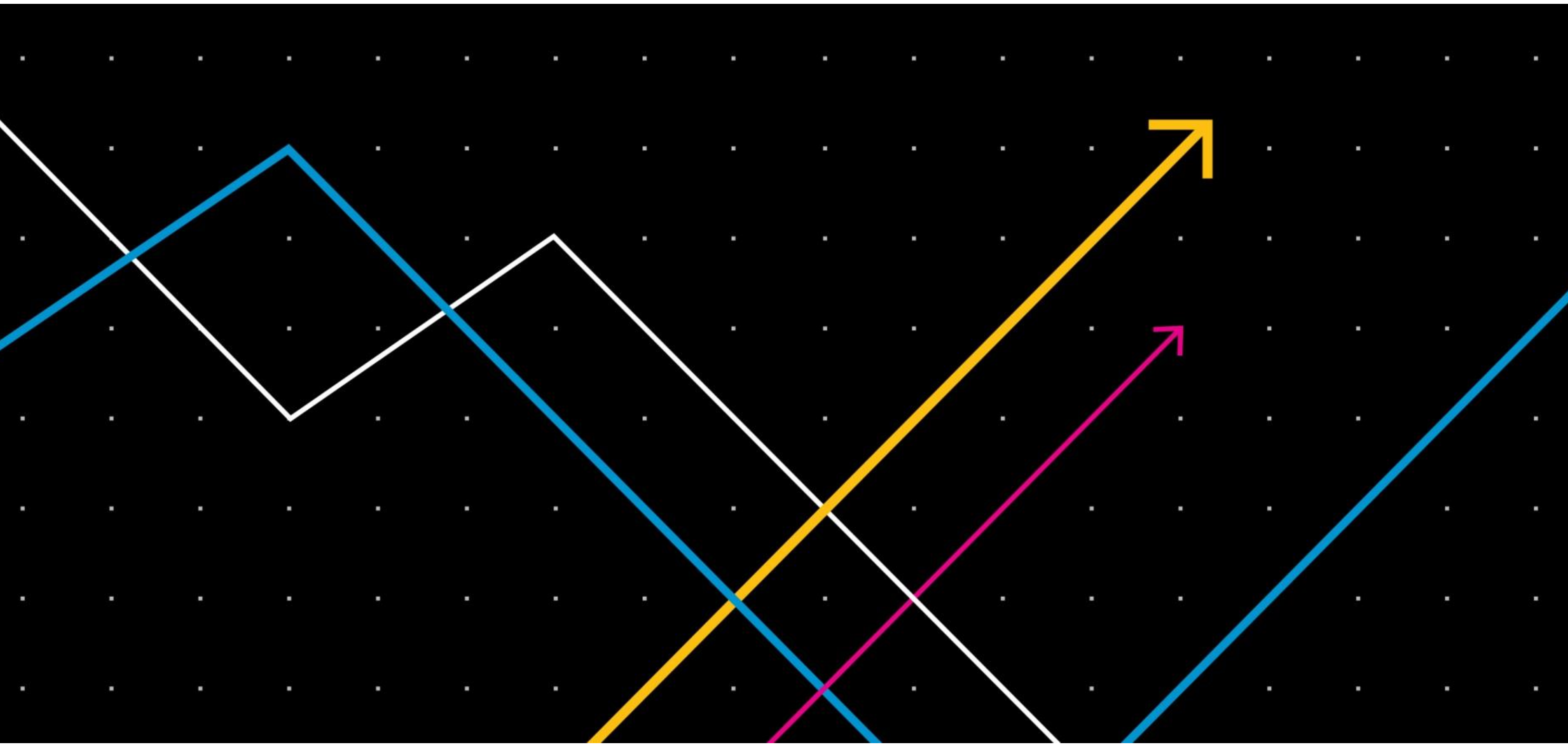


RESEARCH REPORT

A Blueprint for Interagency and Cross-Jurisdictional Data Sharing

Nancy La Vigne, Ellen Paddock, Yasemin Irvin-Erickson, KiDeuk Kim, Bryce Peterson, and Samuel Bieler

July 2017

URBAN INSTITUTE · ELEVATE · THE · DEBATE

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is dedicated to elevating the debate on social and economic policy. For nearly five decades, Urban scholars have conducted research and offered evidence-based solutions that improve lives and strengthen communities across a rapidly urbanizing world. Their objective research helps expand opportunities for all, reduce hardship among the most vulnerable, and strengthen the effectiveness of the public sector.

Contents

Acknowledgments

This research was supported by Cooperative Agreement No. 2012-R2-CX-0001 awarded by the US Department of Justice National Institute of Justice. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban's experts. Further information on the Urban Institute's funding principles is available at www.urban.org/support.

We would like to thank the Metropolitan Police Department, the Prince George's County Police Department, the Metropolitan Transit Police Department, the Washington Metropolitan Area Transit Authority, the Court Services and Offender Supervision Agency, the Maryland Department of Public Safety and Correctional Services, and Walls and Associates for providing the data for this project.

We would like to also thank Eric D. Martin of the National Institute of Justice, Chief Ronald A. Pavlik Jr. and Deputy Chief Kevin Gaddis of the Metropolitan Transit Police, Chief Mark Magaw and Deputy Police Chief Henry Stawinski of the Prince George's County Police Department, Mark Coggeshall and David Huffer of the Court Services and Offender Supervision Agency, Marcy Plimack of the Maryland Department of Public Safety and Correctional Services, Don Walls of Walls and Associates, and Rolf Pendall, Christina Stacey, Akiva Liberman, Helen Ho, Edward Mohr, and Annie Gurvis of the Urban Institute for their support, feedback, and insights.

Executive Summary

Crime analysts today have an unprecedented set of tools available to explore the factors that influence when, where, and why crime occurs. Public, nonprofit, and for-profit organizations are generating and disseminating more high-quality data than ever before, largely because of enhanced storage and retrieval capacity. Similarly, rapid increases in computing power have facilitated the development of more powerful tools to analyze these data. These advances have exciting implications for geospatial analyses of crime and criminal justice data, particularly those that span different sectors and jurisdictions. Integrating data across geographic boundaries and from different fields, such as transportation, land use, and public health, allows researchers and practitioners to better understand the full range of factors that affect public safety. These insights can in turn inform the design of more impactful and responsive public safety strategies.

This blueprint, written to inform the efforts of researchers and analysts in local government agencies and in research settings, offers practical strategies for executing successful data integration projects across agencies and jurisdictions. It combines lessons learned from a wide-ranging literature review with the direct experience of Urban Institute (Urban) researchers, who collected, integrated, mapped, and analyzed interagency and cross-jurisdictional data from the Washington, DC, metropolitan area, a project referred to throughout this blueprint as Metropolitan Crime Mapping. The goal of this blueprint is to encourage similar projects by identifying the opportunities that cross-sectional analysis offers, suggesting strategies to overcome barriers that researchers may encounter, and providing an overview of what the future holds for cross-sector data sharing and analysis.

The Background and Challenges of Data Sharing

Chapter 1 explores the history and theory of cross-jurisdictional and interagency data sharing and cites examples of key insights that emerged as a result of data integration projects and how agencies used this intelligence to inform crime reduction strategies. The chapter then lays out several major data integration challenges that crime analysts will need to navigate, including the following:

- **Resources.** Often the first concern surrounding data integration, resource constraints—including staff time—will determine the scope and sustainability of any data-sharing effort. Many agencies may also face substantial challenges using their data systems for analytic or practical purposes, especially in the early stages of transitioning to digital or automated systems. This limited technical infrastructure may impose further resource demands on data integration efforts.

- **Technological challenges.** A key challenge of data integration is converting datasets so that they "speak the same language" and can be feasibly combined and analyzed. Supporting and maintaining data security is another major challenge.

- **Agency culture and politics.** In many cases, data sharing requires overcoming cultural and political barriers that may include an aversion to information sharing. Agencies can address these issues by working with all partners to identify the shared goals and benefits of a data-sharing and integration project.

- **Staffing and management within individual agencies.** A data-sharing agreement between agencies will have little utility if staffing and management in the partner agencies do not support data sharing or if frequent turnover among those assigned to oversee data sharing weakens lines of communication between partners.

- **Identifying shared goals, benefits, and language.** To bring together the wide array of partners needed, data integration projects will frequently require translational efforts to build a common language and shared set of terms. Each partner will have their own language and jargon; without a shared language, these variations will create friction in a data integration project.

- **Agency staffing and management.** Staff members in charge of executing and overseeing data integration projects need substantive and technical expertise to efficiently manage project demands. Similarly, data integration partnerships may be weakened without the appropriate culture or lines of communication in place to introduce new leadership and management to data-sharing protocols.

- **Central leadership to promote system utility.** Coordination across partners is essential to maintaining the momentum of data-sharing efforts over time. Stable central leadership can improve the effectiveness and efficiency of these ventures, but it can be challenging to identify long-term leadership structures that will be accepted by all project partners.

- **Issues of public access.** In some cases, efforts to share data as a tool for research or practice may be accompanied by a desire to allow greater public access to the data. Though public access can be a strong demonstration of transparency, agencies may fear exposure to public scrutiny or worry that releasing information will have a negative economic impact on their communities.

When considering these challenges, each potential data-sharing partner will need to weigh the costs and utility of integrating data. It is important to note that data-sharing and integration projects range from one-time, ad hoc endeavors to ongoing systems. Though this report draws primarily on Urban's experience conducting a one-time data integration project to highlight new relationships between crime and geographic and environmental factors, agencies should strive to implement structures that support ongoing data sharing in a way that allows them to monitor trends and rapidly respond to new issues that arise.

Key Steps in Data Sharing

Chapter 2 walks through seven steps necessary to implement data sharing, integration, and analysis across agencies and jurisdictions and includes practical lessons on how to execute such efforts more efficiently.

- **Develop a framework for data integration.** Projects should begin by establishing what questions researchers and practitioners would like to answer by sharing and integrating data. This research framework should be shaped by both theory and practice, ensuring that the results of data integration will be relevant to practitioners (problem-driven approach) while allowing for more proactive exploration of new relationships between crime and other variables (theory-driven approach).
- **Organize the research team.** Data integration projects need clear but flexible leadership that provides the structure to make progress as well as the flexibility to pursue new opportunities as they arise. A central project manager who oversees the work of several subject matter experts can help balance these competing priorities.
- **Identify data sources.** When identifying the right data sources to support cross-sector projects, agencies must balance several considerations, including data availability and quality, provider willingness to share data, the utility of the data's content, and the effort needed to analyze data based on how they are formatted and organized. Partners providing data should be engaged regularly and consistently through both organizational and individual relationships. At the same time, the team should take care to avoid redundant requests and be judicious in when and how often they request new data. A timeline developed in collaboration with all partners can provide a basis for accountability and help ensure that efforts move forward.

- **Solicit partners and manage relationships.** Soliciting partners to support the project can be a delicate process, particularly when the data involved could create public relations or operational risks for an agency. To build productive relationships and address these concerns, agencies must demonstrate the value of integrated data projects to prospective partners and ensure that both operational and technical frameworks exist to maintain the security of project data.
- **Create data management structures.** The research team will need to find ways to structure and manage project data so that analyses can be completed efficiently. Data from partner agencies will need to be cleaned, coded, and reconciled. Reconciliation is particularly important, and the team will need to devise a common framework for coding data when jurisdictions use different terms to describe similar observations. Likewise, when data is stored at different geographic levels (e.g., census tracts and block groups), the team will need a strategy for managing data across these different units of analysis. Developing a data dictionary will be an invaluable step in this process.
- **Integrate data.** After structuring data to facilitate efficient management, the team must integrate the data so it can support analysis. Data can be integrated at the individual or place level, but integrating data at the place level is often easier.

Analytic Approaches

Chapter 3 describes analytic approaches to integrated data exploration, providing guidance on how to select appropriate methods from a range of possible analytic strategies. Drawing on Urban's experience with Metropolitan Crime Mapping, the chapter focuses on a wide array of analytic methods, including commonly used approaches such as cross-sectional (point-in-time) and panel (longitudinal/change-over-time) analysis techniques. The chapter also provides an overview of more advanced geospatial techniques, including geographically weighted regression, risk terrain modeling, and Markov transition matrix modeling. Written specifically for an audience of crime analysts and quantitative researchers, the chapter discusses challenges and strategies associated with these analytic approaches and how data integration can provide access to on-demand responses to questions of interest to local governments.

The Future of Interagency Data Integration

Chapter 4 explores the future of data sharing and integration and looks at the opportunities, trends, and challenges that will shape the field going forward. The chapter begins by exploring the changing nature of data use in policy discussions and how these trends may push agencies to explore more data-sharing opportunities. It then assesses how recent developments in cloud computing, data portals, big data processing, and the diffusion of smartphones may help push data integration forward and greatly expand the utility of routinely collected data for informing policy decisions across a wide range of fields. The chapter also discusses issues data integration will pose for civil liberties and privacy and concludes by acknowledging the challenges integration efforts will face in the future, including an ongoing resistance to data-sharing practices among many agencies and a lack of access to necessary resources and technical expertise. Addressing these challenges will be essential to maximizing the potential of integrated data projects.

Introduction

Finding, exploring, and testing relationships between variables is the heart of criminology. Fortunately, continuing advances in information technology are removing many of the data storage and processing barriers that have previously limited researchers' ability to explore new criminological research inquiries. However, though technological barriers to criminological research are tumbling down, organizational barriers to accessing the necessary data remain formidable. The data researchers need are often held by different agencies and multiple jurisdictions. A researcher attempting to build a complete picture of crime and the factors that affect it may need to solicit data from city, county, state, and even federal agencies. These agencies may include police departments, courts, human services agencies, and a host of other government, nonprofit, and for-profit entities. Without a doubt, developing these cross-silo data-sharing partnerships presents unique challenges.

But many lessons can be learned from projects that have undertaken such efforts, offering strategies to mitigate or overcome challenges and execute successful projects. This blueprint explores the literature on cross-jurisdictional, interagency analysis and illustrates the various models that have been employed and documented across the country. Though it draws from a wide array of literature within and outside of the field of criminology, most examples in this blueprint are derived from the experiences of Urban Institute (Urban) researchers who, through a National Institute of Justice grant, collected, integrated, mapped, and analyzed interagency and cross-jurisdictional data from the Washington, DC, metropolitan area, a project referred to throughout this blueprint as Metropolitan Crime Mapping. Using these data, the research team explored questions that could only be assessed with integrated cross-silo data, such as the effect of economic development on crime and the impact of cross-jurisdictional data on the accuracy of predictive crime mapping.

Chapter 1 of this blueprint draws from both theory and practice to provide a brief history of cross-jurisdictional, interagency data mapping and illustrate the associated challenges and advantages. It also describes the various forms that cross-silo data integration can take, from one-time or periodic data sharing to system-wide, real-time data integration. Though the latter model is what agencies should aspire to, this blueprint focuses on the former, as the basic processes associated with one-time data-sharing projects form the basis for a wide array of data integration models.

Chapter 2 describes these processes in detail, laying out the steps necessary to execute cross-jurisdictional, interagency mapping and analysis. These steps include soliciting data-sharing partners, managing those relationships, identifying useful data sources, and creating a structure for managing those data. Chapter 2 also offers strategies for working across agencies and jurisdictions, managing different data definitions and geographies, and interpolating data where necessary. It addresses the time and resource commitments that jurisdictions can expect to make if they embark on a one-time data integration project. This chapter draws from the experiences of the Metropolitan Crime Mapping project and from other examples from the field, including an informal survey of multijurisdictional data integration projects based in metropolitan areas (see Kingsley, Coulton, and Pettit 2014).[1]

Chapter 3 discusses analytic approaches to integrated data exploration and helps readers identify appropriate methods for analyzing data and confirming results. Drawing from the research questions explored in the analysis of data collected from the Washington, DC, metropolitan area, this chapter describes the challenges associated with various analytic approaches and strategies to overcome them. It then discusses how data integration can support rapid responses to questions of interest to local governments. The final chapter looks forward to describe what the future holds for integrated criminal justice analyses and the promising advances in information technology that can facilitate these efforts.

This blueprint is intended to inform the efforts of criminal justice analysts in police and supervision agencies; of city, county, and mayoral staff interested in connecting the dots between populations, services, and outcomes across service providers and jurisdictions; and of researchers looking to develop better, more accurate models that yield findings of value to the research and practitioner communities. Ideally, the experiences documented here can help move the field forward toward a richer, more nuanced exploration of crime in the interest of more effective crime control and prevention.

Chapter 1. Background and Review of the Literature

Cross-jurisdictional, interagency data sharing can have enormous implications for how quickly and effectively law enforcement agencies can detect, respond to, and prevent crime. For crime analysis, data sharing can yield new patterns, reveal new causal factors that drive crime, and inform crime prevention strategies (La Vigne and Wartell 2001; Maltz, Gordon, and Friedman 1990; Santos 2012). This data-driven approach to law enforcement can be further enhanced by analyzing crime data alongside a wide array of socioeconomic indicators, including those from nonjustice agencies. Geographic information system (GIS) mapping has become an increasingly widespread tool for these analyses, allowing agencies to integrate different data sources geographically. The use of GIS in crime and criminal justice research and practice has gained momentum as the technology has become increasingly accessible and user friendly (Chainey and Ratcliffe 2013; Davidson 1981; Raleigh and Galster 2014; Skogan 1986).

Though law enforcement agencies are often at the core of crime data–sharing networks as the main data collectors, a range of other agencies can benefit from such networks as well. Community supervision agencies, for example, can use law enforcement data to enhance their knowledge of police contacts with people on probation or parole and learn of people or activities in the area that may put supervisees at risk of harm or engagement in criminal behavior. Such information has the potential to facilitate more successful reentry outcomes by giving supervision officers a better understanding of the community contexts to which people are returning (IACP, n.d.; La Vigne, Cowan, and Brazzell 2006).[2]

Partnerships with agencies outside the realm of criminal justice may provide valuable information for crime analysis and criminal justice research while engaging a broader group of local leaders in public safety efforts (Hawkins 2006; Wolf 2012). In Camden, New Jersey, for example, police are working with local hospitals to identify patterns and overlapping high-risk populations by comparing police data on arrests and calls for service with emergency department usage, ambulance calls, and other information.[3] Jurisdictions may also find it beneficial to share data with the public to promote greater transparency, accountability, and public trust, and to encourage community participation in public safety (IACP 2015; President's Task Force on 21st Century Policing 2015).[4] But no matter the application, researchers and analysts will generally find that they need both cross-jurisdictional and interagency data to develop the most complete public safety analyses.

Cross-Jurisdictional Data

Politically defined jurisdictional borders, though important for assigning governmental responsibility and for administrative purposes such as census taking, tax collection, and mail delivery, are quite permeable in daily life (Rengert and Lockwood 2009). This creates a unique challenge when developing public policy, as questions of responsibility can lead to “edge effects”—increased rates of crime and criminal victimization along boundaries between relatively homogenous areas—and territorial gaps where policies to address crime risk falling short (Briffault 1996). This is often seen in the context of environmental issues such as wildlife conservation and pollution (Helland and Whitford 2003; Woodroffe and Ginsberg 1998) and is certainly true in the field of crime control and criminal behavior. A substantial body of research supports the idea that crime tends to be concentrated in hot spots that are often much smaller than typical geographic units and may be as small as a few street blocks, a street segment, or a single address (Rengert and Lockwood 2009; Weisburd, Bruinsma, and Bernasco 2009).

Typically, these hot spots are defined through the mapping of crime within a single jurisdiction. Yet crime often works in patterns that ignore jurisdictional boundaries, and people committing these offenses may move freely across city, county, or state lines (Ratcliffe 2003; van Schendel and Abraham 2005). Moreover, there is evidence that some people may strategically commit crimes across jurisdictional boundaries to reduce the likelihood of detection and apprehension (Finklea 2011; van Schendel and Abraham 2005).

The need to analyze crime across political boundaries becomes even more acute when one considers that the areas immediately adjacent to borders are often hot spots themselves. Crime clusters at these borders for several reasons. For example, boundaries frequently overlap with major roads and highways, which provide convenient opportunities for crime and allow people to travel unnoticed. Mixed land uses and geographic features also generate and provide opportunities for crime (Brantingham and Brantingham 1995; La Vigne and Wartell 2001; Song et al. 2015). Crime patterns are neither distributed evenly within political boundaries nor limited to single jurisdictions, and approaching crime through a framework of such boundaries can impede public safety efforts.

Interagency Data

Researchers and practitioners have long known that crime does not occur in a vacuum, even within a single jurisdiction. People and their physical and social environments interact to create opportunities for crime to occur (Brantingham and Brantingham 1993, 259–94; Taylor and Harrell 1996).

Researchers interested in preventing crime by identifying its drivers have noted links between crime and specific environmental and social factors, such as local unemployment rates, the presence of vacant buildings, and certain land use and zoning characteristics (Kinney et al. 2008; Krivo and Peterson 1996; Markowitz et al. 2001; Raphael and Winter-Ebmer 2001). These findings present a compelling case for interagency data sharing and suggest that any understanding of local crime dynamics will be quite limited if it incorporates only conventional crime data.

The History of Interagency and Cross-Jurisdictional Data Sharing

Agencies from different disciplines and across jurisdictional boundaries have shared data, including public safety data, for decades, though typically on an informal, ad hoc basis in response to specific circumstances (Carter 2004, 29–53; Ratcliffe 2012). More recently, some agencies have begun creating more in-depth data-sharing networks in the form of integrated data systems, which are updated on a continuous basis. A 2013 survey conducted by a coalition of national data-sharing advocacy organizations found 30 operational integrated data systems across the country, including 5 at the state level and 25 at the city or county level.[5] These data systems often engage a wide variety of agencies to better understand a specific topic. For example, the University of Chicago-led Integrated Database on Child and Family Programs in Illinois explores the relationships between education, youth, and other data of interest (e.g., employment data) through partnerships with the Chicago Workforce Investment Council, the City Colleges of Chicago, Chicago Public Schools, and the Chicago Department of Family and Support Services.

For law enforcement, information sharing evolved alongside the development of new technologies and increasingly comprehensive and consistent record-keeping practices. Such practices were promoted in the 1920s amid efforts to professionalize policing and included the archiving of arrest records and early individual criminal case files (Archbold 2012). In the 1970s, police departments began to routinely monitor and record citizen calls for service as part of a new emphasis on developing collaborative relationships with citizens and the valuable information they could provide (Archbold 2012; Trojanowicz and Bucqueroux 1998).

Data sharing at the local and regional levels gained momentum in the 1980s and 1990s in response to public and political concerns over drug trafficking and gang activity. At the same time, the growing popularity of computers and the internet facilitated the development of data systems that could be accessed and updated from multiple locations and by multiple users (Carter 2004, 29–53; Nunn 2001; Sheptycki 2004; Schwabe, Davis, and Jackson 2001). These developments accelerated the shift to computerized records, though paper records persist in some jurisdictions.6

In the wake of the September 11 terrorist attacks, federal pressure for law enforcement to engage in interagency and cross-jurisdictional data sharing became much more acute. This pressure was accompanied by an increase in resources to support data sharing as officials at all levels of government sought to improve their disaster management capabilities and close information gaps now seen as national security threats (Kapucu 2006; Williams et al. 2009). The resulting movement toward data sharing has focused on promoting communication and collaboration between local, state, and federal law enforcement and intelligence agencies to prevent terrorism through national data-sharing platforms such as the FBI's National Data Exchange.[7] Perhaps more significantly, the federal push for data integration has also presented a significant challenge to the "need-to-know" culture of intelligence sharing that had previously held sway among many local and federal agencies (Budinger and Smith 2011; Carter 2004, 29–53).[8]

As the culture around data sharing changes, new analytic tools have enabled agencies to interpret and use data in more powerful ways. In particular, crime mapping with GIS has significantly influenced how law enforcement agencies address crime, allowing departments to identify hot spots of crime and more strategically deploy resources. Though efforts to analyze the relationship between crime and place have occurred at least since the 19th century (Weisburd, Bruinsma, and Bernasco 2009), new technologies have changed how this research is conducted. The first experimental computer-based crime-mapping technology was developed in the 1960s but was rudimentary and still prohibitively expensive for most jurisdictions. By the 1980s, however, analytic software could map large and diverse datasets and independently identify patterns (Coppock and Rhind 1991). Computer-assisted crime mapping has since become increasingly widespread in the United States and is now a mainstay of large metropolitan police departments, thanks to advances that have made the technology more accessible and cost-effective (Chainey and Ratcliffe 2013). The utility of this technology has also been greatly enhanced by interagency and cross-jurisdictional data sharing, enabling law enforcement agencies to visualize crime patterns across borders (La Vigne and Wartell 2001; Mamalian and La Vigne 1999; Rich 1995).

Despite recent advances in data sharing and geographic analysis, most jurisdictions still lack a strong infrastructure for interagency and cross-jurisdictional data sharing. The scope and technological capacity of public safety data-sharing networks vary greatly, and these networks often consist of ad hoc agreements between narrow sets of criminal justice–focused stakeholders. Further, efforts to establish regional data-sharing partnerships of the kind described in this blueprint may be relatively informal and, as a result, poorly documented or publicized. The following section draws from existing documentation of regional data-sharing efforts to offer lessons for future endeavors.

Regional Data-Sharing Systems

Data-sharing systems vary in scope, organizational structure, and membership. One of the largest and earliest examples of a formal regional cross-jurisdictional data-mapping system is the Regional Crime Analysis Geographic Information System (RCAGIS), established in 1996 by the US Department of Justice in collaboration with the Baltimore County Police Department and the Regional Crime Analysis System group. The Regional Crime Analysis System, a basic data-sharing system, grew out of a previous partnership between the Baltimore County Police Department, Baltimore Police Department, Maryland State Police, and law enforcement from surrounding counties that led to the successful clearance of a series of high-profile armed robbery cases. A subsequent partnership with the Department of Justice allowed the Regional Crime Analysis System group to greatly expand its geospatial mapping abilities. RCAGIS quickly proved capable of addressing crime, though early implementation issues highlighted opportunities for improvement. For example, agencies had to enter data once into their own systems and then again into RCAGIS, and some member agencies also lacked the technical software capacity to use the mapping functions of RCAGIS (La Vigne and Wartell 2001).[9]

The partners behind RCAGIS came together of their own volition because of an identified need for cross-jurisdictional information. In other cases, systems have been driven by or housed within more formal public entities. Michigan's Courts and Law Enforcement Management Information System (CLEMIS) provides a mechanism for interagency data sharing and includes every law enforcement agency in Oakland and Washtenaw counties as well as several from nearby counties. Though law enforcement agencies form the core of this network, CLEMIS is open to all public safety agencies in Michigan and counts among its members several 911 central dispatch agencies, a prosecutor's office, parks police, the Wayne County Airport Authority, the FBI, Michigan fusion centers, the Michigan State Police, federal Immigration and Customs Enforcement, and the US Secret Service. Since 1982, the database has integrated a wide range of data sources and functions, including police computer–aided dispatch and records management systems, 911 calls, evidence records, crash reports, biometric facial images and fingerprints, crime mapping, and so on (Oakland County 2014). Critically, CLEMIS is also institutionalized as a division of the Oakland County Department of Information Technology, with approximately 30 employees and around-the-clock technical support.[10]

Similarly, San Diego County's Automated Regional Justice Information System was established in 1980 and includes all municipalities in the county, the county government, and their representative law enforcement agencies.[11] Though it integrates many of the same types of data sources as CLEMIS, it places additional emphasis on making data more accessible to officers on the ground through mobile applications such as the Tactical Identification System, which allows officers to compare a photo of a suspect against the local booking database, and the Tactical Automated Response Using GIS-Enabled Technology, which provides geospatial information such as the locations of police incidents and gang activity, the addresses of people on parole and those registered as sex offenders, and more.[12]

In some cases, neighboring jurisdictions may gain access to each other's data when both—independently or through agreement—subscribe to a common data-sharing service that shares data among its members. For example, the Pennsylvania-based CODY Systems platform integrates records management systems, dispatch data, and other criminal justice data sources into one system and includes a mobile access component for officers on patrol. Over 500 clients can also access a specific "vendor-neutral" data-sharing tool that allows officers to query cross-jurisdictional data regardless of differences in data vendors or infrastructure.[13]

These examples of cross-agency and cross-jurisdictional data sharing and integration illustrate the unique challenges agencies face when developing data-sharing networks. At the core of many of these challenges is the reality that agencies come to the table with different budgets, technological and human capacity, data infrastructure, agendas, and cultures. These challenges, and their implications for data sharing and integration, are discussed below.

Challenges

Effective data integration, whether through periodic ad hoc sharing or a fully automated system, requires a common language and common standards and expectations to ensure that data sharing is beneficial to all involved. This section provides more detail about the difficulties agencies are likely to encounter in three specific areas: resources, technology, and interpersonal or political factors.

Resources

One of the greatest challenges agencies can face when implementing a data-sharing system is securing and maintaining the resources necessary to support the system, particularly with the resources needed for computer-based data sharing. Though the up-front costs of software development and deployment often command the greatest investment (La Vigne and Wartell 2001), participating agencies must also ensure that they can fund updates and other maintenance for the system to remain efficient. Additional costs arise from the vast difference in data efficiency between partners: some more well-resourced agencies may have relatively advanced data management systems and analytic tools and well-trained staff, but smaller and particularly rural jurisdictions may still rely on paper documents or be in the early stages of digitization (Jackson et al. 2014; Kaza and Chen 2008). Such agencies may be important partners for public safety purposes but will most likely need to develop their internal systems before they can participate fully in a software-based data-sharing and mapping arrangement. In some cases (e.g., situations where there is a movement to expand the public availability of data), agencies may incur additional legal costs ensuring compliance with open data requirements and other relevant legislation.[14]

Technology

Interoperability is another significant challenge agencies face when attempting to share data. Even if a participating agency has a strong internal data system, that system may be very different from those of other agencies and may not be able to easily communicate with those systems (Kaza and Chen 2008).[15] Initial barriers to interoperability include differences in data labels and terminology: even if data are shared through the same format or systems, agencies may use different terms to describe the same thing or assign different meanings to the same term (Maltz 1999; McCormick et al. 2015; Swartz 2008). For fully integrated databases, interoperability also involves reconciling the different programming languages used to create each agency's unique system (Chisnall 2013).

Data security is also paramount, and systems must adequately protect against online breaches. In some cases, this might require adjusting the level of access for different user accounts (Agrawal, Evfimievski, and Srikant 2003). Other challenges include ensuring timeliness of data, developing efficient entry systems that only require data be entered once, and training staff to properly clean and enter data in those systems and to understand how to best use the system for public safety purposes.[16] Chapter 4, The Future of Interagency Data Integration, provides more information on challenges that arise from the use of new technologies as data sharing and integration tools.

Agency Culture and Politics

Law enforcement agencies are often especially—though not uniquely—protective of their information, creating significant challenges for data sharing (Linden 2003).[17] A summary of lessons learned from the National Neighborhood Indicators Partnership notes that "[f]rom the agency's viewpoint, they have much to lose by granting access to their data, and the easiest response for them in the short-term will be to reject the request without serious consideration."[18] Efforts to create interagency data-sharing partnerships must therefore begin by establishing trust among all parties and by making clear (1) the value of being involved and (2) that sharing data will not unreasonably compromise control over internal data systems (Pardo, Gil-Garcia, and Luna-Reyes 2010).[19]

Identifying Shared Goals, Benefits, and Language

A common characteristic of all successful data-sharing efforts is that they began with agencies cultivating strong relationships—both within their own jurisdictions and across neighboring jurisdictions—and identifying how each partner will benefit from the arrangement (Harris and Romesburg 2002).[20] Partnerships involving both criminal justice agencies and agencies in other fields are especially susceptible to differences in language and terminology creating barriers to mutual understanding. Just as variation in data labels can keep data from being efficiently integrated, variations in terminology can lead to miscommunication and even friction among agencies. It is therefore critical to long-term success to ensure that agencies literally and figuratively "speak the same language" and build strong relationships that facilitate smooth data sharing (Bouhaddou et al. 2008; Cheminais 2009, 1–22).

Agency Staffing and Management

Management issues, both within and across agencies, can put the success of data-sharing networks at risk. For example, IT and analysis units in police departments are often overseen by sworn officers rather than civilian data experts. These officers may have little technical or management experience, and in some cases, frequent turnover among leadership may prevent them from acquiring that experience. Leadership turnover in individual agencies can also weaken the larger partnership if an appropriate process or culture is not in place to ensure that new leadership is rapidly brought on board to any data-sharing agreement (Clingermayer and Feiock 1997; McGillivray and Smith 2004; Nedović-Budić and Pinto 2000).

Central Leadership to Promote System Utility

Designating cross-agency leadership can greatly expand the utility and efficiency of data-sharing systems. In many cases, a few people—or even a single person—are responsible for driving initial data-sharing efforts and may shift naturally into a leadership role during the early phases of collaboration (Cheminais 2009, 1–22). But stable and effective overarching leadership can be difficult for jurisdictions to achieve. In partnerships where leadership involves an unpaid position, perhaps appointed or elected by constituent members, it may be difficult to find someone who is sufficiently agency-neutral, knowledgeable, and willing to provide leadership in addition to his or her regular responsibilities. Creating one or more full-time positions to coordinate data-sharing efforts can address the issues of expertise and neutrality and improve efficiency, but such positions may be challenging for jurisdictions to fund (Emerson, Nabatchi, and Balogh 2012; Gazley 2008; Pardo et al. 2006).

Issues of Public Access

Unique political issues arise when data sharing includes providing greater access to the public.[21] Though this can be an effective way for agencies to demonstrate transparency and encourage public partnership in public safety efforts, agencies may also fear how released information will affect public perception and the economic prospects of areas revealed to have high crime rates. Such data may drive away businesses and potential residents alike, reinforce negative perceptions of neighborhoods with high crime, and suppress commercial activity.[22]

Weighing the Utility of Interagency and Cross-Jurisdictional Data Sharing

Data sharing can carry substantial costs but can also yield enormous benefits to public safety in the form of more efficient access to information, improved coordination in problem-solving and crime prevention activities, and better clearance rates on cases. Most significantly, analysis of multisource data can help jurisdictions reduce crime by identifying patterns that give insights into the root causes of crime, allowing police to more proactively prevent incidents (La Vigne and Wartell 2001). Such systems have also been found to be cost-effective. For example, an independent consultant estimated that the Automated Regional Justice Information System created total annual savings of $13,871,167 for participating agencies.[23] However, the utility of implementing such a system should be considered on a case-by-case basis.

Despite a significant amount of research on the relationship between crime and environmental factors, the utility of this knowledge will depend on how agencies use it. In an ideal scenario, data would be continually collected, integrated, and analyzed by local agencies that are positioned to do so sustainably. Neighborhoods are unique and dynamic ecosystems shaped by continually changing interactions between physical characteristics and resident mobility (Galster and Hedman 2014), and practitioners are best served by data integration that occurs on a systematic basis and works with data specific to the area (Colvin and Goh 2005; Lindsay, Jackson, and Cooke 2010; Maltz and Targonski 2002).

In practice, however, such frequent and well-managed data integration is rare. But several new technologies have created opportunities for more timely integration at lower costs (see chapter 4, The Future of Interagency Data Integration). Thus, the majority of this blueprint draws on Urban's experience conducting a one-time, researcher-driven data integration project to explore how such data may be used. Even as agencies strive for more timely and ongoing data collection and integration, one-time data collection conducted by practitioners or researchers can provide valuable, location-specific insights into the relationship between crime and neighborhood factors while highlighting data sources that could contribute useful information if built into a longer-term data-sharing framework.

Chapter 2. Sharing Data across Agencies and Jurisdictions

Even sophisticated jurisdictions with internal data warehouses in place will frequently need to undertake extensive data collection and management efforts when developing cross-silo data analyses. Because the data required for analysis will often be held by multiple agencies and jurisdictions, analysts developing these projects will need strategies for reaching out to partners, identifying valuable data sources, and developing management systems to effectively use the data they acquire. This process is equal parts data expertise and diplomacy: analysts must understand their partners' needs and requirements for data access and have the expertise to effectively integrate different sources of data.

This chapter begins by exploring how to lay the groundwork for cross-silo projects by determining research questions, identifying the right data sources, selecting partners that can provide access to those sources, and managing those relationships. It then discusses how to organize and structure these data with an emphasis on managing different levels of geographic analysis and addressing differences in data definitions. Finally, the chapter draws on Urban's experience directly collecting and integrating multijurisdictional data from across the Washington, DC, metropolitan area and offers actionable strategies for government analysts and policy researchers interested in undertaking cross-silo research.

Developing a Research Framework

At the core of a successful cross-silo project is a framework that lays out the key principles that will help form specific research questions and subsequent project execution. This framework is typically shaped by a combination of theoretical insights and practical problems facing the participating jurisdictions.

Theory-Based Framework

A theory-driven approach to cross-silo projects seeks to answer questions that explain and link concepts such as the underlying factors driving social cohesion or juvenile violence. A theory-based framework focuses a project on understanding the extent to which these links either support or refute existing theories within the larger context of criminological research. These theories are frequently capable of informing practical public safety efforts, but their first-order objective is to explore relationships between concepts.

Metropolitan Crime Mapping, for example, drew heavily on the theory of "the criminality of place," which suggests that certain small areas ("places") have characteristics that make criminal activity more likely (Brantingham and Brantingham 1995). Generally, such places fall into one of three categories: crime generators, crime attractors, and crime enablers. *Crime generators* are places that attract large numbers of people, such as shopping malls, transit stations, and sporting events. The density of people presents those contemplating a criminal act with ample opportunities for crime, thus making offending more likely. *Crime attractors*, such as bars or nightclubs, draw people motivated to commit crimes because they are known to present good opportunities to do so. *Crime enablers* afford criminal opportunities because of their lack of regulation or enforcement of rules (Brantingham and Brantingham 1995).[24] For the Metropolitan Crime Mapping project, the decision to explore the relationship between crime and metro stations was based partly on the idea that metro stations can serve any of these three roles. Metropolitan Crime Mapping focused on spatial relationships between crime and other data sources, and the theory of criminality of place heavily influenced the decision to combine data that showed, for example, the relationship between crime and businesses or the trends in bike thefts at transit stations. This theoretical framework also informed the decision to explore whether metro stations change from crime generators to attractors based on time of day.

Theory-based frameworks are essential to advancing the larger study of criminology and frequently offer valuable insights to frontline agencies charged with addressing particular public safety problems. Theory-based ventures may offer high-level insights into what should be done to address a problem, but they frequently do not provide specific, actionable strategies for agencies seeking to remedy a particular issue. For this reason, researchers employing primarily theory-based frameworks should take care to consider how their findings might inform the challenges facing the agencies contributing data to their work and use the insights they garner to inform policies and practices.

Problem-Based Framework

A problem-based research framework is developed to assess specific problems in a jurisdiction or to develop actionable strategies to address identified problems. A problem-based framework may be informed by theoretical considerations, but its motivating impulse is to develop policy-relevant solutions. Though some organizations' data frameworks will likely be problem based as a matter of course, there are also advantages to actively considering a problem-based framework. A problem-based framework strongly encourages—and frequently requires—the engagement of agencies charged with actually addressing the problem. Because problem-based frameworks require this engagement and may offer solutions to current challenges, they can help engender buy-in and participation from data providers.

Problem-based frameworks focus on identifying and addressing concrete challenges, but analysts should remain alert to opportunities to explore larger theoretical questions. Connecting jurisdiction-specific solutions to broader contexts allows insights from a specific project to inform larger discussions about criminological theory and policy. These connections can in turn encourage the diffusion of best practices and encourage the development and refinement of theories that will enhance jurisdictions' ability to execute effective policy.

Integrated Frameworks

Cross-silo projects are often guided by a framework that responds to or investigates a problem but does so using a theory-based framework. Using theory to guide a project can inform how researchers identify problems and assess solutions, resulting in more efficient and valid analyses that can be connected to the broader literature to advance the field. Research frameworks that steer the project toward actionable insights will also encourage organizational buy-in among data partners, advance the field's knowledge of best practices, and support effective execution of policy.

While developing the Metropolitan Crime Mapping project, researchers identified several interest areas (e.g., crime and economic development, crime and transportation, community supervision and recidivism, gunfire detection technology and mapping, and cross-jurisdictional data sharing for predictive crime mapping) that, based on criminological theory and prior conversations with agency partners in Washington, DC, would likely benefit from cross-silo research. Researchers reached out to these partners to identify the questions and challenges the project should address. The resulting research questions fell somewhere along the spectrum between theory driven and problem driven. Two examples are described below to illustrate different ways theory and practice might interact to develop a larger framework for research as a whole and for research questions.

CRIME AND TRANSPORTATION

Research questions exploring crime and transportation were developed through an exchange of theory identified by researchers and problems identified by the Washington Metropolitan Area Transit Authority (WMATA). Researchers initially focused on crime at transit stations, a decision rooted in the theory that transit stations are frequently hot spots of crime (Tilley et al. 2004) because of an abundance of suitable targets and unique criminal opportunities. Researchers also planned to hone in on rates of "iCrime," a recent form of pickpocketing or theft of personal electronic devices (e.g., smartphones or tablets) that scholars theorize stems partly from rapid growth in smartphone ownership, the high value of phones, and the increasing social importance placed on connectivity facilitated by such devices (Farrell 2015; Hanna, Rohm, and Crittenden 2011; Roman and Chalfin 2007). Such devices, in addition to being both valuable and ubiquitous, are typically small, portable, and easy to steal in an environment where many people pass through constantly and often come into physical contact during particularly busy times of day—an ideal place to both encounter victims and quickly depart from the scene.

However, after discussions with and preliminary data from WMATA, Urban researchers switched track. Though iPhone thefts were certainly an issue at Washington, DC, transit stations, a more pressing question centered around an increase in bicycle thefts at transit stations during the early stages of the project. Thus, despite theory raising iCrime as an issue of concern, the Metropolitan Crime Mapping project ultimately explored and responded to the identified needs and relative lack of knowledge around local bike theft.

GUNFIRE DETECTION TECHNOLOGY AND MAPPING

Another example of theory and identified problems interacting to shape research is the decision to investigate the implications of gunfire detection technology (GDT). Implemented in Washington, DC, beginning in 2008–09, ShotSpotter gunfire detection technology is a relatively new source of data that researchers and practitioners are still seeking to understand. For researchers, GDT is a compelling area of study because of its practical implications, including a faster and more consistent response to gunfire incidents, better situational awareness, a reliable source of evidence, and a useful data source for predictive crime mapping, and because of the potential value GDT might add to existing types of gunfire data (Aguilar 2013; Choi, Librett, and Collins 2014).

Criminology has long been concerned with the idea of a "dark figure of crime": the great volume of criminal incidents not captured by statistics because of underreporting or other reasons (Coleman and Moynihan 1996; Penney 2014). Gunfire detection technology, which relies on automatic sensors rather than human reporting, could overcome many limitations of traditional crime statistics reporting, which include underreporting by citizens, limited police capacity to detect or correctly identify the location of gunfire, institutional influences affecting what numbers are recorded, and so on. (MacDonald 2002; Myers 1980; Klinger and Bridges 1997; Skogan 1974). Of course, GDT has limitations, and a secondary benefit of studying GDT data is an understanding of these limitations and how to enhance the data's value for research and practice. Thus, the decision to delve into the implications of this technology as a measure of gun violence and possible tool for predictive crime mapping was a product of mutual interest in exploring new opportunities.

Organizing the Research Team

Once the framework is developed, project staff must assemble a team and determine who will oversee the project. Whether the cross-silo effort is a one-time project or an ongoing venture, clear areas of responsibility need to be assigned, particularly when specialized expertise is required to answer specific research questions.

The team structure for Metropolitan Crime Mapping needed to provide adequate oversight and expertise to advance each of the five interest areas while managing integration of data from law enforcement, private sector, and transportation partners. Identifying clear lines of responsibility for different project components was thus critical to effectively managing the project. For each research question, a point person was selected based on their technical expertise and experience in the field. For example, the lead researchers exploring crime and economic development had prior experience analyzing the dynamics of neighborhood economic growth, and the lead researcher examining community supervision and recidivism had extensive experience working with probation and parole datasets. The team's principal investigators took primary responsibility for maintaining and building partner relationships in addition to providing overall supervisory oversight throughout the project. A project manager worked under the principal investigators and was tasked with running day-to-day operations and handling the array of administrative tasks endemic to a project involving the integration of so many data sources.

Identifying Data Sources

Once research questions have been identified and research teams formed, analysts will need to identify the right data sources to support their analyses. This requires balancing several considerations, including data availability and quality, provider willingness to share data, the utility of the data, and the effort needed to analyze data based on how they are formatted and organized. This section delves into these considerations and discusses how to optimize project resources when selecting data sources.

Develop Clear Data Collection Goals and Tailor Analytic Inquiries

Cross-silo analyses involve broad inquiries that could be informed by an almost infinite number of data sources, so analysts need to focus their efforts. Without clear data collection goals, the team will struggle to access and clean an ever-widening array of sources, leading to project delays. It is thus essential early on to strategically identify a specific, discrete array of datasets to collect, ensure that they can support the analysis, and concentrate efforts in this area. This process is particularly important given the structure and volume of criminal justice data, which are distributed among many municipal, county, state, and federal agencies; even collecting basic crime data for a county may involve contacting multiple police agencies with varying levels of data sophistication.

For example, Urban researchers wanted to include crime data from Prince George's County, which shares a border with Washington, DC. The team assumed it would be viable to collect data from most jurisdictions in the county, but this quickly proved to be an overwhelming task. In addition to the Prince George's County Police Department, which handles most crime in the county, 24 of the county's 27 municipalities collect independent crime statistics.[25] Collecting data from all of these municipalities would have required substantial time and resources and returned minimal value for the project as only a few areas of the county actually border Washington, DC, and would have been relevant to an investigation of cross-border crime. Realizing this, the team reoriented and focused its data collection efforts on departments in areas bordering Washington, DC, whose data would allow them to explore cross-border crime trends without expending unnecessary effort or resources.

Remain Flexible to Adapt to New Opportunities

On the other hand, though committing to data sources is important for project efficiency, projects can also benefit from maintaining a flexible outlook and a willingness to modify their plans if new data are identified that improve on existing sources or provide a better way to answer the original research questions. Data collection that is responsive to the overall objectives of the research questions and not inflexibly committed to a set plan will help the team answer research questions in the most effective manner possible. For example, new data sources may supplement or replace sources that prove inadequate or inaccessible. To maximize the value of the identification process while minimizing data collection time, researchers may begin with a list of data sources they can collect while leaving room for new opportunities. Effectively managing the identification of data sources will also help researchers conduct outreach to data partners more efficiently and ensure that they target the right people for outreach and limit the time spent soliciting partners.

Metropolitan Crime Mapping demonstrated the benefits of creating a data collection plan and outlook that is flexible enough to be altered as needed. Urban researchers initially planned to assess the relationship between crime, neighborhood change, and economic development using a combination of housing data and socioeconomic markers, such as family income. During the project, the team discovered a new data source: the National Establishment Time-Series (NETS) Database, which offered business location and revenue data at the establishment level. The team adopted this new data source, and its unprecedented degree of nuance allowed researchers to create incredibly detailed neighborhood profiles and paint a more complete picture of the relationship between crime and economic development, including previously unavailable temporal trends in business activity.

The decision to study gunfire detection technology's impact on predictive crime mapping followed a similar trajectory, as new data sources provided ways to answer research questions from a new angle. The team originally focused on using land use and cross-jurisdictional data to assess whether these sources of data could improve predictive crime mapping. Midway through the project, the team became aware that GDT data for Washington, DC, were publicly available. An initial investigation suggested that GDT data had immense potential to improve predictive mapping and could be readily integrated into existing analyses. Given these advantages, researchers quickly decided to reorient their data analysis plans to focus on the value of GDT as a tool for improving predictive mapping. This in turn provided an opportunity to more effectively answer the original research questions.

Not every dataset is suitable for midcourse adoption. It is often difficult to assess whether a particular data source will help answer the research questions until the team can explore its potential and its limitations. Thus, pursuing new data sources can be a gamble given the amount of time needed to negotiate data-sharing agreements, organize data transfer protocols, and assess the data received. Analysts should attempt to assess the suitability of data through interviews, codebooks, or other avenues before requesting data pulls. In the case of the NETS Database and GDT, several characteristics made these data sources attractive options. First, the data were readily accessible, with information made available early on through codebooks or data provider interviews that effectively summarized the content. This dramatically increased the ease of identifying their value as datasets, though this is unlikely to be the case for much of the data analysts will need to conduct cross-agency data analyses.

Minimize Challenges with Strong Communication and a Timeline

Adequate planning and strong communication with data partners can help minimize the challenges of data identification. Research projects will frequently need to conduct multiple data extractions, particularly if the project is new and the team is exploring new data. Analysts will have to work closely with data providers to refine the scope of their requests and obtain new datasets. Project timelines should account for this iterative process and build in time for the team to review data pulls and return to providers for clarification. At the same time, iterative requests should be kept to a minimum because of the time required for each data pull and the demands they place on data providers.

Good communication can also limit redundant information requests by ensuring that both analysts and data providers understand the potential and limitations of the data and how these characteristics correspond to the research questions. Requesting data samples in advance of full data pulls can improve efficiency by allowing analysts to vet the data and establish a clear scope of inquiry before making more extensive data requests. Strong lines of communication with partners also allow data providers to suggest more effective ways for the team to address its data needs and to reveal new opportunities throughout the life of the project.

Soliciting Partners and Managing Relationships

Cross-silo research projects must establish, strengthen, and maintain relationships with the agencies that own and oversee the data sources they will use. Sharing data, especially crime and public safety data, requires a significant level of trust between partners that takes time to develop (Crank 2014; Goldsmith 2005).[26] But once established, this trust and communication helps promote project success by improving the efficiency of data collection, strengthening project members' ability to adjust strategy during the project, enhancing responsivity to changing local circumstances, and increasing the likelihood that research will be relevant to practitioner partners.

Agencies or researchers seeking to form data-sharing arrangements should look first to preexisting partnerships for the data sources they might offer and for input on shaping research questions that are responsive to practitioner needs. Leveraging such relationships also makes data collection more efficient by minimizing the effort needed to reach a basic level of familiarity, trust, and agreement with partner agencies. For this project, Urban researchers benefited from several strong preexisting relationships with key agencies in the Washington, DC, area, including the Metropolitan Police Department (MPD), the Court Services and Offender Supervision Agency, and the WMATA Metro Transit Police.

But for many data sources, agencies will need to establish new partnerships. Getting a foot in the door can be one of the most difficult challenges teams will face and often requires a great deal of persistence to overcome. Fortunately, a few strategies can help projects reach out to local partners holding potentially useful data. First, agencies or individual researchers should identify whether they have any shared contacts (either personal or organizational) with the target agency. A project analyst or team member might be in communication with someone at the target agency, or the project team as a whole may have an established relationship with that agency. Where no such contacts exist, local coordinating bodies can be a useful starting point for identifying the full range of local organizations and potential partners and for initiating contact with member agencies.

After securing the attention of a potential partner, project staff should establish the value of the data request and resulting analyses to the research inquiry and to the larger field of practice. Each partner should have a clear idea of how their agency stands to benefit from the project so that they see the value of putting in the necessary time and resources and of exposing themselves to the real or perceived risks involved with sharing their data. Emphasizing the value of the project and soliciting agencies for their input on research questions is critical to encouraging collaboration. This is particularly true when a request comes from outside researchers, but it also applies to interagency communication, especially when working with agencies with a cultural reticence toward sharing data. La Vigne and Wartell (2001) provide a useful illustration of both the difficulties and importance of this process through their efforts to secure initial buy-in for the Orange County Gang Incident Tracking System, a data-sharing system designed to help map gang activity across jurisdictional lines. In this case, the specific nature of the problem the system intended to address made buy-in particularly difficult to secure:

> On one end of the spectrum, some of the chiefs with more serious gang problems did not want the database created and the “truth” recorded because they were afraid it would affect tourism and cast their city in a negative light. At the other end, cities with little or no gang problem were not sure they wanted to expend the effort and resources on a system from which they would receive little benefit. (La Vigne and Wartell 2001)

The research team overcame this challenge by asking each agency to identify their own pressing research questions—in addition to the larger goals of the project—and asking what types of data would be useful to them in their work.

For both preexisting and new relationships, key partners must be fully on board and should understand the value data sharing adds to their organization. This will help strengthen buy-in and promote the sustainability of the partnership by reducing the likelihood of a partner ending their participation or declining to collaborate.[27] Additionally, cultivating relationships with each data provider will strengthen lines of communication that are essential to developing familiarity with the unique characteristics of each data source. Understanding how each data source is formatted, organized, labeled, and so on is essential to the integration and analysis that occur later. Finally, strengthening partnerships and establishing mechanisms for ongoing collaboration (e.g., regular meetings, phone calls, or updates) can bolster information-sharing pipelines and open the door to more informal communication, which can be used to explain a particular data problem during analysis or to share analysis strategies (Markovic and Stone 2002; National Institute of Justice 2008). Above all, it is essential to view relationships with other agencies *as relationships* and to not terminate or downgrade communications once data have been received. A continuous feedback cycle with partner agencies will ensure that data providers remain connected to research progress and goals, that partners have ample opportunity to provide input, and that research continues to be responsive to practitioner needs.

Agencies must be assured that their data will be protected, particularly when working with law enforcement data or data containing personal information. Researchers can allay some concerns about data sharing by coming to the table prepared with a plan to protect data security and preserve each agency's ownership of their data. For Metropolitan Crime Mapping, Urban created a secure file transfer protocol portal that allowed agencies to send data files over a secure connection. Internally, data were stored in a secure, confidential server only accessible by project staff.

Partner engagement is critical, particularly when working with law enforcement agencies, and there are several ways to ensure engagement is ongoing and interactive. First, the initiating researchers or agency should become familiar with the primary point of contact at each agency and with the hierarchy of people who ought to be involved in communications. For example, when working with a police department, the chief will likely be copied on most communications even if the mechanics of data sharing are primarily handled by a civilian analyst (as is often the case in larger departments) or sworn officer. Second, it may be prudent to solidify relationships up front through a formal agreement or memorandum of understanding to ensure that expectations, including what data will be shared, the influence each agency will have over any final research products, and other key aspects of the relationship, are mutually understood and agreed on. Accommodating certain partner requirements will likely mean adjusting project timelines. For example, if a partner wishes to review any final research products, then a reasonable amount of time must be allotted for that review. One of Urban's key strategies for managing relationships consistently was for a single senior researcher to take the lead on engaging all partners.

Creating Data Management Structures

After securing access to data, the next step is to prepare for data collection and the cleaning, coding, and organizing of data obtained from multiple agencies. This process can be particularly onerous given the likelihood that data from each agency will be formatted differently. On the agency's end, data will typically be prepared by an analyst or officer, who may do this work as part of their job but will more often be taking on this responsibility in addition to their everyday duties. This task may be relatively simple in agencies with well-organized databases and an infrastructure that supports quick data downloads, as is often seen in major police departments. However, data collection will typically be more complex and time consuming, as data may need to be initially compiled from different locations within a single agency. Thus, it is important, particularly when requesting data as an external researcher, to be cognizant of the burdens imposed by each data pull. Researchers should allow a reasonable amount of time for data to be collected and recognize that agencies will likely be unable to clean their own data because of resource constraints. In addition, a researcher with access to all relevant datasets is better positioned than the agency itself to clean and reorganize data in a manner consistent with the project's integration and analysis plans. This may not be true, however, for ongoing data-sharing partnerships. Given the intensive and time-consuming nature of cleaning data before integration, longer-term data-sharing networks can operate more effectively by investing time up front to establish clear protocols for data labelling, organization, and so on.

One of the most important elements of a data management plan is a data dictionary that explains specific data labels and describes the contents, format, and structure of each dataset. A data dictionary can also help define the bounds of the data to be included in analyses, such as the years and geographic areas covered. For the Metropolitan Crime Mapping project, which focused primarily on geospatial data integration through GIS mapping, it was particularly relevant to clarify the level of geography of each dataset. For example, data on reported crimes and calls for police service were coded with either a street address or x and y coordinates, providing slightly different but similarly specific points of reference with which to map data. Given that researchers are very likely to encounter datasets that use different geographic units of measure, a data dictionary that explains these differences can greatly improve efficiency. Finally, a data dictionary is important for onboarding new staff to the project, thereby preserving institutional memory and project sustainability, and for ensuring that all team members maintain awareness about the data they have access to.

In many cases, datasets that are available only for higher levels of geography, such as census tracts, are provided in this manner because of policy-related restrictions rather than availability. Agencies often place restrictions on the use of their data, and each partner in a data-sharing arrangement should become familiar with those restrictions. In some cases, restrictions may lead agencies to deny all requests to share a certain type of data altogether or require that the data be stripped of several details or potential identifiers. In other cases, agencies may release data at a certain level of specificity but require that any published research products protect confidentiality by reporting only aggregate information or by excluding a specific dataset altogether. Agencies with strong institutional restrictions on data use will likely need to routinely review the research to ensure that their data are used appropriately and in a way that protects the privacy of their clients.

When developing a data dictionary, researchers or agencies collecting data should recognize that data definitions often vary across agencies or jurisdictions and plan accordingly. Crime labels are a notable example: when examining data from the Metropolitan Police Department and the Prince George's County Police Department, researchers discovered that MPD had several unique and nuanced subcategories of larceny (e.g., larceny I and larceny II, theft of auto part I and theft of auto part II, bike theft I and bike theft II, etc.). These designations were not used by other jurisdictions and thus required additional understanding and analysis to integrate. MPD also had a subcategory of crime called snatch/pickpocket, a somewhat unique label under the broader category of robbery. Researchers noted that some jurisdictions may have different interpretations of the physical contact involved in a snatch/pickpocket and whether it should be classified as robbery or larceny. Agencies and researchers must understand that different police departments, supervision agencies, and so on do not always define terms in the same way, and data definitions should be discussed with each partner to ensure compatibility.

On occasion, data will arrive in such poor condition that the cost of cleaning and preparing that dataset will be extremely high. In these situations, the research team may need to assess whether the value of the dataset merits the investment needed to clean it. Urban researchers encountered such a challenge when one dataset was delivered as thousands of PDF files that could not be readily converted or scraped to produce usable data. Harvesting information from these files and transferring it to a more usable format would have required an immense amount of time-consuming labor or a software program specifically designed to read and analyze PDF files. The team decided that the potential value of the information did not justify the immense amount of time and effort required to extract the data or the money needed to acquire the software; similar situations should be considered on a case-by-case basis.

Integrating Data

Collecting all the targeted data together in one place is a significant achievement and major step toward data integration, but several steps remain before analysts can proceed with integrative analyses. Data sources must be cleaned and made compatible with one another across a common mode of integration. Two common methods employed by the Metropolitan Crime Mapping project were individual-level (case-based) integration and location-based integration. Regardless of what method a project uses, data integration begins by navigating issues of compatibility.

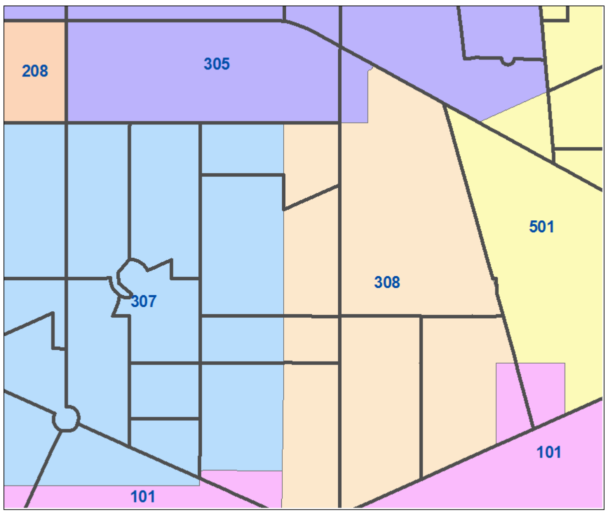
Compatibility is a key prerequisite to data integration and can have different meanings for different types of integration. For data integration that takes place through ad hoc, manual means, ensuring compatibility is a matter of aligning definitions, units, and formatting so that different data sources can be compared via a common unit or axis point, such as a geographic point or a single person. For data integration that occurs through an automated, computer-based network, compatibility also encompasses variations in programming language, back-end data system infrastructure, software used to protect or store data, and so on.[28] In both cases, the goal is to ensure that different data sources "speak the same language" (in terms of definitions, units, etc.) so that integration has the potential to reveal legitimately meaningful relationships.

For the Metropolitan Crime Mapping project, the team exploration of the relationships between community supervision and recidivism used individual-level integration to merge data from case files because the outcome in question, recidivism, was individually based. The unit of analysis was the person on supervision, and the team needed a unique identifier for each client across all datasets to merge data. This arduous process involved merging over 40 community supervision data tables with data on individual supervisees. Some tables employed a common unique identifier to refer to each person, and these tables were easy to link together. However, many tables did not use such reliable identifiers and instead had to be matched to each person through complicated cross-referencing of multiple tables and charts.

In contrast, integrating information geographically—simply merging data based on a common geographic reference, such as a census block—was relatively efficient in most cases. But some data sources used disparate geographic units of analysis, and researchers needed to impute estimates from the original data for them to align with the geographic unit of interest. The Urban team sought to isolate relationships between economic development and crime by controlling for various neighborhood factors. However, business establishment data were stored at the block group level from 2000–10; most census data during this time were aggregated to the census tract level, a much larger geographic unit. To provide an estimate suitable for use as a control, researchers used linear interpolation to generate demographic estimates that corresponded to the smaller block group unit. This approach may not be appropriate for a primary variable of interest. In another example, researchers wished to integrate census data at the census tract level with another dataset that presented data by police service area (PSA); the two geographic units differed in size and their borders did not align in any way (see figure 1 below). This presented unique challenges. Unlike integrating, for example, state-level data with county-level data, where county data can be aggregated up to provide state data, completely incompatible geographies cannot simply be combined. In this case, a more complex strategy was needed: weighting the census block group data based on the proportion of shared area across PSAs and census block groups (see appendix A, Spatial Data Integration).

FIGURE 1

Example of Mismatch between Census Tract and Police Service Area Boundaries



Note: Census block boundaries (black lines) are superimposed over police service areas (colored blocks). Integrating geographic units presents situations where, for example, a police service area contains multiple whole and partial census blocks or where a census block crosses multiple police service areas.

Perhaps most significantly, integrating data geographically demonstrated the immense value of GIS not only as a visualization tool but as a highly effective platform for data integration. Most GIS software can map x and y coordinates and addresses without additional labor on the part of the analyst, greatly simplifying the cleaning and reformatting of data. In addition, the visual nature of GIS as a data integration tool helps make the results of analyses more intuitive to researchers and practitioners. Whereas a table of numbers is fundamentally abstract and requires a significant level of cognitive processing to understand, the human mind can more quickly understand and make sense of visual relationships, such as position, shape, size, distance, and so on (Few 2013). These visual relationships show patterns that can then be analyzed quantitatively through rigorous statistical analysis and yield meaningful insights in response to research questions.

Lessons Learned

There is undoubtedly a great wealth of knowledge to be gained from integrating and analyzing cross-agency data, and the results of such integration can highlight useful patterns and relationships between crime and community dynamics. But there are also a number of lessons to learn from the process. This chapter discusses the strategies essential to any data integration partnership: developing thoughtful relationships with data providers and maintaining active communication. Researchers can benefit greatly from a forward-thinking research plan that employs effective strategies such as developing questions that are likely to benefit practitioners, creating a well-developed structure for managing data, anticipating and mitigating data-related challenges, and maintaining clear communication pipelines with partners to ask questions as they arise. Though this report describes a one-time data-sharing endeavor led by researchers, it is worth noting that data integration will be most sustainable and useful if it is built into the daily functions of the agencies collecting the data.

Chapter 3. Analytic Approaches to Integrated Data Exploration

Technological advances have made it easier and cheaper to collect, merge, manipulate, visualize, and analyze data. As a result, a historic volume of data is available to end users through public and commercial platforms. National and local datasets also include data at a much finer geographical resolution, such as the neighborhood, block group, or address level. With the growing adoption of data-processing and sharing platforms by local government agencies and the increased communication and partnership among agencies, a variety of end users can now make use of these data sources to overcome organizational silos and collaboratively address issues in their jurisdictions. The analytic approaches applicable to this larger and finer supply of data extend well beyond standard exploration and mapping techniques.

This chapter explores an array of methods for analyzing interagency and cross-jurisdictional data at many geographic levels. It begins with an overview of the purpose of these analytic tools, followed by a description of key methodologies and tools that are broadly applicable for analysts in criminal justice and public policy. The chapter provides a brief overview of how each method was used in the Metropolitan Crime Mapping project and explains when it is most appropriate to use each method.

Background

The choice of analytic methods and approaches is guided primarily by the research inquiry. Given that data needed to answer complex criminal justice questions are frequently distributed across multiple agencies and geographies, the Metropolitan Crime Mapping project sought to explore the benefits and challenges associated with data integration. In doing so, the project developed and answered a broad range of research questions that required interagency and cross-jurisdictional data analysis. Research staff employed two methods of analysis—exploratory and customized—each with a different set of uses. Exploratory methods allow researchers to prepare, describe, and explore data to facilitate quick explorative analyses and are some of the most commonly used methods among researchers and practitioners. Customized methods allow researchers to tailor analytic parameters to address specific questions or deal with confounding factors. Though less regularly used than exploratory methods, these methods can save time and effort for researchers who frequently conduct a very specific type of analysis or who frequently work with different types and sources of data that require different analytic approaches. Metropolitan Crime Mapping used both methods, sometimes alone and sometimes in combination, to pursue its research inquiries. This approach allowed researchers to describe, classify, explain, visualize, and predict crime patterns and trends in the Washington, DC, metropolitan area using cross-jurisdictional and interagency data. The following section describes a key set of analytic approaches and tools, lays out their characteristics, and recommends when to use them.

Statistical and Geospatial Analysis Designs and Methods

Cross-Sectional (One Point in Time) and Panel (Longitudinal) Designs and Data Analysis

Cross-sectional studies are one of the most common research designs and are called "cross-sectional" because information on the unit of analysis is collected at only one point in time. This design is ideal when a researcher would like to describe the relationship between an outcome (e.g., local economic health) and a factor (e.g., gunfire). Though useful for describing associations, this design is not well suited for establishing a causal relationship between an outcome and a factor because observations pertain to only one point in time and an order of events cannot be determined. However, cross-sectional studies are ideal for describing the prevalence of a problem and quickly exploring the relationship between various factors.

Panel data refers to information on the same units of analysis collected at multiple points in time. Compared to cross-sectional models, panel designs provide more opportunity to explore the temporal order of outcomes. Fixed-effects panel designs are particularly useful when some variables cannot be included in the analysis and do not vary over time. Fixed-effects models allow researchers to control for unobserved heterogeneity in statistical models when this heterogeneity is constant over time and is known to correlate with independent variables, which may include the geography of an area or other unobserved and static characteristics of the neighborhood.

Panel study designs are ideal for projects with sufficient data and resources to observe the relationship between indicators over a time series and control for the time-invariant effects of exogenous factors. However, data collection and processing are more costly and labor intensive than in cross-sectional models.

Cross-sectional ordinary least squares regression and fixed-effects panel models were used in the Metropolitan Crime Mapping project to examine the impact of gunfire on the health of local businesses. Though it is widely assumed that crime affects local economies, parsing out specific effects is difficult because of the lack of neighborhood-level economic data. Metropolitan Crime Mapping sought to close this gap with ordinary least squares models that explored the relationships between economic growth, gunfire, and crime. The regression models drew on official crime data, data from GDT systems, and a previously underutilized source of economic growth data at the microgeographic level: the NETS Database.[29] The joint utility of official crime data, ShotSpotter data, and NETS indicators allowed researchers to explore the relationship between gun violence and business health at the establishment and block level through cross-sectional and longitudinal models while ordinary least squares and fixed-effects panel models allowed them to control for confounding effects.

Negative Binomial Regression

Negative binomial regression (NBR) is an ideal tool for exploring the spatial influence of criminogenic features when the outcome of interest exhibits a random but clumped distribution. As part of NBR, analysts can use descriptive analysis tools such as the Kolmogorov–Smirnov test and skewness test, commonly found in statistical software packages, or a spatial autocorrelation test, commonly found in GIS software packages, to observe patterns. In such analyses, NBR performs significantly better than other regression models in fitting clustered observations.

These attributes made NBR appropriate for the analysis of bike thefts and violent crimes at Washington Metropolitan Transit Authority stations. To model the criminogenic influence of transit stations as transportation nodes and their place function within the broader neighborhood and city context, NBR was performed using sources such as American Community Survey data, WMATA data, Walkscore.com data, the NETS Database, and crime data from the WMATA Metro Transit Police. NBR allowed the research team to use interagency and cross-jurisdictional data to explore station-, neighborhood-, and city-based and time-variant characteristics that make certain transit stations prone to attracting different crimes.

Geovisualization

Geovisualization is the process of visually depicting geographic information. It is a very common method of analysis and provides a relatively inexpensive but powerful means of communicating complex spatial information in a simple format for a wider audience. These data visualizations can help to explore and make inferences about potential spatial relationships, among other functions. For Metropolitan Crime Mapping's exploration of bike thefts and violent crime at transit stations, geovisualization was instrumental in depicting the spatiotemporal crime patterns and trends at WMATA stations in the study area.

Multivariate Logistic Regression Analysis

Regression models are generally appropriate for assessing continuous variables. However, multiple logistic regression is a more robust method for modeling binary (yes/no) outcome variables, such as whether a person has recidivated. With a logarithmic transformation of a binary outcome variable, logistic regression allows analysts to model the nonlinear relationship between an outcome event and multiple independent variables in a linear fashion. Logistic regression is best suited for research questions on the relationship between different risk factors and the probability of a certain outcome event.

Logistic regression analysis allowed the Metropolitan Crime Mapping team to explore the relationship between recidivism and individual mobility, a person's housing situation at the beginning of their supervision term, and the characteristics of their community. The evidence on the relationship between residential mobility and recidivism is mixed, so the Metropolitan Crime Mapping project examined how moving from one housing situation or neighborhood to another affected recidivism. Specifically, a series of multivariate logistic regressions were conducted to estimate the impact of individual residential mobility, housing type, and neighborhood characteristics on recidivism. Additionally, a subsample of analysis allowed the research team to examine whether moving between different housing situations and the resulting changes in community characteristics affected recidivism.

Geographically Weighted Regression

Brunsdon, Fotheringham, and Charlton (1996) developed geographically weighted regression (GWR) in acknowledgment of Tobler's (1970) first law of geography: "Everything is related to everything else, but near things are more related than distant things." This concept, known as spatial autocorrelation, led researchers to recognize that traditional regression models cannot capture the relationships between some variables and outcomes without acknowledging that nearby phenomenon are often highly correlated. GWR addresses this problem by estimating regression coefficients at each geographic data point. In traditional multiple regression models, regression coefficients for one variable are fixed over the course of the study.

GWR is essentially a traditional regression framework that takes into consideration local spatial relationships. It is used in geospatial modeling to explore "spatial nonstationarity" (Thapa and Estoque 2012, 85). When diagnostic statistics, such as a Global Moran's I, reveal high levels of spatial autocorrelation, GWR can allow researchers to estimate the local parameters of relationships and not just the global parameters (as with traditional multiple linear regression modeling). Thus, GWR recognizes that nearby units of analysis tend to be more similar than units farther away and weights each unit differently based on its location relative to a target unit. GWR is an ideal analysis tool for exploring the nonstationary relationships between variables over a common geography.

The Metropolitan Crime Mapping project used GWR to explore the validity and reliability of gunfire detection technology as a source of information about gun violence. Specifically, the team explored the relationship between GDT activations, local patterns of reported crime data, and citizen calls. GWR findings allowed researchers to assess how traditional measures of gun violence, such as reported gunshots and gun assaults, related to gunshots detected through GDT systems. GWR also provided localized estimates that allowed researchers to explore how the relationship between these variables changed at different points across the city, which a broader tool would not have detected.

Risk Terrain Modeling

Risk terrain modeling (RTM) was developed by Caplan and Kennedy (2010) based on the understanding that criminal outcomes are both event and context dependent. According to RTM, the interaction of several criminogenic factors at the microgeographic level can be studied to reveal consistent patterns of interactions that facilitate crime (Piza, Kennedy, and Caplan 2011). Computing the conditions that underlie these patterns is a key component of RTM, which has the ability to weight the criminogenic spatial influence of different factors.

Crimes occur in places where the presence of people motivated to commit crime intersects with ready targets for criminal opportunity. Therefore, exploring crime opportunities and patterns in an area requires considering the impact of those same opportunities and patterns in adjacent areas. These influences can extend beyond the boundaries of a county, city, or state. Such a variety and volume of data require geospatial analysis tools and processes like RTM that can be customized to process and analyze data in the most efficient way.

The Risk Terrain Modeling Diagnostics (RTMDx) utility automates the steps of RTM: operationalizing the spatial influence of risk map layers, selecting and validating the risk map layers with existing outcome data, weighting the risk map layers in relation to one another, and testing the predictive validity of the resulting risk terrain model. The utility allowed Metropolitan Crime Mapping researchers to select only the most appropriate risk factors for robberies and aggravated assaults, with the optimally operationalized criminogenic spatial influences of these risk factors (see appendix B for detailed steps). The end result is the "best" risk terrain model, created with a combination of powerful statistical methods, namely cross-validation, a custom elastic net model of penalized Poisson regression, and a custom bidirectional stepwise regression with Bayesian information criterion scores. Generally, such an analysis requires substantial programming skills, but tools such as RTMDx allow analysts to easily deploy sophisticated statistical methods to conduct crime risk prediction.

Markov Transition Matrix Modeling

To assess the impact of including Washington, DC, data and both MPD and Prince George's County Police Department data in crime forecasts, the Urban research team developed a Markov transition matrix model that employed a multistage model that incorporates information iteratively, which allowed the team to include data from different periods of time. This was particularly relevant for this project because the crime and gunshot data included in the analysis were all collected over different (but overlapping) time periods: crimes reported to the MPD between 2000 and 2013, activations of the MPD's GDT system from 2010 to 2013, and crimes reported to the Prince George's County Police Department from 2010 to 2013.

Conclusion

Public safety researchers and practitioners have made great progress collecting, integrating, and analyzing data to learn about the relationship between crime and place. The Metropolitan Crime Mapping project was essential to these efforts and provided several examples of important methods and tools for projects looking to analyze data from different agencies and jurisdictions.

Though several statistical techniques can be used for cross-jurisdictional data analysis, the tools highlighted by the Metropolitan Crime Mapping project may help analysts increase the quality and effectiveness of their analyses without substantially increasing the resources required. Such tools may also allow analysts to better integrate new data into their analyses: Metropolitan Crime Mapping used a wide array of methods to integrate new neighborhood-level data sources, including the NETS Database, the American Community Survey, and ShotSpotter data, alongside traditional data sources. These methods will continue to grow more effective, and it will be up to the developers of those methods and the analysts who employ the tools to engage in partnerships that maximize the impact of new developments in the field.

The increasing use of GPS-enabled devices could benefit efforts to collect and analyze granular, real-time data about communities, including data on crimes and disorder. Future study designs should consider how such geographic information can be used to analyze crime patterns. Further, given the growing number of data- and information-sharing initiatives, now might be an opportune time to invest in customizable tools and utilities that make it easier for researchers to merge, analyze, share, and report on an increasing volume of available cross-jurisdictional and interagency data. The future of integrated data exploration will require partnerships between academics, researchers, and practitioners in which each stakeholder effectively communicates the needs (e.g., data and infrastructure) and the developments (e.g. new analytic methods) in one field with stakeholders from other fields.

Chapter 4. The Future of Interagency Data Integration

The main content of this blueprint thus far has been a detailed discussion of cross-silo data integration and analysis, where researchers or agencies request data that are then shared physically or digitally through active management. This form of data sharing is the most common among analysts and researchers. But recent advances in information technology, as well as in the accessibility of and culture surrounding data, could significantly reshape the landscape of data integration. Many of these developments, from cloud-based file sharing (e.g., popular services such as Dropbox, Google Drive, and OneDrive) to smartphone technology and all of its attendant applications, are now pervasive in daily life. Most people today share personal information in large quantities, both intentionally (e.g., through social media) and unintentionally (e.g., through consumer and other transactional data). But in the context of policy-oriented cross-silo data integration, research and practice have remained relatively static, particularly at the local and regional level. The technologies that facilitate data access and dissemination in daily life among private citizens have yet to permeate interagency data integration efforts in meaningful or widespread ways despite the advantages they may offer. At the same time, digitizing existing information systems and data-sharing networks has the potential to make these systems both easier to use and lower in cost in the long term.

The final chapter of this blueprint takes a prospective look at how to bridge this gap. It harnesses the knowledge of experts to translate and synthesize what is technologically possible for local interagency data sharing and integration in 2017 into language that is directly relevant to practitioners. The chapter begins by describing recent changes in the role of interagency data work, including the technologies and major events that have helped foster a shift toward greater data sharing and integration. It then delves into key developments, such as cloud computing, data portals, smartphones, and version control, with major implications for how data are shared and used between agencies and jurisdictions. The latter part of the chapter focuses on changes in practice and culture, such as the open data movement and easier access to expertise, that have created opportunities and momentum for greater data sharing and integration. Finally, it concludes by noting the tension that agencies and researchers will encounter around issues of privacy and civil rights and explores how to mitigate these concerns by engaging the public as a partner in data integration efforts.

This chapter was informed by interviews with several IT and data experts who were asked to provide their insights on what the future of data sharing might hold given the range of new technologies and tools available today. It shifts focus from researcher-driven data integration to discuss a more fully integrated data-sharing network between agencies. This shift is deliberate, as data sharing will inform practice most effectively if agencies collect and share data independently while researchers and universities continue to provide valuable resources for development and analysis expertise. Only when data integration becomes part of an agency's routine operations will such a project be sustainable and valuable across a wide array of criminal justice and social services.

The Shifting Role of Data Sharing

In the 21st century, major crises such as the September 11 terrorist attacks and Hurricane Katrina floods led to a change in the role of data sharing, particularly its role in effective crisis response. Intense scrutiny of government actions in the wake of these incidents highlighted gaps in information and communication that several experts have asserted could have saved many lives—through prevention in the former case and effective emergency response in the latter (Banipal 2006; Comfort and Kapucu 2006; Lee and Rao 2007; Popp et al. 2004). These disasters taught a hard lesson: that many of the problems and solutions that arise in such crises are complex and cross sectoral and require the cooperation of many agencies across multiple jurisdictions.

Of course, it is easy to speculate that data sharing would have prevented or improved a situation, but such arguments remain just that—speculation. What is clear is that if data are to be useful in a crisis situation—or even in response to more minor but time-sensitive problems—pipelines capable of rapidly accessing data must be established. Policymakers and leaders simply do not have time in a crisis situation to begin developing data-sharing partnerships. Thus, though the majority of this blueprint has described a one-time, researcher-driven data integration effort, agencies would be best served to move toward ongoing partnerships that allow them to acquire data rapidly through strong lines of communication, interpersonal relationships and memorandums of understanding with other agencies, or virtual networks that multiple users can use to securely access, enter, and analyze data through a central portal. This is particularly relevant to law enforcement, which frequently must respond to smaller-scale crises of crime and violence that still require rapid response. Maintaining a strong data-sharing network also allows analysts to provide rapid, on-demand answers to decisionmakers' questions, permitting them to create policy that is informed and best serves the public interest. More frequent analysis of different sources of data may also lead agencies to identify emerging issues and make midcourse corrections where necessary.

Trends in the Technology of Data Sharing

Before discussing specific advances in technology that might affect interagency data integration in the near future, it is helpful to paint a picture of what the future might look like.

Officer Ramirez is out on patrol when she receives a call about a disturbance in a nearby neighborhood. Arriving at the given address, she finds a man in the midst of a mental health crisis. From speaking to the man and a few neighbors standing nearby, she learns his name and quickly types it into an application on her smartphone. Her query is run through several databases that scan information from local police departments, social services agencies, hospital emergency rooms, and other sources. She quickly receives a list of results and finds that the man has been associated with several similar incidents and was referred previously to a local mental health care organization with dual diagnoses of severe mental illness and substance dependency. Officer Ramirez calls the organization, which promptly dispatches staff to the scene and helps her deescalate the situation. Officer Ramirez later uses the same app to log key details of the situation for an incident report she will complete back at the department. As she types, the information is transmitted back to the police department database in real time through a secure wireless connection. Some of this information is immediately uploaded into the same local data hub that Officer Ramirez was able to search for crisis response information moments before. Partner agencies can access and update data simultaneously and track changes made by others. The hub's welcome screen offers easy ways for staff members to search information, access frequently used pages, and even run basic analyses or visualize data through an interface without needing advanced statistical training. At the same time, analysts with those skills can download data from the hub in a variety of formats that allow for more advanced analyses.

The data system described above may seem futuristic, but the technology for every aspect of it already exists. The challenge now is not inventing new technologies but identifying and adapting existing ones, combining these tools into an infrastructure that facilitates data sharing while maintaining data security. The following section will break down major advancements into key technological components and outline the implications each has for data sharing.

Cloud Computing as a Launchpad for Data Sharing

One of the new technologies most critical to data sharing is the ability to store and exchange information virtually in the cloud. Cloud computing is an internet-based form of computing that shares resources (e.g., storage and information) among a pool of devices that can access these resources on demand. The cloud provides a vehicle by which criminal justice and other agencies might share and store data through a central hub, accessible from virtually anywhere with internet access at any time, at a very low cost. This mobility has enormous implications for on-the-ground staff of data-sharing organizations. For example, the cloud could allow police officers on patrol to easily enter or access information through a mobile device and similarly allow probation and parole officers to record key information about people on supervision or situations in the field, minimizing paperwork and the number of times they must return to a central office to report information. These gains in efficiency have been estimated to yield between 25 and 50 percent in cost savings in computing operations at the state level (West 2010).[30]

Despite growing awareness of the cost savings offered by cloud-based systems, concerns about data security and privacy are a key barrier preventing adoption of cloud technology (Pearson and Benameur 2010; Popović and Hocenski 2010). These concerns are stoked, one might plausibly assume, by the frequent appearance of hacking and cybersecurity breaches in the media.[31] However, several cloud-specific security solutions exist, and cloud storage systems are often more secure than conventional hardware storage systems. Additionally, cloud storage protects key information from hardware failures that risk destroying all information on a given device. The combined appeal of cost-effectiveness and improved efficiency and security has led several federal agencies to begin using the cloud to securely store and exchange information within their own agency and with other agencies. FedRAMP, formally established in December 2011, is "a government-wide program that provides a standardized approach to security assessment, authorization, and continuous monitoring for cloud products and services" (FedRAMP 2014).[32] FedRAMP's array of physical, digital, and administrative data safeguards provide a standard model for cloud-based data security that might be replicated to some extent in other areas.[33] Cloud storage has become increasingly popular among federal agencies but has yet to be widely adopted by local and regional agencies, with some notable exceptions. For example, SF OpenData is a cloud-based clearinghouse for data published by the city and county of San Francisco.[34]

Data Portals and Mashups

Data portals and mashups are other ways to bring data together and present it in a uniform, user-friendly, and usually interactive way. Both offer ways to aggregate data sources into a single interface, but they differ slightly in the underlying technology and in how data are presented. Data portals typically display different types of information side by side, each through its own "portlet," so they are comparable but separate (e.g., a data dashboard). A data portal can thus provide a useful interface through which multiple agencies can synthesize shared information. Several existing data portals may provide a useful resource for agencies and researchers seeking to explore or integrate different data sources. These data portals appear in a variety of contexts, from advocate-driven info hubs to data dashboards. Some are interactive and allow users to conduct simple analyses with data from different sources using a single tool (see the American FactFinder and the Google Public Data Explorer).

Mashups also aggregate data from multiple sources, but they differ from data portals in that they integrate multiple data sources "from wherever they are stored and whatever format they are stored in," typically into a hybrid graphical display that addresses a particular need or question (e.g., integrating data on available real estate directly onto a digital map) (Sanders 2014). Data portals and mashups are sometimes compared to a salad bowl and a melting pot, respectively. Data portals, the salad bowl, present data side by side. Mashups, the melting pot, blend the presentation.[35] Both tools are customizable and allow users to define what information they want to display. Creators of mashups often make the interface even more interactive for other programmers by making the code behind the original data source or mashup accessible through open application programming interfaces. To greatly oversimplify a metaphor, this is similar to the packaging of a commercial food providing not only a list of ingredients but also the (typically proprietary and protected) full details of how the product is made for the benefit of culinary innovators who might combine that formula with another and improve on and innovate around that recipe. Platform creators give up some of the exclusiveness of their creation by making it easier to imitate, but they facilitate a great degree of innovation and gain critical feedback on their own product. Open application programming interfaces can help accelerate and reduce the costs of creating a mashup because they reduce up-front labor on the part of the developer, who can draw on existing code instead of building a data integration tool from scratch (Floyd et al. 2007; Merrill et al. 2002; Sanders 2014).

Use of a data portal or mashup does not preclude use of the other system, and in some cases the two systems may actually complement each other. For example, a mashup application may be used as a widget in a data portal to visually integrate data in that portal. Drawing on Urban's Metropolitan Crime Mapping project, a specific illustration of what this might look like would be if all data sources from the Washington, DC, transit authority, police department, community supervision agency, and so on were (1) drawn together in a single portal, where they could be viewed side by side, that also (2) included a mashup application that enabled users to easily integrate data over a map of Washington, DC.

Big Data and the Internet of Things

Big data processing has made available several new data sources that could, in theory, be integrated with local agency data to yield important insights into crime and public safety. Essentially, big data processing is just that: computers processing or extracting useful information from extremely large datasets. In many cases, this involves analyzing unstructured data or data that may contain important patterns but are not organized in a standardized format, such as a spreadsheet, that is amenable to analysis. Examples of unstructured data include books, video and audio recordings, social media feeds, and so on.

There are several ways that big data could be used to better understand local public safety dynamics. One example is the storage and use of police body-worn camera footage. As more police departments outfit their officers with body-worn cameras as a proactive and visible measure to promote officer accountability, those officers are collecting a tremendous volume of data. It would be enormously resource intensive—and virtually impossible—to systematically analyze this much footage through traditional means (i.e., human analysts watching and manually coding or otherwise analyzing video content). Video content analysis technologies can automatically analyze footage, but they remain limited in what information they can detect, the quality of video required for effective analysis, and the speed at which they perform analysis (Chen, Mao, and Liu 2014; Hampapur et al. 2009). Thus, limited availability of suitable tools remains a major obstacle to analysis of body-worn camera footage despite the potential of these relatively objective records of police-citizen interactions to help hold departments accountable to their communities and communities accountable to their police departments. But big data analysis methods are developing rapidly and could resolve this challenge in the near future.

Another potential source of data and data integration possibilities is the wide range of devices and objects that connect to wireless networks and emit signals, sometimes called "the Internet of things." Cell phones and computers are obvious examples, but the concept more directly applies to ubiquitously "wired" systems of objects brought online to achieve some purpose: trash cans that contact city services or a central receiving computer system to alert you when they are full, buildings that improve energy efficiency by knowing when the air conditioning is running while windows are open, and fire hydrants that alert the city when they are broken. New York University's Center for Urban Science and Progress is behind one of the most extensive such efforts and has partnered with the city to launch the first "quantified community" in a new Manhattan development called Hudson Yards. When completed, the 28-acre complex will include thousands of sensors that collect information on air quality, pedestrian traffic, energy production and consumption, and other measures and return the information to secure servers. This information can then be analyzed to help improve city services, with projected savings of $20 billion by 2020.[36]

The quantified city project is just one example of a larger move to create “smart cities” fueled in part by a 2015 White House initiative that will invest over $160 million in research to address intersectional city challenges related to crime, traffic congestion, local economy, climate change, and so on.[37] A key aspect of this initiative is collecting and sharing information among local agencies and universities through a web of partnerships called the MetroLab Network.[38] Though MetroLab itself does not focus on crime and public safety, it is easy to envision a similar network that extends to detect maintenance issues, such as broken streetlights, widely believed to influence crime and perceptions of public safety (Pain et al. 2006; Skogan 2012).

A different but more directly crime-related example of an automated, sensor-based data source is gunfire detection technology. GDT systems such as ShotSpotter use a network of optical and acoustic sensors to detect gunfire and transmit the incident data to a central processing center for rapid analysis and response by police departments. Cell phones and other personal devices also generate data automatically (e.g., on the volume of communication and geographic data such as preferred routes or destinations) in addition to information deliberately sent by the owner.

However, government collection of this data has understandably been met with a considerable amount of controversy and criticism from civil rights and privacy advocates, most notably in the case of the National Security Agency’s surveillance of personal phone data. Several local police departments have also been subject to criticism and lawsuits for accessing such information—without a warrant—through “stingray” cell phone tracking devices. Automatically generated cell phone data are certainly freely available, but agencies must weigh the tangible benefits gained from acquiring these data against the considerable and legitimate public concern over the threat collecting these data poses to civil rights and personal privacy.

Smartphones as Tools for Data Collection and Usage

In general, smartphones have positive potential as a means of collecting and accessing data. Smartphones provide a vehicle by which practitioners on the move can both enter new information and retrieve stored information that may inform actions in the field. Many cloud computing platforms, which provide the necessary data storage space and infrastructure, also offer easy ways for clients to present data in a mobile interface. Smartphones can also operate as data collection tools, as many smartphones now come with a range of capabilities, from geolocation to cameras, that can enhance the detail and quality of data entered into a shared data system. Patrol officers could, for example, use phones to mark the location of a crime, and people on probation or parole could use their phones to check in with their supervising officers remotely, reducing the frequency of in-person visits where appropriate.

Mobile devices also offer an important opportunity for bidirectional communication with the public. Law enforcement and other agencies can use social media or dedicated applications to inform citizens of key updates, but smartphones are also a convenient and relatively anonymous way for citizens to contact local agencies about concerns or issues (e.g., through a "tip line" app or text service). Many such apps have been developed to enable citizens to exchange information with local government agencies about local issues. For example, the BlueLight 911 app facilitates 911 calls through an app that lets users share their precise location, picture, and medical condition to 911 dispatch agents.[40] Other apps allow citizens to report a wide variety of nonemergency issues, such as potholes and graffiti (Kingsley, Coulton, and Pettit 2014).

When used as tools for community engagement, smartphones have implications for the accessibility of information and the perceived accessibility of agencies themselves. Many users rely on smartphones as their sole or only stable means of communication, often forgoing other devices such as landline phones, desktop or laptop computers, and so on (Kingsley, Coulton, and Pettit 2014). The accessibility of smartphones also makes them a compelling tool for both one-time survey efforts as well as ongoing community-wide data collection on specific topics. For example, smartphones can be used to collect data on citizen perceptions of policing at a lower cost than traditional surveys while reducing many of the barriers that might prevent citizens from reporting negative perceptions of the police through conventional means, such as a fear of being identified or the lack of time or willingness to go through a formal complaint process. These data might include citizens' direct observations of and interactions with police as well as their overall sense of "police legitimacy." Data gathered in this manner and collected consistently over time could help give voice to people whose experiences might otherwise go unheard, validate citizens' perceptions as a legitimate measure of community policing success, and help police departments assess whether they are improving on this measure.

Simultaneous Access and Version Control

Today, several platforms, ranging from databases to web-editable documents, allow multiple users to access and edit data or information on a central hub simultaneously and from different locations without waiting for space to free up. The futuristic scenario described earlier with Officer Ramirez envisioned an automated, multiuser, interagency database that would allow staff at participating agencies to access and update information anywhere at any time. This can have important implications for day-to-day work—for example, a case manager working with the person who was in mental health crisis could enter case notes as Officer Ramirez enters her incident information simultaneously into the same system. But a major challenge with simultaneous, multiuser access is that for the system to work efficiently, different users must be able to tell what changes are made to the shared database, when, and by whom. Without such a system, the case manager might not realize that Officer Ramirez has entered new and pertinent information about a person they are working with, or Officer Ramirez might fail to see a correction to previously incorrect information. Or both parties might enter redundant information about the same person or situation. Traditionally, users needed to proactively communicate these changes through e-mail, phone calls, and so on, which can be an unreasonable expectation of staff members who might be overworked or in dynamic jobs that frequently require sudden shifts in activity in response to emerging situations. Version control software addresses this problem by recording changes as they are made and allowing users to recall previous versions of a shared product. The latest version control tools, including popular systems such as Git, Subversion, and Mercurial, are designed to track changes to specific lines of code in multiuser software source code projects, and similar principles have been applied to other types of files as well.

Interoperability

Though technology affords new opportunities to improve the speed, efficiency, and timeliness of interagency data sharing, it also introduces new challenges. This is particularly true when creating a computer-based data-sharing system that directly integrates agency data sources across a shared network. Many of these challenges relate to the concept of interoperability, or how well data from different agencies are able to “speak to” and integrate with one another. On one level, this is a matter of human protocol as much as technical compatibility in terms of replicability and ease of transfer (i.e., could someone new to your dataset look at it and understand the labelling and organization well enough to draw useful information from it?). On another, perhaps more challenging level, interoperability is a technical issue related to systems’ use of different programming languages and infrastructures. Similar to how human languages have evolved in different contexts, many different programming languages have been created in an effort to find the most efficient way to transmit clear instructions to a computer in different contexts. But unlike most human languages, there are not necessarily physical territories in which certain shared languages or organizational systems are used; agency A might have an office right next to agency B but use a data system built on a completely different language. This presents a major obstacle for local interagency and cross-jurisdictional data sharing, particularly because many state and local agencies rely on legacy systems that were programmed in a simple, homegrown language or a language that is no longer widely used. Programs written in different languages have historically been difficult to integrate, just as it would be difficult to share information with someone who does not speak the same language as you without undergoing the painstaking process of learning and translating the original information into their language. In both cases, a third-party translator is key. For data-sharing networks, this role is played by data mediation or transformation tools, which convert source data into a predetermined destination format. A shared database that retrieves data directly from their original source can be built to include such a mediation tool so that source data can be displayed in a common format and attempts to search or retrieve information from that source by another agency can be translated into language matching that of the original system. Without a data transformation tool that can convert data from their original source format, data may be shared as they are (perhaps through a data portal that displays different data sources separately) but are unlikely to be usable by other agencies and cannot be directly integrated or combined into a single display. Transformation also allows partners in a data-sharing network to share and retrieve information without needing to know the ins and outs of each database.

There are many slow and painstaking ways to make data interoperable, and some past efforts to digitize interagency data sharing have asked partners to enter information twice: first into an internal database and then, using a different set of organizational protocols, into the shared database. However, investing time and resources up front to develop a strong data-sharing infrastructure can (1) greatly reduce the time and effort required in the long term and (2) result in a tool that is more user-friendly and more likely to be used by practitioners. Much of this infrastructure development is highly technical work that will require the expertise of programmers who specialize in developing such databases. This presents significant challenges to agencies with tight budget constraints and competing priorities. But data-sharing infrastructures also include a broad array of human elements, from culture to policy, whose dynamics have critical implications for how data are shared.

Trends in the Culture and Practice of Data Sharing

At the crux of many recent changes in data-sharing technology and culture is the idea of open data. Open Knowledge International defines open data as "data that can be freely used, re-used and redistributed by anyone."[41] Because of the immense amount of data collected by governments, the open data movement has been accompanied by a push for open government, both because of the immense amount of data collected by governments and for more philosophical reasons, including the idea described by Kinsley, Coulton, and Pettit (2014) that "technology can improve the transparency of processes and information so that citizens can hold governments accountable." Other arguments for open government include the idea that government data are a public good that should be available to taxpayers, that the data can bring value to society through research, and that open government would encourage citizen engagement in governmental decisionmaking. In both public and private contexts, open data can serve as a mechanism for accountability and quality control by opening data up to greater public review of the content and of the data themselves. But this scrutiny also makes many agencies apprehensive about releasing data because of the potential public response or because of how the data might be used.

In the context of law enforcement and public safety, many agencies may feel that open data is antithetical to security and that greater public access to data will inevitably compromise information that should remain closely guarded. However, two misconceptions must be challenged here. First, open data is often only thought of as public release of information, but there is also immense value in making data more open internally, for example, by giving patrol officers greater access to databases normally accessed mostly by analysts. This can enhance internal accountability and provide a means of checking data accuracy for those most likely to detect errors. Second, open data does not necessarily involve opening all data. Data containing personal identifying information are and should remain confidential, but releasing aggregate data and patterns can help encourage transparency, enhance understanding of local crime and risk, and promote public engagement in public safety work.

The Value of Democratizing Data

In addition to its role in promoting accountability and transparency and improving relationships with the public, open data can also invite public participation in crime prevention and provide some checks on data inconsistencies or errors. Another key benefit of open data is its impact on research. On one hand, researchers can access much more data as it becomes publicly available, which through analysis they might transform into new and useful findings. This can help contribute to the cross-germination of ideas and lead to useful innovations. On the other hand, when researchers share their own data and research processes in the form of journal articles or other publications, it holds data providers more accountable for the accuracy and quality of data by expanding the number of users and inviting input and inquiries around data definitions, quality, accuracy, and so on. In many cases, this may lead researchers to refine their own understanding of previous findings or open up new lines of inquiry. Principles of open information also play a beneficial role in web and software development in the form of open-source code, which promotes universal access to a product's code to encourage future improvements, adaptations, and innovations. The availability of source code on the Internet allows many developers to draw heavily on existing code when creating a new program rather than building it entirely from scratch.

One particularly compelling benefit of open data is that it allows jurisdictions to crowdsource analyses and invite a broader range of people to examine and analyze data. Though far from replacing the need for in-house data analysis, crowdsourcing presents a unique opportunity for jurisdictions to maximize their limited resources while inviting deeper—and free—data analyses that might highlight patterns or connections that would have gone undiscovered. Take, for example, the work of blogger Ben Wellington, whose blog, *I Quant NY*, regularly combs New York City data to identify unexpected patterns and relationships. In a popular post, Wellington examined parking ticket data around fire hydrants and noticed that one particular hydrant had generated $33,000 in parking tickets in one year. Curious, Wellington visited the hydrant and realized that there was a very wide bike lane between the hydrant and the parking lane, which made it less of an obvious parking deterrent and presumably misled drivers to believe that they were parked sufficiently far from the hydrant. The city subsequently remarked the area to make it clear that parking is not allowed.[42] Outside of the generosity (or boredom) of bloggers, cities can also take a more proactive approach to accessing needed expertise through directed crowdsourcing. Possible strategies include posing questions to the public online or through a social media campaign, holding a competition or hackathon to promote focused analysis on a specific topic area, or contacting data analysis nonprofit organizations such as DataKind, Black Girls Code, Girls Who Code, or Code for America to co-identify data-related problems and areas for analysis.

Challenges of Open Data

As shifting cultural preferences and increasingly widespread Internet access make public a great deal of previously unpublished information, the open data movement is also exposing data providers and consumers to new challenges. Regularly cleaning and publishing data can be costly and time intensive, particularly for agencies that collect large volumes of data. In the case of government-provided data related to safety and security, there is an ongoing debate about what information should be open in the name of transparency and what must remain private to protect security, confidentiality, or sensitive intelligence. Other challenges previously mentioned include concerns over increased public scrutiny and the possibility that data will be used in unanticipated and potentially harmful ways.

Open data presents very different challenges for data consumers. An immense amount of data is available today through the Internet, but barriers related to accessibility, data quality, cost, and the volume of data itself reduce the utility of these data to the average citizen.

Accessibility, in the context of data sharing, tends to be narrowly defined, with "access" viewed as a matter of Internet access or, in the case of internal data, appropriate security permissions. This type of functional access to data is, of course, a prerequisite for computer-based data sharing and an area that continues to see disparity in terms of a "digital divide." Explaining this divide, Zickuhr and Smith (2012) cite a 2011 survey conducted by The Pew Charitable Trusts that examined Internet access across racial groups and found that about 70 percent of black respondents and 68 percent of Latino respondents reported using the Internet, compared with 80 percent of whites. But meaningful access to data is more complex than simply being able to see the data; access is also a function of the user's ability to understand and interpret this data. This draws attention to the larger issue of data literacy: even those who can access and use the Internet in some capacity often do not know how to make significant use of these data. Data illiteracy is an issue not only for data consumers but for staff of agencies that provide data, particularly when these agencies seek to increase data accessibility internally. Making data available to staff who are not trained analysts has its benefits, but they are unlikely to be realized if those staff do not understand how data are collected, organized, analyzed, and so on. Similar challenges arise when sharing data between agencies.

But in many cases, data can be made more accessible by presenting them in a way that does not require highly specialized knowledge to understand. One way to achieve this is by using data labels. Currently, open data portals often upload data with original variable names that are difficult to interpret or altogether unintelligible for some prospective users and may even consist of abbreviated codes for some concepts (e.g., "bike theft at metro stop X" may be given the short variable name "bt_metrox"). This may save time at the analysis end, but it makes that data extremely confusing to an outsider. That user could consult a data dictionary to help interpret this variable, but doing so would require extra effort they may not be willing to put forth. Thus, intuitive and simple data labels, typically referred to in the field as metadata, are a relatively easy way for data providers to make their data more comprehensible to the reader.

Data labelling scratches the surface of a second issue related to open data: data quality. Nearly all data are flawed in some way, but there are measures that can help ensure that the data agencies record in their databases come as close as possible to representing the reality they intend to capture. Training staff on how to properly collect and enter data can clarify expectations and ensure consistency within an agency. Because many local government agencies task virtually all staff with some role in developing or entering data, training should address staff in a variety of roles and may involve developing agency standards for certain issues, such as how much and what type of information should a probation officer enter into a person's case plan? How much detail should be included in notes about a police call for service? How should a case manager keep track of referrals to service providers? Should they also track how frequently a client attends services or treatment? Data standards also come into play when integrating data across agencies or jurisdictions, particularly when the intent is to create a system over a shared computer network that retrieves data directly from sources. Creating a common set of data definitions, formatting protocols, and organizational guidelines can smooth data integration and greatly reduce confusion, redundancy, and gaps between datasets. Where data security is a concern, standards can also ensure that data partners are uniformly aware of security measures and are taking sufficient precaution to protect the information.

Even in an admittedly ambitious ideal scenario where staff members are well trained and fully understand agency and partnership guidelines, data are managed by humans, and humans make mistakes. Routine data checks can help ensure that information is as accurate as possible and entered properly. A check might be as simple as regular "ground truthing": making a common sense comparison of the conditions portrayed by data with the observable reality on the ground (Cytron 2014). It can also be a more intensive process using analytic software that automatically checks for inconsistencies or multiple staff reviews of the same dataset. Staff members should always be aware that data are not neutral, and assumptions go into what is collected and how.

Access to Expertise

Thus far, this chapter has discussed a range of technological and cultural developments that can make data sharing faster, more efficient, more useful, and more accessible. Further, these technologies can serve as a platform for conducting informative data integration analyses quickly and on a routine basis. But as data sharing becomes a more high-tech enterprise, it creates a stronger need for appropriately trained staff. Expertise is expensive and no small investment for many departments; however, without this investment, many of the more advanced methods of data sharing will remain out of reach. Agencies have several options when deciding how to staff up in support of data sharing and integration, and our discussions with IT experts highlighted a few possibilities.

According to Urban Institute data scientist Alex Engler, creating a successful system for sharing data even semiregularly requires, at a minimum, a chief data officer.[43] The chief data officer sits at the center of a data-sharing network and ensures that it runs efficiently; his or her job is to work with stakeholder agencies and build data pipelines so that information flows easily and in a timely manner. Urban senior fellow and data visualization specialist Jonathan Schwabish offers a slightly different model, which he calls “dataflow,” that says data can most effectively inform practice when an agency (1) assembles a team of data analysts, (2) embeds this team near the decisionmakers, (3) works to change its culture, and (4) starts with modest, achievable successes that reaffirm the value of data and build confidence in the use of new technologies.[44] Schwabish adds that building a team of data analysts gives them a greater presence in the agency, facilitates collaboration, and allows agencies to hire people with specialized skills, such as a statistician who can do complex analyses, a programmer or IT expert who understands the structure of the data-sharing system, and design or data visualization experts who can present the analyses in a way that is understandable and useful to the larger agency. (See also Patil and Mason 2015.)

Changing the Culture around Data Sharing

One of the greatest and most important challenges to data sharing is agency culture. Data sharing advocates must find tangible ways to illustrate the value of data sharing to staff in every agency and at every level, emphasizing the myriad ways that data can improve performance as well as the costs of not using data. Recent federal efforts to promote data sharing have helped bring this concept into the mainstream, and President Obama’s Memorandum on Transparency and Open Government and the establishment of Data.gov sought to model this work at the national level.[45] At the local level, having a data champion, someone who believes in and will advocate for the value of data sharing, can be the key to making interagency data sharing a reality. Data visualization, storytelling, and marketing can also help translate abstract data into pictures, maps, and other compelling—and comprehensible—visuals.

Conclusion

Data integration has great potential to support crime and safety work given the constant development of advanced technological tools and the growing cultural momentum to embed data more thoroughly into the daily work of public agencies. Data integration allows jurisdictions to examine relationships between constantly changing neighborhood dynamics and crime, and the field is well on its way to creating models that predict crime more accurately and adaptively than ever before. This type of predictive public safety work opens new opportunities for agencies to act proactively and intelligently to prevent crime using strategies that anticipate crime displacement and incorporate a growing understanding of potential root causes. Strategic partnerships with agencies outside the field of criminal justice will allow police to respond more holistically to incidents by drawing on information from social services and health agencies, among others. For example, a police officer responding to a call for service could learn in real time whether members of the household have had interactions with social services agencies that might indicate a reason to proceed with special care or attention, such as the presence of a child or person with a disability.

However, cautions surrounding data integration remain. Concerns over privacy and the protection of civil rights will continue to be a major issue for data collection and use. When it comes to personal data, there is a deeply embedded and well-founded concern over data security, and agencies entering data-sharing partnerships must take great care to develop firewalls and data access protocols to protect against breaches. At the same time, local agencies can reduce tensions by engaging the public as a partner in data collection and integration. Agencies can make data and analyses more available for public use and develop pathways by which members of the public can directly contribute their information and analysis.

Appendix A. Spatial Data Integration

If data exists on different geographic levels, it would be difficult to make comparisons without some sort of standardization or adjustment. For nonstandard census geographic areas (e.g., a police service area, a radius around subway stations) in particular, it would be challenging to make meaningful interpretations of the data. There are analytic approaches to addressing this issue. One such method relies on areal summation to estimate variables of interest (e.g., demographic and socioeconomic statistics) using the same geographic unit. This appendix describes the mechanics of this methodology and its performance in different test settings.

Problem

When point data (e.g., the location of an arrest) are available, one can aggregate the data up to any level of geography required for analysis. But if data are only available at the areal unit level (e.g., census blocks, police beats), aggregation is more complicated. For example, one may be interested in census demographic information for a police service area (PSA) that may not follow the boundaries of standard census geography. Analysts are sometimes interested in understanding the demographic profile of a select geographical area around a point of interest (e.g., the number of low-income households within a half mile of a subway station) that intersects multiple census-level polygons. In such instances, analysts need to estimate variables of interest because they are not directly available at the desired level of geography.

Methodology

Figure A.1 shows an example where the boundaries of a PSA do not follow census geography. This incompatibility prohibits analysts from simply aggregating census blocks or block groups to the PSA level. The Metropolitan Crime Mapping project thus developed a method for estimating key measures at the PSA level (and other levels of geography required for analysis) by weighting the census block group–level data. The weighting was based on the proportion of areas shared by two geographic layers and follows these steps:

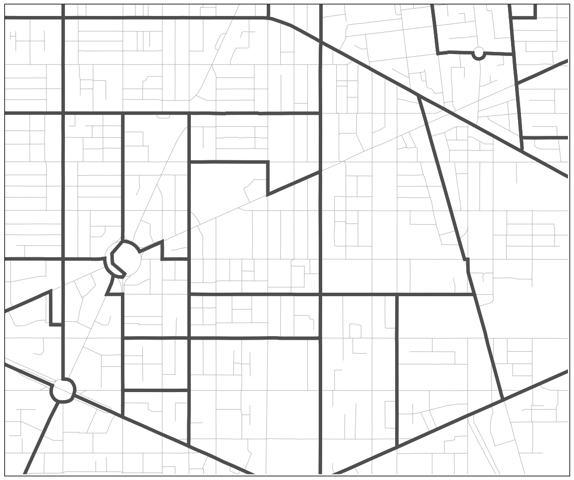
1. Assign Census 2010 block-level centroids to layers A and B.[46]
2. Assign population levels to each shape in layers A and B according to the location of each block centroid.

3. Merge layers A and B.
4. Assign population levels to each intersected feature created by overlaying layers A and B according to the location of each block centroid. Calculate the proportion of the population of each layer A shape that is contained in each layer B shape.
5. Calculate the proportion of the population of each layer B shape that is made up of each layer A shape.
6. Weight each variable in layer A by the appropriate weight calculated in steps 4 and 5.
7. Aggregate estimates calculated in step 6 for each shape in layer B.

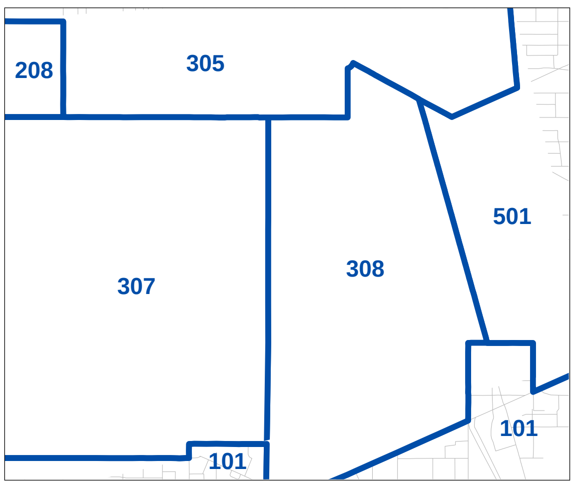
These steps can be applied to a buffered feature, such as a half-mile radius around metro stops. This method can also be used to resolve any discrepancies in geographic boundaries. For example, figures A.2 and A.3 illustrate how the boundaries of the 2000 and 2010 censuses changed. There are several types of boundary changes: (1) reassigned block group IDs without boundary change, (2) a block group in 2000 split into multiple block groups in 2010, (3) multiple block groups in 2000 merged into a single block group in 2010, and (4) segments of multiple block groups in 2000 collapsed into one or more block groups in 2010.

FIGURE A.1

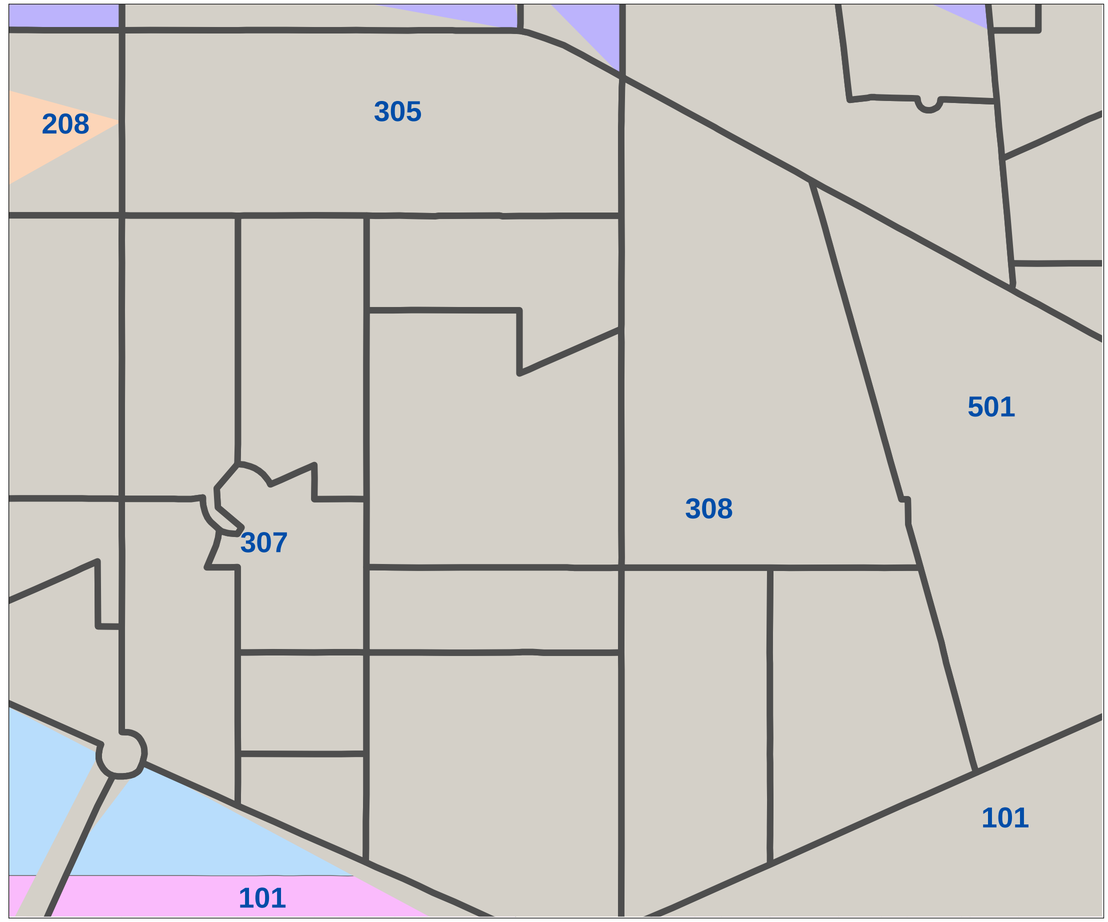
Boundary Incompatibility



Census block groups



Police service areas

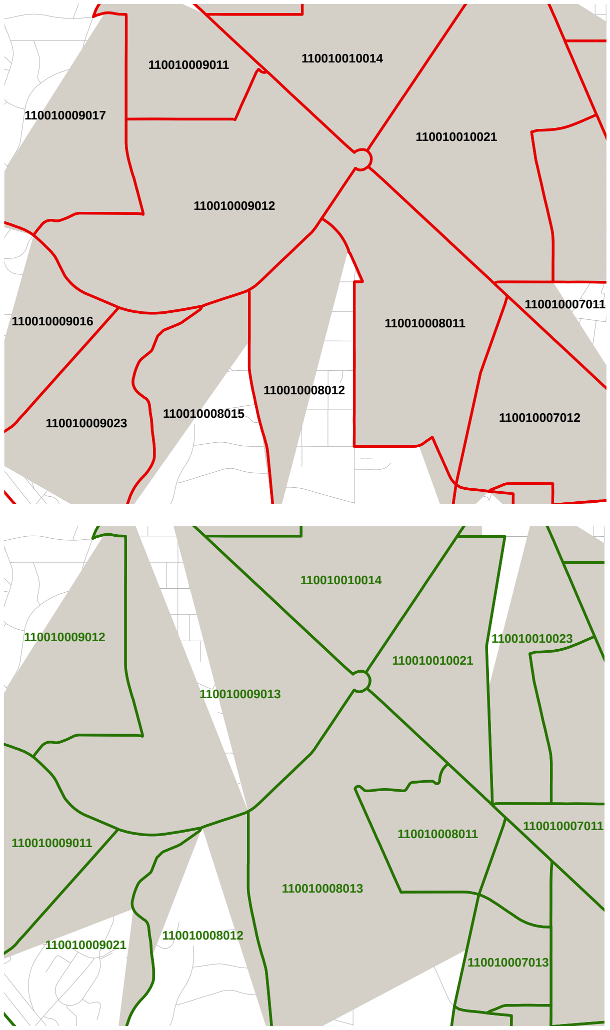


Census block group and police service area overlay

Source: Urban Institute analysis.

FIGURE A.2

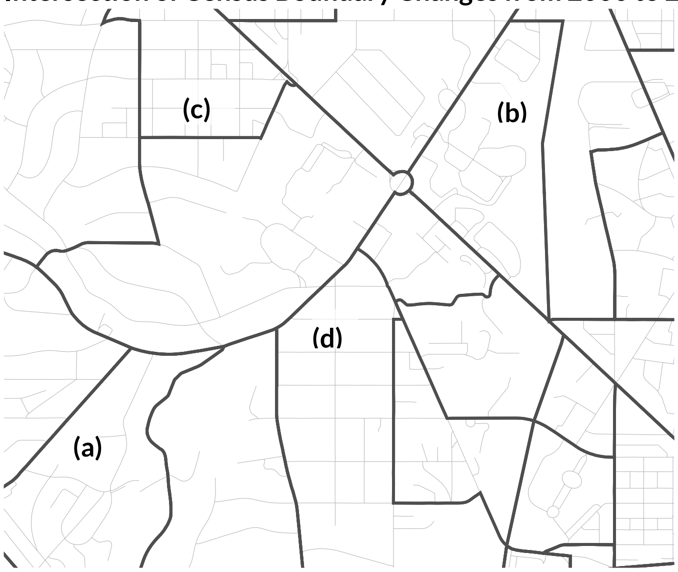
Census Boundary Change from 2000 to 2010 at the Block Group Level



Source: Urban Institute analysis.
Note: The Census Bureau provides a crosswalk file to resolve incompatibilities between 2000 and 2010 census boundaries. However, the crosswalk file is only available at the census block and tract levels. Given how much these boundaries changed, manipulating the block-level crosswalk to create a crosswalk analogue for block groups is not a straightforward process.

FIGURE A.3

Intersection of Census Boundary Changes from 2000 to 2010



Source: Urban Institute analysis.

Performance

Because weights are derived from populations, this method would work best if the unit of interest (e.g., people, household income, and crimes) was equally distributed according to the weighting unit (e.g., population, area, and households), which rarely is the case for spatial data. The performance of this method was tested to assess how reliably it can approximate data in practice.

This test used point data of crime incidents, which were aggregated to two areal units—census block groups and police service areas—so the actual number of crimes that occurred in each area were known. The test involved estimating the number of crimes per PSA using the block group-level data and vice versa; examining the correlation between the estimated number of crimes and the actual number of crimes; and visualizing the deviation, measured by (actual value - estimated value), which demonstrates how sensitive this method is to particular geographic areas.

Our test demonstrated that using smaller geographic areas to estimate data for larger areas (i.e., “small to large”) is less prone to error than the reverse method (i.e., “large to small”). Table A.1 shows that using police service areas to estimate the number of crimes in block groups leads to a higher average squared deviation (1.38) than the small-to-large method (0.02). It also shows that the correlation between the actual number of crimes in each block group and the estimated number of crimes in each block group derived from police service areas (r=0.54) is lower than the correlation of actual and estimated crimes in police service areas (r=0.98). The latter correlation indicates that police service area data were almost perfectly estimated from block group data.

TABLE A.1

Testing Error in Areal Summation

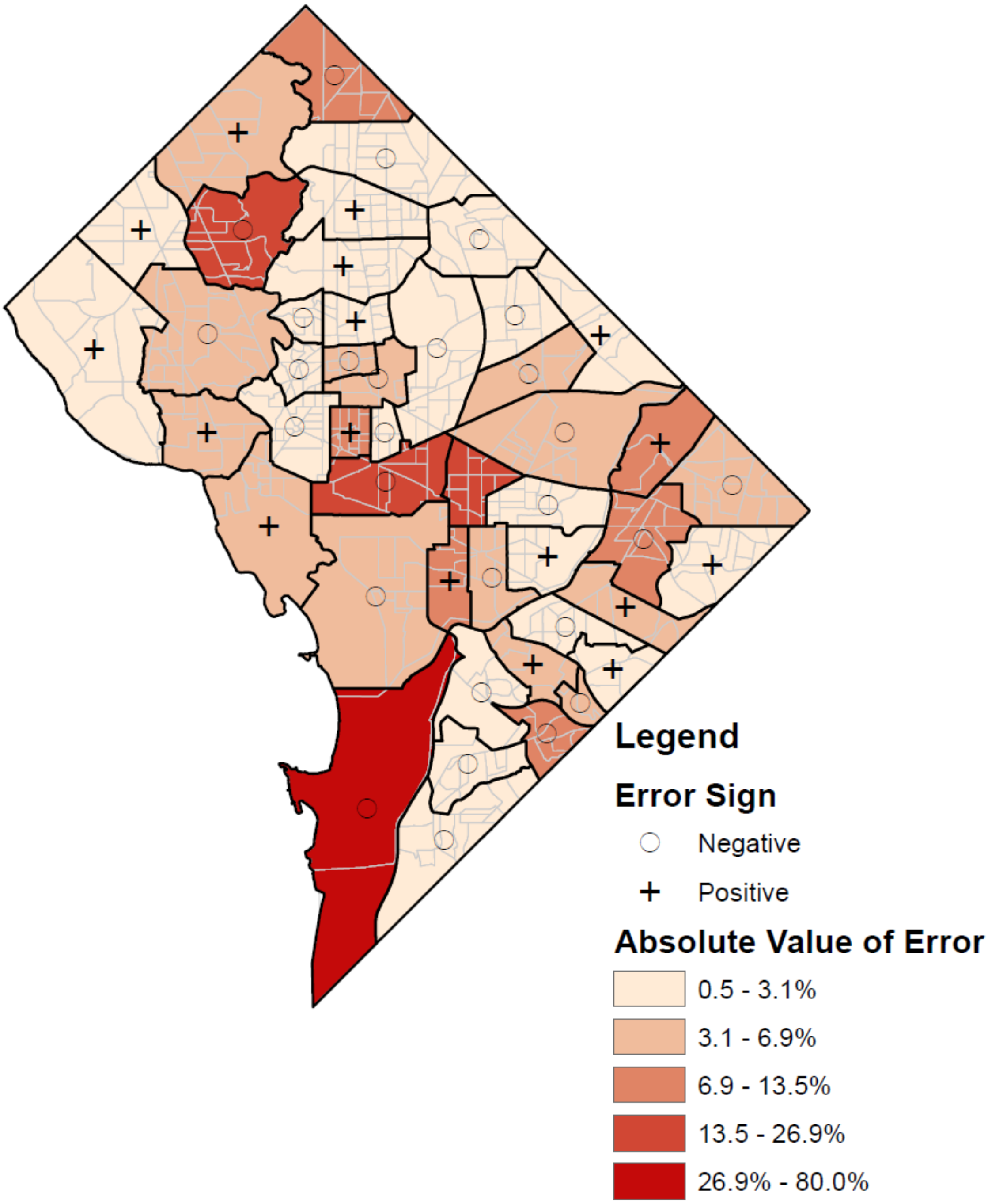
	(Actual Value - Estimated Value)^2					
	Sample size	Mean	Standard deviation	Minimum	Maximum	Correlation
Large to small						
PSA to CBG	448	1.38	6.51	0.00	87.05	0.54
Small to large						
CBG to PSA	46	0.02	0.09	0.00	0.64	0.98

Notes: CBG = census block group; PSA = police service area.

These results are also shown in figures A.4 and A.5 below. Figure A.4 maps the error in the estimated number of crimes in police service areas based on block group counts. Figure A.5 displays the error for the estimated number of crimes in block groups based on police service area counts. Once again, using a larger area to estimate data for a smaller area leads to greater error. The largest error in Figure A.5 (80 percent) was observed in a large police service area with low crime (four incidents). All other errors for the small-to-large calculation were less than 27 percent. The maps also suggest that error from this estimation method is not spatially concentrated. In other words, there is no conspicuous clustering of positive or negative errors or the edge effect (i.e., greater error in the units bordering Washington, DC, compared to units near the city center).

FIGURE A.4

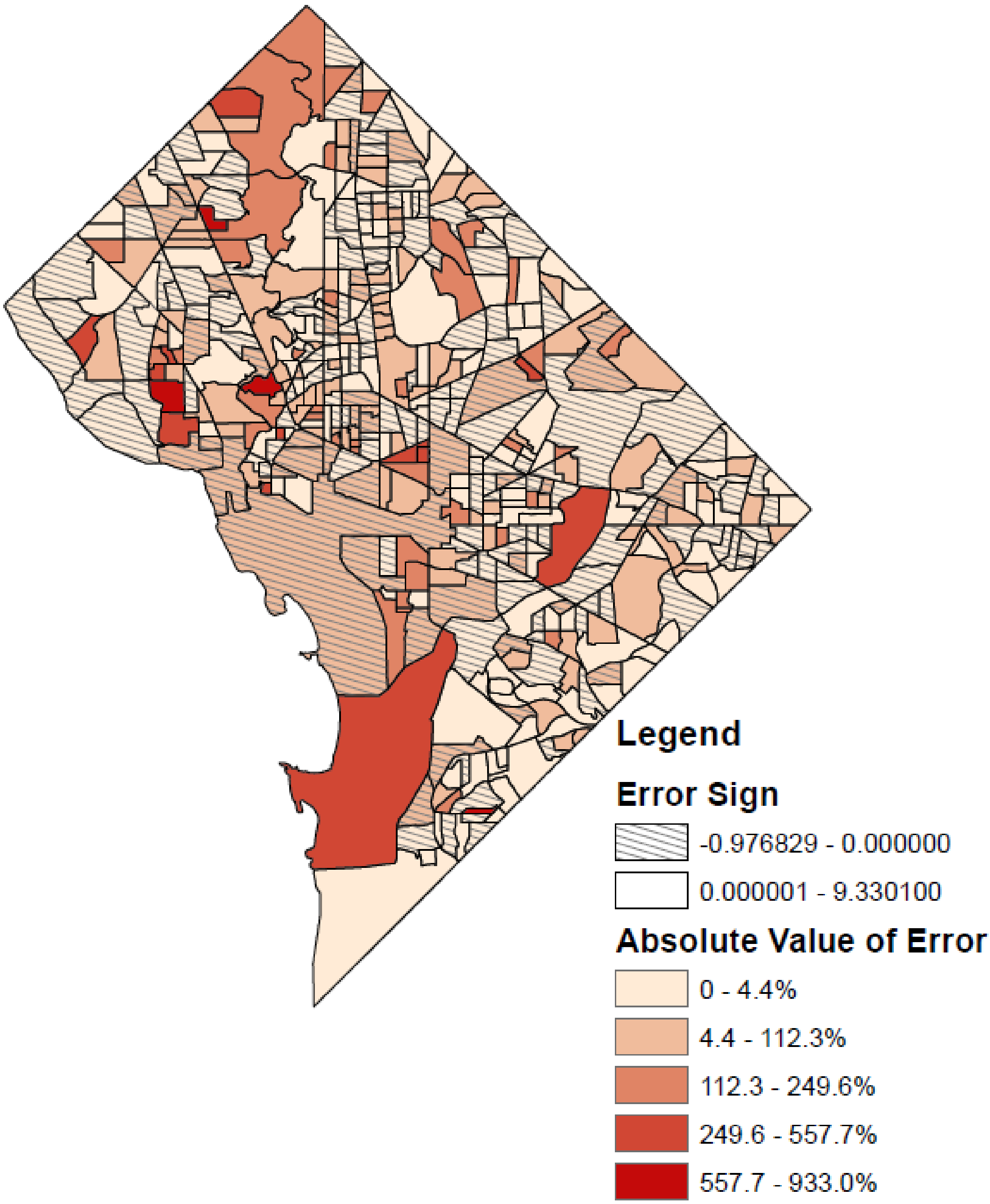
Crime Estimates for Police Service Areas Based on Block Group Counts



Source: Urban Institute analysis.

FIGURE A.5

Crime Estimates for Block Groups Based on Police Service Area Counts



Source: Urban Institute analysis.

Appendix B. Detailed Risk Terrain Model Steps

The risk terrain models in this project were developed with the RTMDx utility using the following steps:

1. The point shapefile of the dependent variable and the boundary polygon shapefile of the study extent were entered into the RTMDx utility.
2. Model risk factors with varying spatial influences were created from the input land use and other crime features. All geographic calculations were completed within the study extent using raster cells of fixed size. These risk factors measured whether the raster cells in the study extent were within a threshold distance of the feature or in an area of high density of the feature.
3. For the calculations in step 2, raster cells that fall within the threshold proximity were represented as 1 (highest risk) and the cells outside this proximity were represented as 0 (not-highest risk). Density variables were also reclassified into highest density and not-highest density regions. Highest density regions were regions with a density more than two standard deviations above mean density. These regions were represented as 1, whereas regions with a lower density were classified as not-highest density regions and represented as 0. These values were assembled into a table with rows representing cells and columns representing binary variables, and the count of the dependent variable at each raster cell was calculated.
4. Cross-validation was used to build a custom elastic net model of penalized Poisson regression with two fixed L2 penalties and optimized L1 penalties.
5. The elastic net penalization used in step 4 reduced a large set of model factors with different spatial operationalizations to a smaller set of factors by filtering them with statistical testing in simple linear modeling and then balancing the prediction model's fit with complexity by pushing variable coefficients toward zero. In each model, model factors that stood up to shrinkage in the penalized model were accepted as useful risk factors and passed to step 6.
6. This last step used a custom bidirectional stepwise regression with Bayesian information criterion scores to find the best risk terrain model for each model. The regression was repeated with two stepwise regression models: one assuming a Poisson distribution and the other assuming a negative binomial distribution. Based on the Bayesian information criterion scores, the best risk terrain model was chosen between these two regression models with different distributions. During this process, relative risk values, calculated by rescaling factor coefficients

based on the minimum and maximum risk values in the best risk prediction model, were produced for the risk factors included in the best model.

Notes

1. "Catalog of Established and Emergent Integrated Data Systems," National Neighborhood Indicators Partnership, last modified June 2, 2015, http://www.neighborhoodindicators.org/library/catalog/catalog-established-and-emergent-integrated-data-systems.

2. "Navigating the Compact," Interstate Commission for Adult Offender Supervision, accessed August 27, 2015, http://www.interstatecompact.org/About/NavigatingtheCompact.aspx.

3. "Camden, N.J., Police to Share Data with Hospitals for Insight into Better Health, Public Safety," *Emergency Management*, January 15, 2015, http://www.govtech.com/em/health/Camden-NJ-Police-Share-Data-with-Hospitals.html; Larry Kaplan, "Public-Private Data Partnership in Camden Locates Health Hotspots in Neighborhoods," *Nonprofit Quarterly*, January 13, 2015, http://nonprofitquarterly.org/2015/01/13/public-private-data-partnership-in-camden-locates-health-hotspots-in-neighborhoods/; Megan Smith and Roy L. Austin, "Launching the Police Data Initiative," *The White House Blog*, May 18, 2015, https://obamawhitehouse.archives.gov/blog/2015/05/18/launching-police-data-initiative.

4. Jenni Bergal, "Touting Prevention, Police Put Crime Info Online," *Stateline* (blog), January 20, 2015, http://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2015/1/20/touting-prevention-police-put-crime-info-online.

5. "Catalog of Established and Emergent Integrated Data Systems," National Neighborhood Indicators Partnership.

6. Dan Kedmey, "Catching Criminals with the Cloud," *Time*, May 14, 2015, http://time.com/3858364/caught-by-the-cloud/; Jeffrey Benzing and Halle Stockton, "Allegheny County Police Still Keep Many Records on Paper," *PublicSource*, March 7, 2015, http://publicsource.org/allegheny-county-police-still-keep-many-records-on-paper/.

7. Charlie Bush, "Enhancing Criminal Justice and Homeland Security Capabilities: N-DEx Fulfilling Its Vision to Support Law Enforcement," *Police Chief*, February 2013, http://iacpmag.wp.matrixdev.net/enhancing-criminal-justice-and-homeland-security-capabilities-n-dex-fulfilling-its-vision-to-support-law-enforcement/; Robert L. Mitchell, "It's Criminal: Why Data Sharing Lags Among Law Enforcement Agencies," *Computerworld*, October 24, 2013, http://www.computerworld.com/article/2486359/government-it/it-s-criminal--why-data-sharing-lags-among-law-enforcement-agencies.html.

8. "Information Sharing," US Department of Homeland Security, last modified April 13, 2016, http://www.dhs.gov/topic/information-sharing; Michael Haslip, and Paul Wormeli, "Interagency Information Sharing: The National Information Exchange Model," *Police Chief*, April 2007, http://iacpmag.wp.matrixdev.net/interagency-information-sharing-the-national-information-exchange-model/.

9. "Regional Crime Analysis Geographic Information System (RCAGIS)," National Archive of Criminal Justice Data, accessed February 27, 2017, http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/3372.

10. "About CLEMIS," Oakland County, accessed August 27, 2015, https://www.oakgov.com/clemis/Pages/about/default.aspx.

11. "ARJIS Agencies," Automated Regional Justice Information System, accessed August 27, 2015, http://www.arjis.org/SitePages/ARJISAgencies.aspx.

12. Pamela Scanlon, "Remote Biometrics Empower San Diego Officers," *Police Chief*, June 2014, http://www.policechiefmagazine.org/remote-biometrics-empower-san-diego-officers/.

13. "CODY Integrated Public Safety Software Solutions," CODY Systems, accessed August 27, 2015, http://www.codysystems.com/integrated-software-solutions.

14. Tod Newcombe, "Managing the Hidden Costs of Open Data," *Government Technology*, August 5, 2015, http://www.govtech.com/data/Managing-the-Hidden-Costs-of-Open-Data.html.

15. Rehan Chawdry, "Increasing Public Safety Data Sharing Interoperability," Hendon Media Group, accessed July 3, 2017, http://www.hendonpub.com/resources/article_archive/results/details?id=1698.

16. Michael Chau, Homa Atabakhsh, Daniel Zeng, and Hsinchun Chen. 2001. "Building an Infrastructure for Law Enforcement Information Sharing and Collaboration: Design Issues and Challenges."

17. Doug Wyllie, "Technology Isn't the (Biggest) Problem for Information Sharing in Law Enforcement," *PoliceOne*, April 30, 2009, http://www.policeone.com/police-products/communications/articles/1816539-Technology-isnt-the-biggest-problem-for-information-sharing-in-law-enforcement/.

18. "NNIP Lessons on Local Data Sharing," National Neighborhood Indicators Partnership, last modified February 14, 2016, http://www.neighborhoodindicators.org/library/guides/nnip-lessons-local-data-sharing.

19. Ibid.

20. Ibid.

21. "ARJIS Agencies," Automated Regional Justice Information System.

22. Newcombe, "Managing the Hidden Costs of Open Data."

23. "ARJIS Agencies," Automated Regional Justice Information System.

24. Ronald V. Clarke and John E. Eck, "Know How Hot Spots Develop," Crime Analysis for Problem Solvers in 60 Small Steps, accessed February 27, 2017, http://www.popcenter.org/learning/60steps/index.cfm?stepNum=17.

25. "Crime in the United States, 2010," Federal Bureau of Investigation, accessed September 23, 2015, https://ucr.fbi.gov/crime-in-the-u.s/2010/crime-in-the-u.s.-2010.

26. Michael Chau, Homa Atabakhsh, Daniel Zeng, and Hsinchun Chen. 2001. "Building an Infrastructure for Law Enforcement Information Sharing and Collaboration: Design Issues and Challenges."

27. "Interagency Data Sharing – A Primer," National Geospatial Advisory Committee, June 2011, https://www.fgdc.gov/ngac/ngac-interagency-data-sharing-primer-june-2011.pdf. "NNIP Lessons on Local Data Sharing," National Neighborhood Indicators Partnership.

28. See chapter 4, The Future of Interagency Data Integration, for a prospective review of emerging technologies and their implications for data sharing.

29. The NETS Database is the only time series database that provides information on establishment mobility patterns, sales growth performance, job creation and destruction, and changes in primary markets. This unique database contains 24 years of historical information (1990–2013) on over 50 million establishments with over 300 data fields. Current location data (zip code, city, state, metropolitan statistical area, longitude and latitude of the establishment and its headquarters) for each establishment, combined with standard industrial classification codes up to eight digits, enable researchers to observe the impact of any outcome on various indicators of business growth for very specific industries at micro levels of analysis (e.g., establishment, block, block group, census tract, zip code).

30. Jeffrey Stinson, "States Turn to Cloud Computing for Cost Savings," *Fiscal Times*, October 14, 2014, http://www.thefiscaltimes.com/2014/10/14/States-Turn-Cloud-Computing-Cost-Savings.

31. Daniel Humphries, "A Rating of 2014's Top Cybersecurity Issues," *Software Advice* (blog), http://www.softwareadvice.com/security/industryview/top-issues-report-2014/.

32. "Program Overview," FedRAMP, accessed February 22, 2017, https://www.fedramp.gov/about-us/about/.

33. Kimberly Nelson, “The Cloud – Not a Trend but an Enabler of 2015 City Tech Trends,” *CityNext Blog*, January 8, 2015, http://www.microsoft.com/en-us/citynext/blogs/the-cloud-not-a-trend-but-an-enabler-of-2015-city-tech-trends/default.aspx.

34. “SF OpenData,” City and County of San Francisco, accessed February 22, 2017, https://data.sfgov.org/.

35. Alexiei Dingli, “Introduction to Web Science: Web 2.0,” PowerPoint presentation, accessed July 3, 2017, http://staff.um.edu.mt/alexiei.dingli/IntroductionToWS/Lectures/2_1_3_Web2.0.ppt.

36. Joe Anuta, “Hudson Yards to Be First ‘Quantified Community,’” *Crain’s*, April 14, 2014, http://www.crainsnewyork.com/article/20140414/REAL_ESTATE/140419932/hudson-yards-to-be-first-quantified-community; Brian Libby, “Quantifying the Livable City,” *CityLab*, October 21, 2014, http://www.citylab.com/tech/2014/10/quantifying-the-livable-city/381657/; Sam Rega, “In New York City and Chicago, the Smart City is Here—and It’s Keeping Track of Everything.” *Business Insider*, June 3, 2015, http://www.businessinsider.com/introducing-the-smart-city-2015-6.

37. Office of the Press Secretary, “FACT SHEET: Administration Announces New ‘Smart Cities’ Initiative to Help Communities Tackle Local Challenges and Improve City Services,” news release, September 14, 2015, https://www.whitehouse.gov/the-press-office/2015/09/14/fact-sheet-administration-announces-new-smart-cities-initiative-help; Bernard Marr, “How Big Data and the Internet of Things Create Smarter Cities,” *Forbes*, May 19, 2015, http://www.forbes.com/sites/bernardmarr/2015/05/19/how-big-data-and-the-internet-of-things-create-smarter-cities/.

38. “A City-University Collaborative for Urban Innovation,” MetroLab Network, accessed February 27, 2017, http://metrolab.heinz.cmu.edu/.

39. WiredBlue, “Latest Version of MyPD Launched at IACP 2014,” news release, October 28, 2014, http://www.officer.com/press_release/12014376/latest-version-of-mypd-app-launched-at-iacp-2014.

40. “2015 Multi-City Innovation Campaign,” BlueLight, accessed February 27, 2017, http://2015cityinnovation.populr.me/bluelight.

41. “What is Open Data?” Open Knowledge International, accessed March 30, 2017, http://opendatahandbook.org/guide/en/what-is-open-data/

42. Ben Wellington, “Success: How NYC Open Data and Reddit Saved New Yorkers Over $55,000 a Year,” *I Quant NY* (blog), June 2, 2014, http://iquantny.tumblr.com/post/87573867759/success-how-nyc-open-data-and-reddit-saved-new; Ben Wellington, “Update: Single Fire Hydrant Nets NYC $33,000 a Year, not just $25,000,” *I Quant NY* (blog), April 24, 2014, http://iquantny.tumblr.com/post/83770853308/update-single-fire-hydrant-nets-nyc-33000-a.

43. Alex Engler, programmer analyst I, Urban Institute, interview, October 9, 2015.

44. Jonathan Schwabish, senior research associate, Urban Institute, interview, October 15, 2015.

45. “Transparency and Open Government: Memorandum for the Heads of Executive Departments and Agencies,” The White House, January 21, 2009, https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government.

46. Blocks are the smallest units of analysis available from the census.

References

Agrawal, Rakesh, Alexandre Evfimievski, and Ramakrishnan Srikant. 2003. “Information Sharing across Private Databases.” In *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*, 86–97. New York: Association for Computing Machinery.

Aguilar, Juan R. 2013. “Gunshot Location Systems the Transfer of the Sniper Detection Technology from Military to Civilian Applications.” Paper presented at the 47th International Carnahan Conference on Security Technology, October 8–11, 2013, Medellin, COL.

Archbold, Carol A. 2012. *Policing: A Text/Reader*. Thousand Oaks, CA: SAGE Publications.

Banipal, Kulwinder. 2006. “Strategic Approach to Disaster Management: Lessons Learned from Hurricane Katrina.” *Disaster Prevention and Management: An International Journal* 15 (3): 484–94.

Bouhaddou, Omar, Pradnya Warnekar, Fola Parrish, Nhan Do, Jack Mandel, John Kilbourne, and Michael J. Lincoln. 2008. “Exchange of Computable Patient Data between the Department of Veterans Affairs (VA) and the Department of Defense (DoD): Terminology Mediation Strategy.” *Journal of the American Medical Informatics Association* 15 (2): 174–83.

Brantingham, Patricia L., and Paul J. Brantingham. 1993. “Environment, Routine and Situation: Toward a Pattern Theory of Crime.” In *Routine Activity and Rational Choice: Advances in Criminological Theory*, vol. 5, edited by Ronald V. Clarke and Marcus Felson. New Brunswick, NJ: Transaction.

———. 1995. “Criminality of Place: Crime Generators and Crime Attractors.” *European Journal on Criminal Policy and Research* 3 (3): 1–26.

Briffault, Richard. 1996. “The Local Government Boundary Problem in Metropolitan Areas.” *Stanford Law Review* 48 (5): 1115–71.

Brunsdon, Chris, A. Steward Fotheringham, and Martin E. Charlton. 1996. “Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity.” *Geographical Analysis* 28 (4): 281–98.

Budinger, Zoe Baird, and Jeffrey H. Smith. 2011. “Ten Years After 9/11: A Status Report on Information Sharing.” Statement before the US Senate Committee on Homeland Security and Governmental Affairs, Washington, DC, October 12.

Caplan, Joel M., and Leslie W. Kennedy. 2010. *Risk Terrain Modeling Manual: Theoretical Framework and Technical Steps of Spatial Risk Assessment for Crime Analysis*. Newark, NJ: Rutgers Center on Public Security.

Carter, David L. 2004. *Law Enforcement Intelligence: A Guide for State, Local, and Tribal Law Enforcement Agencies*. Washington, DC: US Department of Justice, Office of Community Oriented Policing Services.

Chainey, Spencer, and Jerry Ratcliffe. 2013. *GIS and Crime Mapping*. Hoboken, NJ: John Wiley and Sons.

Cheminais, Rita. 2009. *Effective Multi-Agency Partnerships: Putting Every Child Matters into Practice*. Thousand Oaks, CA: SAGE Publications.

Chen, Min, Shiwen Mao, and Yunhao Liu. 2014. “Big Data: A Survey.” *Mobile Networks and Applications* 19 (2): 171–209.

Chisnall, David. 2013. “The Challenge of Cross-Language Interoperability.” *Communications of the ACM* 56 (12): 50–56.

Choi, Kyung-Shick, Mitch Librett, and Taylor J. Collins. 2014. “An Empirical Evaluation: Gunshot Detection System and Its Effectiveness on Police Practices.” *Police Practice and Research* 15 (1): 48–61.

Clingermayer, James C., and Richard C. Feiock. 1997. “Leadership Turnover, Transaction Costs, and External City Service Delivery.” *Public Administration Review* 57 (3): 231–39.

Coleman, Clive, and Jenny Moynihan. 1996. *Understanding Crime Data: Haunted by the Dark Figure*. Buckingham, UK: Open University Press.

Colvin, Caran A., and Angeline Goh. 2005. “Validation of the Technology Acceptance Model for Police.” *Journal of Criminal Justice* 33 (1): 89–95.

Comfort, Louise K., and Naim Kapucu. 2006. “Inter-organizational Coordination in Extreme Events: The World Trade Center Attacks, September 11, 2001.” *Natural Hazards* 39 (2): 309–27.

Coppock, J.T., and D.W. Rhind. 1991. “The History of GIS.” In *Geographical Information Systems: Principles and Applications*, edited by Paul A. Longley, David J. Maguire, Michael F. Goodchild, and David W. Rhind, 21–43. Hoboken, NJ: John Wiley and Sons.

Crank, John P. 2014. *Understanding Police Culture*. Abingdon, UK: Routledge.

Cytron, Naomi. 2014. “What Counts?” In *What Counts: Harnessing Data for America’s Communities*, edited by Naomi Cytron, Kathryn L. S. Pettit, and G. Thomas Kingsley, 428–40. San Francisco, CA: Federal Reserve Bank of San Francisco. http://www.whatcountsforamerica.org/wp-content/uploads/2014/12/WhatCounts.pdf.

Davidson, R.N. 1981. *Crime and Environ*ment. London: Croom Helm.

Emerson, Kirk, Tina Nabatchi, and Stephen Balogh. 2012. “An Integrative Framework for Collaborative Governance.” *Journal of Public Administration Research and Theory* 22 (1): 1–29.

Farrell, Graham. 2015. “Preventing Phone Theft and Robbery: The Need for Government Action and International Coordination.” *Crime Science* 4 (4).

FedRAMP (Federal Risk Authorization Management Program). 2014. “FedRAMP Forward: 2 Year Priorities.” Washington, DC: FedRAMP. https://www.fedramp.gov/files/2015/03/FedRAMP-Forward-2-Year-Priorities.pdf.

Few, Stephen. 2013. “Data Visualization for Human Perception.” In *The Encyclopedia of Human-Computer Interaction*, edited by Mads Soegaard and Rikke Friis Dam. 2nd ed. Aarhus, DNK: Interaction Design Foundation.

Finklea, Kristin M. 2011. *The Interplay of Borders, Turf, Cyberspace, and Jurisdiction: Issues Confronting U.S. Law Enforcement*. Washington, DC: Congressional Research Service.

Floyd, Ingbert R., M. Cameron Jones, Dinesh Rathi, and Michael B. Twidale. 2007. “Web Mash-ups and Patchwork Prototyping: User-Driven Technological Innovation with Web 2.0 and Open Source Software.” In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*, 86. New York: Institute of Electrical and Electronics Engineers.

Galster, George, and Lina Hedman. 2014. “Neighborhoods, Neighborhood Effects, and Residential Mobility: A Holistic View and Future Directions.” In *Strengthening Communities with Neighborhood Data*, edited by G. Thomas Kingsley, Claudia J. Coulton, and Kathryn L.S. Pettit. Washington, DC: Urban Institute.

Gazley, Beth. 2008. “Beyond the Contract: The Scope and Nature of Informal Government–Nonprofit Partnerships.” *Public Administration Review* 68 (1): 141–54.

Goldsmith, Andrew. 2005. “Police Reform and the Problem of Trust.” *Theoretical Criminology* 9 (4): 443–70.

Hampapur, Arun, Russell Bobbitt, Lisa Brown, Mike Desimone, Rogerio Feris, Rick Kjeldsen, Max Lu, Carl Mercier, Chris Milite, Stephen Russo, Chiao-fe Shu, and Yun Zhai. 2009. “Video Analytics in Urban Environments.” In *Proceedings of the 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, 128-133. New York: Institute of Electrical and Electronics Engineers.

Hanna, Richard, Andrew Rohm, and Victoria L. Crittenden. 2011. “We’re All Connected: The Power of the Social Media Ecosystem.” *Business Horizons* 54 (3): 265–73.

Harris, Kelly J., and William H. Romesburg. 2002. *Law Enforcement Tech Guide: How to Plan, Purchase and Manage Technology (Successfully!): A Guide for Executives, Managers and Technologists*. Washington, DC: US Department of Justice, Office of Community Oriented Policing Services.

Hawkins, Dan. 2006. *Tech Guide for Communications Interoperability: A Guide for Interagency Communications Projects*. Washington, DC: US Department of Justice, Office of Community Oriented Policing Services.

Helland, Eric, and Andrew B. Whitford. 2003. “Pollution Incidence and Political Jurisdiction: Evidence from the TRI.” *Journal of Environmental Economics and Management* 46 (3): 403–24.

IACP (International Association of Chiefs of Police). n.d. *Police-Corrections Partnerships Collaborating for Strategic Crime Control*. Alexandria, VA: International Association of Chiefs of Police.

———. 2015. *IACP National Policy Summit on Community-Police Relations: Advancing a Culture of Cohesion and Community Trust*. Alexandria, VA: International Association of Chiefs of Police.

Jackson, Brian A., Victoria A. Greenfield, Andrew R. Morral, and John S. Hollywood. 2014. “Police Department Investments in Information Technology Systems.” Santa Monica, CA: RAND Corporation.

Kapucu, Naim. 2006. “Interagency Communication Networks During Emergencies Boundary Spanners in Multiagency Coordination.” *The American Review of Public Administration* 36 (2): 207–25.

Kaza, Siddharth, and Hsinchun Chen. 2008. “Evaluating Ontology Mapping Techniques: An Experiment in Public Safety Information Sharing.” *Decision Support Systems* 45 (4): 714–28.

Kingsley, G. Thomas, Claudia J. Coulton, and Kathryn L. S. Pettit, eds. 2014. “Progress in Data and Technology.” In *Strengthening Communities with Neighborhood Data*. Washington, DC: Urban Institute.

Kinney, J. Bryan, Patricia L. Brantingham, Kathryn Wuschke, Michael G. Kirk, and Paul J. Brantingham. 2008. “Crime Attractors, Generators and Detractors: Land Use and Urban Crime Opportunities.” *Built Environment* 34 (1): 62–74.

Klinger, David A., and George S. Bridges. 1997. “Measurement Error in Calls-for-Service as an Indicator of Crime.” *Criminology* 35 (4): 705–26.

Krivo, Lauren J., and Ruth D. Peterson. 1996. “Extremely Disadvantaged Neighborhoods and Urban Crime.” *Social Forces* 75 (2): 619–48.

La Vigne, Nancy, and Julie Wartell. 2001. *Mapping Across Boundaries: Regional Crime Analysis*. Washington, DC: Police Executive Research Forum.

La Vigne, Nancy, Jake Cowan, and Diana Brazzell. 2006. *Mapping Prisoner Reentry: An Action Research Guidebook*. Washington, DC: Urban Institute.

Lee, Jinkyu, and H. Raghav Rao. 2007. “Exploring the Causes and Effects of Inter-Agency Information Sharing Systems Adoption in the Anti/Counter-Terrorism and Disaster Management Domains.” In *Proceedings of the 8th Annual International Conference on Digital Government Research: Bridging Disciplines & Domains*, 155–63. Digital Government Society.

Linden, Russell M. 2003. *Working Across Boundaries: Making Collaboration Work in Government and Nonprofit Organizations.* Hoboken, NJ: John Wiley and Sons.

Lindsay, Rachael, Thomas Jackson, and Louise Cooke. 2010. “Mobile Access to Information Systems in Law Enforcement: An Evaluation of Its Implications for Data Quality.” In *Proceedings of the International Conference of Information Management and Evaluation*, 188. Sonning Common, UK: Academic Conferences and Publishing International.

MacDonald, Ziggy. 2002. “Official Crime Statistics: Their Use and Interpretation.” *Economic Journal* 112 (477): F85–106.

Maltz, Michael D. 1999. *Bridging Gaps in Police Crime Data: A Discussion Paper from the BJS Fellows Program*. Washington, DC: US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

Maltz, Michael D., and Joseph Targonski. 2002. "A Note on the Use of County-Level UCR Data." *Journal of Quantitative Criminology* 18 (3): 297–318.

Maltz, Michael D., Andrew C. Gordon, and Warren Friedman. 1990. *Mapping Crime in Its Community Setting: Event Geography Analysis*. New York: Springer.

Mamalian, Cynthia A., and Nancy La Vigne. 1999. "The Use of Computerized Crime Mapping by Law Enforcement: Survey Results." Washington, DC: US Department of Justice, Office of Justice Programs, National Institute of Justice.

Markovic, John, and Christopher Stone. 2002. "Crime Mapping and the Policing of Democratic Societies." New York: Vera Institute of Justice.

Markowitz, Fred E., Paul E. Bellair, Allen E. Liska, and Jianhong Liu. 2001. "Extending Social Disorganization Theory: Modeling the Relationships between Cohesion, Disorder, and Fear." *Criminology* 39 (2): 293–319.

McCormick, Kathleen A., Joyce Sensmeier, Patricia Dykes, Erin N. Grace, Susan A. Matney, Kathleen M. Schwartz, and Marla J. Weston. 2015. "Exemplars for Advancing Standardized Terminology in Nursing to Achieve Sharable, Comparable Quality Data Based Upon Evidence." *Online Journal of Nursing Informatics* 19 (2).

McGillivray, Fiona, and Alastair Smith. 2004. "The Impact of Leadership Turnover on Trading Relations Between States." *International Organization* 58 (3): 567–600.

Merrill, William, Katayoun Sohrabi, Lewis Girod, Jeremy Elson, Fredric Newberg, and William J. Kaiser. 2002. "Open Standard Development Platforms for Distributed Sensor Networks." In *Unattended Ground Sensor Technologies and Applications IV*, edited by Edward M. Carapezza. Bellingham, WA: SPIE.

Myers, Samuel L. 1980. "Why Are Crimes Underreported? What is the Crime Rate? Does It 'Really' Matter?" *Social Science Quarterly* 61 (1): 23–43.

National Institute of Justice. 2008. "Effective Police Communications Systems Require New 'Governance.'" Washington, DC: US Department of Justice, Office of Justice Programs, National Institute of Justice.

Nedović-Budić, Zorica, and Jeffrey K. Pinto. 2000. "Information Sharing in an Interorganizational GIS Environment." *Environment and Planning B: Urban Analytics and City Science* 27 (3): 455–74.

Nunn, Samuel. 2001. "Police Information Technology: Assessing the Effects of Computerization on Urban Police Functions." *Public Administration Review* 61 (2): 221–34.

Oakland County. 2014. "Courts and Law Enforcement Management Systems: What is it?" Pontiac, MI: Oakland County. https://www.oakgov.com/clemis/Documents/What%20is%20CLEMIS%202014%20March.pdf.

Pain, Rachel, Robert MacFarlane, Keith Turner, and Sally Gill. 2006. "'When, Where, If, and But': Qualifying GIS and the Effect of Streetlighting On Crime And Fear." *Environment and Planning A* 38 (11): 2055–74.

Pardo, Theresa A., Anthony M. Cresswell, Fiona Thompson, and Jing Zhang. 2006. "Knowledge Sharing in Cross-Boundary Information System Development in the Public Sector." *Information Technology and Management* 7 (4): 293–313.

Pardo, Theresa A., J. Ramon Gil-Garcia, and Luis F. Luna-Reyes. 2010. "Collaborative Governance and Cross-Boundary Information Sharing: Envisioning a Networked and IT-Enabled Public Administration." In *The Future of Public Administration Around the World: The Minnowbrook Perspective*, edited by Rosemary O' Leary, David M. Van Slyke, and Soonhee Kim, 129–39. Washington, DC: Georgetown University Press.

Patil, D.J., and Hilary Mason. 2015. *Data Driven: Creating a Data Culture*. Sebastopol, CA: O'Reilly Media, Inc.

Pearson, Siani, and Azzedine Benameur. 2010. "Privacy, Security and Trust Issues Arising from Cloud Computing." In *2nd International Conference on Cloud Computing Technology and Science Proceedings*, 693–702. New York: Institute of Electrical and Electronics Engineers.

Penney, Terry L. 2014. "Dark Figure of Crime (Problems of Estimation)." In *The Encyclopedia of Criminology and Criminal Justice*, edited by Gerben Bruinsma and David Weisburd. New York: Springer.

Piza, Eric L., Leslie W. Kennedy, and Joel M. Caplan. 2011. "Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies." *Journal of Quantitative Criminology* 27 (3): 339–62.

Popović, Krešimir, and Željko Hocenski. 2010. "Cloud Computing Security Issues and Challenges." *MIPRO 2010: 33rd International Convention Conference Proceedings*, 344–49. New York: Institute of Electrical and Electronics Engineers.

Popp, Robert, Thomas Armour, Ted Senator, and Kristen Numrych. 2004. "Countering Terrorism through Information Technology." *Communications of the ACM* 47 (3): 36–43.

President's Task Force on 21st Century Policing. 2015. *Final Report of the President's Task Force on 21st Century Policing*. Washington, DC: US Department of Justice, Office of Community Oriented Policing Services.

Raleigh, Erica, and George Galster. 2014. "Neighborhood Disinvestment, Abandonment, and Crime Dynamics." *Journal of Urban Affairs* 37 (4): 367–98.

Raphael, Steven, and Rudolf Winter-Ebmer. 2001. "Identifying the Effect of Unemployment on Crime." *Journal of Law and Economics* 44 (1): 259–83.

Ratcliffe, Jerry H. 2003. "Suburb Boundaries and Residential Burglars." Canberra: Australian Institute of Criminology.

———. 2012. *Intelligence-Led Policing*. Abingdon, UK: Routledge.

Rengert, George F., and Brian Lockwood. 2009. "Geographical Units of Analysis and the Analysis of Crime." In *Putting Crime in its Place*, edited by David Weisburd, Wim Bernasco, and Gerben J. N. Bruinsma, 109–22. New York: Springer.

Rich, Thomas F. 1995. "The Use of Computerized Mapping in Crime Control and Prevention Programs." Washington, DC: US Department of Justice, Office of Justice Programs, National Institute of Justice.

Roman, John, and Aaron Chalfin. 2007. "Is There an iCrime Wave?" Washington, DC: Urban Institute.

Sanders, Greg. 2014. "Mashup City: Tools for Urban Life and Urban Progress in the Internet Age." In *Strengthening Communities with Neighborhood Data*, edited by G. Thomas Kingsley, Claudia J. Coulton, and Kathryn L.S. Pettit. Washington, DC: Urban Institute.

Santos, Rachel Boba. 2012. *Crime Analysis with Crime Mapping*. Thousand Oaks, CA: SAGE Publications.

Schwabe, William, Lois M. Davis, and Brian A. Jackson. 2001. *Challenges and Choices for Crime-Fighting Technology Federal Support of State and Local Law Enforcement*. Santa Monica, CA: RAND Corporation.

Sheptycki, James. 2004. "Organizational Pathologies in Police Intelligence Systems: Some Contributions to the Lexicon of Intelligence-Led Policing." *European Journal of Criminology* 1 (3): 307–32.

Skogan, Wesley. 1974. "The Validity of Official Crime Statistics: An Empirical Investigation." *Social Science Quarterly* 55 (1): 25–38.

———. 1986. "Fear of Crime and Neighborhood Change." *Crime and Justice* 8: 203–29.

———. 2012. "Disorder and Crime." In *The Oxford Handbook of Crime Prevention*, edited by Brandon C. Welsh and David P. Farrington, 173–88. Oxford: Oxford University Press.

Song, Justin, Martin A. Andresen, Patricia L. Brantingham, and Valerie Spicer. 2015. "Crime on the Edges: Patterns of Crime and Land Use Change." *Cartography and Geographic Information Science* 44 (1): 51–61.

Swartz, Nikki. 2008. "Agencies Must Standardize Data Labels." *Information Management* 42 (5).

Taylor, Ralph B., and Adele V. Harrell. 1996. *Physical Environment and Crime*. Washington, DC: US Department of Justice, Office of Justice Programs, National Institute of Justice.

Thapa, Rajesh Bahadur, and Ronald C. Estoque. 2012. "Geographically Weighted Regression in Geospatial Analysis." In *Progress in Geospatial Analysis*, edited by Yuji Murayama, 85–95. Tokyo: Springer Japan.

Tilley, Nick, Johnathan Smith, Stephen Finer, Rosie Erol, Corrine Charles, and John Dobby. 2004. *Problem-Solving Street Crime: Practical Lessons from the Street Crime Initiative*. London: Home Office, Research Development and Statistics Directorate.

Tobler, Waldo. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46 (2): 234–40.

Trojanowicz, Robert C., and Bonnie Bucqueroux. 1998. *Community Policing: How to Get Started*. Cincinnati, OH: Anderson.

Troy, Austin, and J. Morgan Grove. 2008. "Property Values, Parks, and Crime: A Hedonic Analysis in Baltimore, MD." *Landscape and Urban Planning* 87 (3): 233–45.

van Schendel, Willem, and Itty Abraham, eds. 2005. *Illicit Flows and Criminal Things: States, Borders, and the Other Side of Globalization*. Bloomington: Indiana University Press.

Weisburd, David, Gerben J.N. Bruinsma, and Wim Bernasco. 2009. "Units of Analysis in Geographic Criminology: Historical Development, Critical Issues, and Open Questions." In *Putting Crime in its Place*, edited by David Weisburd, Gerben J.N. Bruinsma, and Wim Bernasco, 3–31. New York: Springer.

West, Darrell M. 2010. "Saving Money through Cloud Computing." Washington, DC: Brookings Institution.

Williams, Christine B., Martin Dias, Jane Fedorowicz, Dax Jacobson, Sonia Vilvovsky, Steve Sawyer, and Michael Tyworth. 2009. "The Formation of Inter-organizational Information Sharing Networks in Public Safety: Cartographic Insights on Rational Choice and Institutional Explanations." *Information Polity* 14 (1–2): 13–29.

Wolf, Robert V. 2012. *Law Enforcement and Public Health: Sharing Resources and Strategies to Make Communities Safer*. Washington, DC: US Department of Justice, Office of Community Oriented Policing Services.

Woodroffe, Rosie, and Joshua R. Ginsberg. 1998. "Edge Effects and the Extinction of Populations inside Protected Areas." *Science* 280 (5372): 2126–28.

Zickuhr, Kathryn, and Aaron Smith. 2012. *Digital Differences*. Washington, DC: Pew Research Center.

About the Authors

Nancy La Vigne is the director of the Justice Policy Center at the Urban Institute. She is a widely published researcher on prisoner reentry, criminal justice technologies, crime prevention, policing, and the spatial analysis of crime and criminal behavior. Her work appears in scholarly journals and practitioner publications and has made her a sought-after spokesperson on several related subjects.

Ellen Paddock is a research associate in the Justice Policy Center. Her work focuses on the overlap between public health and criminal justice and on public safety strategies that effectively reduce crime while minimizing collateral damage to vulnerable communities. Before joining Urban, Paddock completed internships with a youth gang intervention program, a reentry job training program, and an interagency criminal justice resource center. She graduated from Duke University in May 2014 with a BA in public policy and Arabic.

Yasemin Irvin-Erickson is a senior research associate in the Justice Policy Center. Her research primarily focuses on crime prevention and social protection through the lens of inclusive growth and sustainable development.

KiDeuk Kim is a senior research associate in the Justice Policy Center, where he leads multidisciplinary research teams to examine issues related to criminal justice interventions and policies. He is also a visiting fellow at the US Department of Justice, Bureau of Justice Statistics. Before joining Urban, he was a senior researcher for the State of New York, providing strategic planning and research oversight for numerous task forces on criminal justice issues. His current research focuses on understanding decisionmaking in the criminal justice system and evaluating criminal justice programs. His research has appeared in prominent academic journals and has been cited in several media outlets.

Bryce Peterson is a senior research associate in the Justice Policy Center. His research focuses on correctional policy, children of justice-involved parents, video surveillance and body-worn camera technologies, federal and state justice statistics, and prison population forecasting. He conducts both quantitative and qualitative research and has received grants from the National Institute of Justice, the National Institute of Corrections, and the Bureau of Justice Assistance, among other funders.

Samuel Bieler is a former research associate in the Justice Policy Center, where he worked on issues of gun violence, crime mapping, and social impact innovation.

STATEMENT OF INDEPENDENCE

URBAN INSTITUTE · ELEVATE · THE · DEBATE

2100 M Street NW
Washington, DC 20037

www.urban.org