Helping Law Enforcement Use Data from Mobile Applications

A Guide to the Prototype Mobile Information and Knowledge Ecosystem (MIKE) Tool

Edward Balkovich, Don Prosnitz, Steven C. Isley, Anne Boustead, Bonnie Triezenberg
This project was supported by Award No. 2011-IJ-CX-K058, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.
Preface

This is the second of two reports describing an electronic surveillance study sponsored by the National Institute of Justice. Our aim in this study was to clarify the new sources of electronic surveillance data; educate stakeholders about what data are available, where they are located, and what legal principles govern law enforcement use of these data; and inform the policy debate about electronic surveillance. We constrained our study to criminal investigations by state, local, and tribal law enforcement—the constituency of the National Institute of Justice. Our study produced a tool (named the Mobile Information and Knowledge Ecosystem [MIKE]) that would help interested stakeholders—law enforcement, commercial enterprises, regulators, legislators, and the public (including advocacy groups)—better understand the relationships between the data, its sources, and applicable legal constraints. This prototype is not currently widely distributed.

This volume, written for practitioners interested in using MIKE, describes the prototype that we developed. It is intended to serve as users’ manual for MIKE, by explaining its principles, illustrating its potential uses, and describing how to add content. Although this tool is not widely available, our goal in writing this users’ manual was to provide vital information to entities interested in adopting or extending this tool in the future. It should be of interest to those in law enforcement, policy analysis, and commerce interested in electronic surveillance and using our prototype.

RAND Justice Policy

The research reported here was conducted in the RAND Justice Policy Program, which spans both criminal and civil justice system issues with such topics as public safety, effective policing, police-community relations, drug policy and enforcement, corrections policy, use of technology in law enforcement, tort reform, catastrophe and mass-injury compensation, court resourcing, and insurance regulation. Program research is supported by government agencies, foundations, and the private sector.

This program is part of RAND Justice, Infrastructure, and Environment, a division of the RAND Corporation dedicated to improving policy- and decisionmaking in a wide range of policy domains, including civil and criminal justice, infrastructure protection and homeland security, transportation and energy policy, and environmental and natural resource policy.

Questions or comments about this report should be sent to the project leader, Edward Balkovich (Edward_Balkovich@rand.org). For more information about RAND Justice Policy, see www.rand.org/jie/justice-policy or contact the director at justice@rand.org.
# Contents

Preface ........................................................................................................... iii  
Figures and Tables .......................................................................................... vii  
Summary .......................................................................................................... xi  
Acknowledgments ............................................................................................ xiii  

## CHAPTER ONE

**Introduction** ................................................................................................. 1  
Description of MIKE ........................................................................................ 2  
Organization of This Volume .......................................................................... 3  

## CHAPTER TWO

**User’s Guide to MIKE** ................................................................................ 5  
Introducing MIKE ............................................................................................. 6  
Investigative Use Case .................................................................................... 9  
Policy Use Case ............................................................................................... 27  
Commercial Use Case ................................................................................... 40  

## CHAPTER THREE

**Curator’s Guide to MIKE** .......................................................................... 49  
Curating and Updating MIKE ......................................................................... 49  
Populating and Developing MIKE ................................................................... 71  

## CHAPTER FOUR

**Architect’s Guide to MIKE** ....................................................................... 81  
Architectural Goals and Attributes ................................................................ 81  
Prior Approaches ........................................................................................... 82  
Architectural Overview of MIKE .................................................................... 84  

## APPENDIX

**Stakeholder Reactions to the Wiki** ............................................................. 93  

Abbreviations .................................................................................................. 99  
Glossary .......................................................................................................... 101  
References ...................................................................................................... 109
Figures

2.1. MIKE’s Homepage ................................................................. 6
2.2. MIKE’s Description of Corporate Information ...................... 7
2.3. MIKE’s Description of a Mobile App ....................................... 8
2.4. MIKE’s Description of Technical Information ....................... 9
2.5. MIKE’s Description of Legal Information ............................ 10
2.6. Initial Query of MIKE ............................................................ 12
2.7. Results of the Query ............................................................. 13
2.8. The Information About a Gmail Account .............................. 13
2.9. Discovering What Apps Were on the Device ......................... 14
2.10. Discovering the Apps That Were Downloaded to the Phone .... 14
2.11. The Google Play App .......................................................... 15
2.12. Information Describing the Data Google Play Collects .......... 16
2.13. Information About 2703(d) Court Orders .......................... 17
2.14. The Snapchat App ............................................................. 18
2.15. The Instagram App ............................................................ 19
2.16. The Instagram Corporation ................................................ 19
2.17. Querying All Mobile Apps to Discover Those that Collect Location Information .................. 20
2.18. Foursquare’s Data Contexts ................................................ 21
2.19. The Foursquare Radar App ................................................ 22
2.20. Google Calendar ................................................................. 23
2.21. Yahoo Basic Subscriber Information ................................... 23
2.22. SquareCash/Transaction Page .............................................. 24
2.23. Results of the Query for All Mobile Apps that Collect Location Information .................. 25
2.24. Information About Geoloqi ................................................ 26
2.25. Locating Policy Issues Addressed by MIKE ......................... 29
2.26. Discovering Policy Issues .................................................. 29
2.27. Legal Protections for Location Information ......................... 30
2.28. ECPA/Basic Subscriber Information .................................. 31
2.29. ECPA/Administrative Subpoena ...................................... 32
2.30. Texas Code of Criminal Procedure ................................... 33
2.31. Order for Installation and Use of Mobile Tracking Device ... 34
2.32. United States v Jones ......................................................... 35
2.33. Search Results ................................................................. 36
2.34. Search Results ................................................................. 37
2.35. Location Category ............................................................ 38
2.36. Foursquare App ............................................................... 39
2.37. Geoloqi ............................................................................ 40
2.38. Cell Site Location ............................................................. 40
# Tables

<table>
<thead>
<tr>
<th>1.1.</th>
<th>Where to Find Topics of Interest</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.</td>
<td>Level 1 Information Sources</td>
<td>74</td>
</tr>
<tr>
<td>3.2.</td>
<td>Level 2 Information Sources</td>
<td>74</td>
</tr>
<tr>
<td>3.3.</td>
<td>Level 3 Information Sources</td>
<td>75</td>
</tr>
<tr>
<td>4.1.</td>
<td>Relationships Specified in the Entity-Relationship</td>
<td>88</td>
</tr>
<tr>
<td>A.1.</td>
<td>Concerns Regarding Surveillance of the Mobile App Ecosystem</td>
<td>96</td>
</tr>
<tr>
<td>A.2.</td>
<td>Necessary Components of the Tool</td>
<td>97</td>
</tr>
<tr>
<td>A.3.</td>
<td>Reactions to MIKE</td>
<td>98</td>
</tr>
</tbody>
</table>
Mobile devices, apps, and services—collectively referred to as the *mobile ecosystem* by Silicon Valley—are very helpful to state and local law enforcement in criminal investigations. Virtually everyone carries a cell phone, most of which are smartphones. The phone, the apps it uses, and the services it accesses collect and retain enormous amounts of information. The metadata describing these collections and the stored content of these collections are much richer and subject to different access rules than conventional wire line wiretaps.

Despite this apparent trove of useful information for criminal investigations, law enforcement organizations face several challenges in using it. Technology and commercial relationships evolve quickly, and law enforcement organizations have trouble keeping up with the level of detail needed to make investigative choices backed by well-formulated surveillance requests. They also face considerable uncertainty about the law and jurisprudence that govern access to commercially collected data. Moreover, there is a growing awareness of and availability of countermeasures, such as encryption, that can be used to evade law enforcement surveillance.

Our aim for this project was to facilitate understanding of the information contained within the mobile app ecosystem and the ways law enforcement can access this information. From the beginning, this project was designed to meet the needs of a wide variety of stakeholders—not only the law enforcement officers who may request information but also the commercial entities that must comply with these requests, the policymakers who craft rules governing these requests, and the advocacy groups that may argue for changes in existing policy, regulation, and law. We sought to convey information about both the sources and flows of information through the mobile app ecosystem and the legal rules that control how information may be shared and spread.

To meet these goals, we developed and prototyped a tool to describe the mobile ecosystem and its relationship to laws regulating law enforcement access, which we call the Mobile Information and Knowledge Ecosystem (MIKE). This prototype is not widely distributed. The companion volume, *Electronic Surveillance of Mobile Devices: Understanding the Mobile Ecosystem and Applicable Surveillance Law* (Balkovich et al., 2015), motivated the need for MIKE and provided a high-level description of it. That report also describes our view of the poles of the current policy debate: the golden age of surveillance versus going dark. This debate was one of the driving concerns behind this project. This volume is a detailed technical description of the tool (MIKE), extensive examples of its intended uses, and discussion of methods for adding additional structure and content. We also provide a brief description of stakeholder impressions of the tool.

MIKE is intended for the use of law enforcement, commercial firms, policy analysts, and others. We believe that it assists with navigating and understanding this complex environment. Our tool is interactive and dynamically extensible. Interactivity makes it possible for law
enforcement, commerce, and policy analysts to provide, navigate, and analyze technical, commercial, and legal information related to electronic surveillance. Dynamic extensions make it possible to easily incorporate rapidly evolving technology and commercial relationships and relate them to more slowly changing law and jurisprudence.

Preliminary assessments of this prototype by representatives of these three stakeholder communities—law enforcement, commercial firms, and policy analysts—who have had an opportunity to explore MIKE and its potential uses suggest that our approach has utility in all three domains. For example, law enforcement representatives pointed out that MIKE could be useful not only in guiding investigations but as a training tool allowing users to explore how to use electronic surveillance in a hypothetical investigation. But a key concern of all these stakeholders is how such a potentially valuable tool could be sustained. While we have not definitively answered this question, we have identified a number of plausible options for consideration. The next step would be to analyze these options, fleshing out the requirements for how to further develop the current prototype in ways that would support a particular approach to sustainment.

MIKE is based on an extension of the software used to run Wikipedia, so users are often already familiar with its design and layout. This document functions as a guide to the structure and intended use of MIKE. We begin by discussing the aims and research questions addressed by this project, then briefly describe the results of our research. Each of the remaining chapters in the document is intended to serve as a guide for a different group that may be involved in MIKE’s further development, operation, and/or use. Chapter Two describes how MIKE can be used and is intended for those who adopt and utilize MIKE. Chapter Three describes how MIKE can be edited and maintained and is intended for those who are responsible for curating MIKE. These chapters may be read independently. Chapter Four describes the architecture and design of MIKE and functions as a guide for those interested in further developing MIKE. We conclude with appendixes defining the terminology used in this document and discussing stakeholder reactions to MIKE.
The authors want to thank a number of people who helped with this project. First is our colleague Karlyn Stanley. A number of law enforcement partners assisted us as expert advisers about the conduct of investigations and use of electronic surveillance, including representatives of the District of Columbia Metropolitan Police Department, the Indiana State Police, and the Las Vegas Metropolitan Police Department. In addition, we benefited from a number of individuals who helped us to understand commercial concerns and perspectives and from the perspectives of policy analysts, advocates, and decisionmakers active in this area. We would also like to thank our reviewers, Joshua Baron of the RAND Corporation and Marcus K. Rogers of Purdue University, for their helpful comments and feedback.
Mobile devices, applications (apps), and services—collectively referred to as the mobile ecosystem by Silicon Valley—are very helpful in criminal investigations by state and local law enforcement, and also assist public safety responses. Virtually everyone carries a cell phone, most of which are smartphones. The phone, the apps it uses, and the services it accesses collect and retain enormous amounts of information. The metadata describing these collections and the stored content of these collections can be much richer than and subject to different access rules from those for conventional wiretaps.

Despite this apparent trove of useful information for criminal investigations, law enforcement organizations wishing to tap into it face several important challenges. Technology and commercial relationships evolve quickly, and law enforcement organizations have trouble staying current with information needed to make investigative choices backed by well-formulated surveillance requests. These organizations also face considerable uncertainty about the law and jurisprudence that governs their access to commercially collected data. Moreover, there is a growing awareness of and availability of countermeasures that can be used to evade law enforcement surveillance, such as encryption.

Facing this technical complexity and legal uncertainty, law enforcement may unnecessarily constrain itself and thereby risk public safety or may overreach and potentially violate privacy rights and civil liberties. Neither is good. Law enforcement needs to be as well informed as possible when making decisions about the use of electronic surveillance. While some argue that the mobile ecosystem has created a “golden age” of surveillance (Swire and Ahmad, 2011), the growing use of countermeasures and implementations of devices and services could eventually prevent effective, lawful use of electronic surveillance.

The complex and evolving implementation of mobile devices and services, and the opacity of the mobile ecosystem they constitute, makes it difficult for law enforcement and other stakeholders to remain fully aware of the potential sources of electronic surveillance data relevant to a criminal investigation. Commercial entities in this ecosystem (such as the developer of the mobile device and its operating system, the carrier providing connectivity to the device, and the service providers the user of the phone wittingly and unwittingly accesses using installed

---

1 According to Pew Research Center, 90 percent of Americans over the age of 18 own a cell phone, and 64 percent own a smartphone (Pew Research Center, 2014).

2 Throughout this document, we adopt a broad definition of electronic surveillance, including not only information law enforcement collects directly using electronic devices but also information commercial entities collect and that law enforcement then obtains and uses. For further discussion and justification of this definition of electronic surveillance, see Balkovich et al., 2015.

3 For example, Apple’s approach to iOS security (Apple Inc., 2016).
apps) reside in multiple jurisdictions and are governed by federal and state laws and international agreements. Lawful access to surveillance requires being mindful of what each of these bodies of law allows in particular contexts (e.g., search incident to an arrest). When there are alternative sources within the ecosystem for types of data (e.g., location information), the cost and response time of commercial entities may inform law enforcement’s choice. Law enforcement requests and court orders for data must be formulated and presented in a manner that is easy to comply with and provides law enforcement with a focused body of data that is interpretable. For example, an ill-formed request can overwhelm law enforcement with data, even if the request is not deemed to overreach. Each commercial entity has its own preferences for the format for requests and orders, the cost of compliance, response time, and the structure and scope of the data it will make available in response to the request or order. Failure of law enforcement to comply with these standards can slow the response and delay the investigation.

**Description of MIKE**

To help stakeholders understand and address the complexity of the mobile ecosystem, we developed a tool—the Mobile Information and Knowledge Ecosystem (MIKE)—that systematically relates the commercial entities that collect data about mobile device users to the types of data they collect, the rules governing law enforcement access to that data, and the commercial procedures for lawfully accessing that data in a way that simplifies or at least makes this complexity manageable. We did this by constructing a map relating these concepts to enable users to navigate, analyze, and apply the relationships between these concepts to conduct effective electronic surveillance with low impact on commercial entities and to analyze policy issues related to electronic surveillance.4

As a proof of concept demonstration, we prototyped a working version of such a map, using MIKE, and populated it with data describing about 40 apps and their service providers. MIKE is based on an extension of the software used to run Wikipedia, so users are often already familiar with its design and layout. It is intended for the use of law enforcement officers who are interested in using information collected by the mobile app ecosystem to investigate crimes, policymakers who must regulate law enforcement use of this data, and commercial entities that must respond to requests from law enforcement for information. This prototype has not yet been distributed widely.

We believe this prototype demonstrates a way to systematically capture the relationships among data types, attributes of the law, and commercial entities relevant to electronic surveillance of mobile devices. MIKE is primarily intended for the use of law enforcement, but commercial firms and policy analysts might also use it. Unlike prior descriptions of the mobile app ecosystem, such as static maps,5 MIKE is based on a wiki and is interactive and dynamically extensible. Interactivity makes it possible for law enforcement, commerce, and policy analysts to provide, navigate, and analyze technical, commercial, and legal information related to

---

4 In everyday usage, a map relates such geographical concepts as streets and rivers to one another using a two-dimensional space. Maps can also show how facts about abstract concepts relate to one another, without reference to geography. Our map relates businesses, apps, services, commercially collected data, and electronic surveillance law in a multidimensional space that can be explored using any of these abstractions as the starting point.

5 LUMAscape is an example of a static map; we discuss prior models later.
electronic surveillance. User-friendly dynamic extensions make it possible to incorporate new technology and commercial relationships and relate them to changing law and jurisprudence.

After developing a preliminary version of MIKE, we turned our attention to understanding its implementation and use. We explored MIKE’s utility by demonstrating its use to stakeholders and asking for their feedback (see Appendix A). Preliminary assessments from representatives of these three stakeholder communities—law enforcement, commercial firms, and policy analysts—who have had an opportunity to explore the prototype and its potential uses suggest that our approach has utility in all three domains.6 But a key concern of all these stakeholders is how to sustain such a potentially valuable tool. While we have not definitively answered this question, we have identified a number of plausible options for consideration. The next step would be to analyze these options, fleshing out the requirements for further developing the current prototype in ways that would support a particular approach to sustainment. Since this tool has not yet been implemented, we also leave for future work any evaluation of the outcomes of this tool, such as how it would affect electronic surveillance use, policy development, and law enforcement efficiency.

It is important to note that MIKE is not intended to be a case management system. It is intended to provide reference material for an investigator but it is not and should not be used as a repository for case specific information collected in an investigation. While MIKE’s reference materials include citations to relevant electronic surveillance law, these should not be construed as offering automated legal advice. It is not intended to be a substitute for the guidance of legal counsel in an investigation.

Organization of This Volume

The work described in the companion volume (Balkovich et al., 2015) motivated the creation of MIKE. That volume provides a high-level description of MIKE, how it is used, stakeholder reactions to it, and the issues stakeholders surfaced. This volume provides a detailed guide to the structure and intended use of MIKE.

This guide employs a stakeholder-focused, multiple-view approach to documentation, per recommended best practice for documenting software architecture (Software Engineering Institute, undated). We have not assumed that all readers will read all sections of this guide or read the sections in any particular order. Table 1.1 outlines the topics that emerged from our analysis of who might use this guide and what they might be interested in, then identifies which portions of the guide address these topics. Each section and chapter identified in the table can be read independently, although all document users should read “Introducing MIKE” in Chapter Two.

In the chapters that follow, we assume (depending on the specific chapter) that the reader has a basic understanding of (1) software tool development, its maintenance, and use; (2) basic mobile device technology, communications, apps, and services; and/or (3) familiarity with electronic surveillance and governing law. While this volume was not written to provide an introduction to any of the above, the glossary provides specific definitions for terminology used in this volume and may be useful for those less familiar with these topics. Note, however,

---
6 These assessments are anecdotal and are opinions about the potential utility of a tool like MIKE. Assessing the effectiveness of MIKE was beyond the scope of this project.
that the meaning of terminology used in these fields can depend on the context—our glossary addresses only the specific context of MIKE.

<table>
<thead>
<tr>
<th>Stakeholder Interests</th>
<th>See</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquirer—Department of Justice</td>
<td></td>
</tr>
<tr>
<td>Understanding project goals and status</td>
<td>Chapter One</td>
</tr>
<tr>
<td>Understanding the architecture and constraints</td>
<td>Chapter Four</td>
</tr>
<tr>
<td>Developers</td>
<td></td>
</tr>
<tr>
<td>Understanding the architecture and constraints</td>
<td>Chapter Four</td>
</tr>
<tr>
<td>Users</td>
<td></td>
</tr>
<tr>
<td>Understanding how to use MIKE</td>
<td>Chapter Two</td>
</tr>
<tr>
<td>Conducting an investigation using e-surveillance data</td>
<td>“Investigative Use Case” (Chapter Two)</td>
</tr>
<tr>
<td>Understanding the access rules for e-surveillance data</td>
<td>“Investigative Use Case” (Chapter Two)</td>
</tr>
<tr>
<td>Exploring policies related to e-surveillance</td>
<td>“Policy Analysis Use Case” (Chapter Two)</td>
</tr>
<tr>
<td>Responding to e-surveillance requests</td>
<td>“Commercial Entity Use Case” (Chapter Two)</td>
</tr>
<tr>
<td>Curators</td>
<td></td>
</tr>
<tr>
<td>Understanding how to update MIKE</td>
<td>“Curating and Updating Mike” (Chapter Three)</td>
</tr>
<tr>
<td>Finding and validating technical data</td>
<td>“Populating and Developing Mike” (Chapter Three)</td>
</tr>
<tr>
<td>Updating or entering technical data</td>
<td>“Populating and Developing Mike” (Chapter Three)</td>
</tr>
<tr>
<td>Analyzing content of legal data</td>
<td>“Source of Legal Information” (Chapter Three)</td>
</tr>
<tr>
<td>Updating or entering legal data</td>
<td>“Source of Legal Information” (Chapter Three)</td>
</tr>
<tr>
<td>Updating structure and relationships</td>
<td>“Editing the Structure of the Map” (Chapter Three)</td>
</tr>
</tbody>
</table>
This user’s guide is designed for three classes of individuals who may particularly benefit from the implementation and use of MIKE. First, law enforcement officers seeking to use the information contained in the mobile app ecosystem during an investigation can use MIKE to understand where to find that information and what they must do to obtain it. Second, policy analysts and advocates can use MIKE to understand information sharing in the mobile app ecosystem more generally, to appreciate the current policy landscape, discover potential policymaking needs, and assess the implications of proposed policy changes. Finally, commercial entities may use MIKE to understand how and why they may be asked to comply with information requests from police and what they can do to streamline this process.

We assume that readers of this document have processes for conducting investigations, analyzing policy, and supporting law enforcement. We further assume that readers will be able to relate our stylized examples to existing processes to understand the utility of MIKE. Chapter Four is a reference guide for MIKE and provides an architectural overview.

This chapter describes MIKE and illustrates ways to use it. MIKE is intended to be an interactive tool, and it is therefore difficult to fully convey its use in linear text. The narratives describe the user’s thought processes and interactions with MIKE, relying on screen shots to show queries and responses within MIKE. While we name specific companies and mobile apps and used their published privacy policies to construct the cases, we did not confirm that the indicated data were actually collected or would be provided to law enforcement under the indicated condition. These cases are intended only to illustrate how MIKE could be used, not to provide examples of its actual use.

After introducing MIKE and providing an overview of its content, we offer cases that demonstrate how each class of users might use MIKE. The first case is an example of how MIKE could be used during the investigation of a crime. Since MIKE may be helpful during many stages of an investigation, this use case follows the course of a single investigation. While this is a simplified and contrived example, it suffices to illustrate the many ways MIKE could contribute to an investigation. The second case demonstrates MIKE’s use as a policy analysis tool and focuses on policies regarding location information, since this is currently a particularly salient area for policymakers. The last case illustrates how a small app developer confronting its first request from law enforcement might use MIKE.
Introducing MIKE

We developed MIKE as a tool to describe how information is captured, stored, and shared within the mobile app ecosystem and the legal protections that govern law enforcement access to the data. MIKE was introduced in a companion volume (Balkovich et al., 2015). Much of that introduction is repeated here to allow this chapter to stand alone.

MIKE was constructed to function like a map, guiding users through the complicated web of information underlying the mobile app ecosystem. However, MIKE was not developed to function as a case management system or to provide legal advice. Rather, as will be shown in this user’s guide, MIKE can be used to help navigate the entities, information, and laws that constitute the mobile app ecosystem.

MIKE is based on an extension of the software used to run Wikipedia, so users are often already familiar with its design and layout. As shown in Figure 2.1, there are several different ways to find information within MIKE: A search bar at the top can be used to seek information about specific terms; quick links in the middle of the page provide easy access to lists of concepts and entities; and a query function further down the page allows the user to tailor a search for particular types of information. MIKE’s homepage also provides access to help functions.

Figure 2.1
MIKE’s Homepage
MIKE contains pages pertaining to corporate, technical, and legal information. Figure 2.2 depicts a page describing a particular corporate entity: in this example, Google Inc. This page provides information about the data the entity collects, the products it creates, and its subsidiaries. Much of this information is conveyed in a structured format, so that information from different pages can be easily compared and combined to generate insights about all the entities within MIKE. Purple text denotes links to other pages, which a user can follow to find more information. There is also space for unstructured information about the entity, so that contributors to MIKE share additional knowledge about corporate entities.

A user who wants to know more about the services a commercial entity administers can follow the links provided under the “Owner Of” heading. For example, a user interested in

Figure 2.2
MIKE’s Description of Corporate Information
learning more about the Gmail can follow the link provided. The user will then be directed to MIKE’s page describing the Gmail mobile app (Figure 2.3). This page provides information about the technical requirements for using the app, its funding model, the approximate number of people who use the app, and the data types the app collects. Similar pages are available for all mobile apps included in MIKE.¹

A user who wants to know more about the information types Gmail collects can follow the link for “Account Information” at the top of the list under “Data Contexts” in Figure 2.3. This directs the user to one of MIKE’s technical information pages (Figure 2.4). These pages describe the type of information collected, how this information is stored, and whether and how it is encrypted and provide additional sources of information in case the user wants to learn more about the data. “Storage Mode” is intended to indicate whether an app stores data on the device or on a server used to implement the app (or both).² Additionally, this page describes the legal requirements law enforcement must comply with to obtain this information. The appropriate legal requirements were automatically generated based on the type of entity collecting the data and the type of data collected, among other things.

Figure 2.3
MIKE’s Description of a Mobile App

¹ MIKE, as described here, is a proof-of-concept prototype. Anyone implementing MIKE might find it useful to introduce pages describing individual versions of operating systems for mobile devices, e.g., iOS 9.1.

² It is conceivable that the approach to encryption on the phone is different from that used on a server. It would probably be useful for an implementation of MIKE to incorporate such details.
MIKE also contains pages describing the legal document that can be used to obtain various types of information. These pages can be accessed through links on the data context pages, or by searching MIKE for the name of the document. Figure 2.5 depicts one such page. These pages include both structured and unstructured information about the laws governing law enforcement access to specific types of information. Wherever possible, the text of the relevant statute is quoted to assist in further research.

To describe the contents and structure of MIKE, we have shown one way of navigating MIKE. However, MIKE was designed to be flexible to meet the needs of a variety of stakeholders. In the remainder of this section, we describe how three key groups of stakeholders—law enforcement, policy analysts and advocates, and commercial entities—could use MIKE.

**Investigative Use Case**

This case specifically illustrates the complexity of navigating information describing mobile devices, apps, and the relevant law and jurisprudence. It also demonstrates the value of being able to effectively and easily navigate a comprehensive description of the mobile ecosystem.

The example is based on a hypothetical murder investigation. It demonstrates how MIKE might be used to support the investigation by identifying potentially relevant sources of data collected by the mobile ecosystem. The major features of the example are as follows:

- The victim is found in a back alley near a casino. The victim's name is initially unknown. He has no wallet, an empty phone holster on his belt, and a business card with the name and address of a local hostel in an inside pocket. Robbery is the presumed motive.
Data collected in the mobile ecosystem will be used to first identify the victim and then to precisely determine the time of death. The strategy is first to determine which apps were on the victim’s (stolen) phone and then to retrieve data from the app developers and service providers.

This data are then used to determine the victim’s relationships and identify a potential suspect.

Tracking the suspect’s location and financial transactions (again digitally) leads to a high likelihood that the suspect is indeed the perpetrator.

This example shows how MIKE can help an investigator determine where to find relevant data within the mobile ecosystem and what legal issues she must address to access the data. The example is highly stylized to illustrate many of MIKE’s features and does not presume that this is the only way the investigation might proceed or that the indicated pathways are the only (or even the best) way to obtain the data the investigator seeks.

**Case Description**

**Background**

A body turns up early in the morning in a back alley in close proximity to local casinos. The victim has no watch, jewelry, wallet (leading to a presumption of robbery as the motive), or other means of identification but does have a business card for a local hostel in an inside pocket. The victim has an empty cell phone holster, so it is assumed that his phone has also been stolen.
When shown a picture of the victim, the hostel clerk confirms that the victim was a guest who paid in cash. The name the victim provided on the hostel register comes up blank on all primary database searches (including the web), and a fingerprint search also comes back with no results. Interviews with other guests provide no additional information except that the victim seemed to be carrying a lot of cash.

The investigator knows that the victim had a smartphone and that, if she can somehow match a digital record of the phone to the victim, many investigative paths may open up. The hostel has free Wi-Fi. On request, the hostel voluntarily provides the router logs for the 24 hours immediately preceding the murder. The router reveals that eight smartphones accessed the Wi-Fi during the 24 hours preceding the murder and provides their Media Access Control (MAC) addresses and phone IDs. This information also establishes that the device was an Android phone.

A combination of voluntary submissions of the smartphones by most of the guests and employees, and the timing of when the victim was present at the hostel allow the investigator to isolate the MAC address and the “Android_ID” for the missing phone. Further examination of the router’s logs shows that the particular phone accessed the Gmail email server using the internet Post Office Protocol (POP). Given the two potential IDs (MAC and Android_ID) and evidence that the unknown victim might have had a Gmail account, what can the investigator learn?

**Starting the Investigation**

The starting point is to use MIKE to discover what can be learned about Gmail. Here “Gmail” is entered into MIKE’s search bar (upper right of Figure 2.6). As can be seen, entering text into the search bar generates a drop-down list of pages in MIKE that contain that text in the title. The user can also click the “Search” button for more-complete results.

After clicking on “Gmail,” the user is directed to MIKE’s page describing that app. The results show that Gmail does in fact maintain MAC addresses and device IDs, along with log information about Gmail accounts (the “Data Contexts” table in the lower center of Figure 2.7).

We entered the types of information apps gather (data contexts) into MIKE by extracting them from a reading of the relevant privacy statements rather than from technical examinations of the app. Future versions of MIKE could include data gathered by automated technical analyses of apps. Researchers at Carnegie Mellon University have taken a promising approach to this (Amini et al., 2013).

Clicking on “Account Information” under “Data Contexts” reveals the information Google collects about Gmail accounts—real names and phone numbers (“Has Data Type” in

---

3 Facebook offers a new facial recognition app that could potentially be used to try to identify the victim, but the example assumes that no match is made using this publicly available capability.

4 Third-party data retention rules are relevant to requests like this. The data in such a log can be overwritten faster than law enforcement can request access to it. This gives rise to calls for more-uniform data retention rules with attendant cost implications for the commercial entity retaining the data.

5 The Android_ID is randomly generated at first boot of the smartphone. It can be changed via a factory reset of the device.

6 For a definition of data contexts and a discussion their use within MIKE, see Chapter Four.
the upper left table in Figure 2.8). The investigator also learns that an administrative subpoena is needed to access this information (upper right table, “Legal Information”).

The investigator is also interested in determining what other apps might have been on the phone identified by its MAC and Android_ID, which is presumed to belong to the victim. Therefore, she uses the search bar (upper right of Figure 2.8) to search for “applications.” The search returns a list of pages that include the word “applications,” as shown in Figure 2.9. “App Name” is on the list (center of the search results) and indicates that a number of apps keep track of apps on a phone.

Clicking on “App Name” reveals several apps that track other apps on the phone (Figure 2.10). Given that the investigator knows the device is an Android phone, the relevant one is Google Play (center left in the “Apps and Corporations With This Data” table).

---

7 Our use of the term real name reflects the aspirations of service providers who ask users to identify themselves. In practice, it is easy to use a fake name and difficult for a service provider to detect that the name is fake.

8 Comparable app-tracking apps also exist for iPhones. These apps make note of and retain the apps that have been downloaded to the device, even if the apps have since been deleted. This is typically done to manage licenses and enable the reinstallation of a deleted app. These app-tracking apps are only aware of the apps downloaded from the store associated with the app, e.g., Google or Apple.
As MIKE content grows, the user may be confronted with searching through a list of names that could be overwhelming. A comprehensive list is useful for purposes of discovery or training. However, it may be useful to add query tools in an implementation of MIKE to help narrow such a list when a more specific list is required.
Figure 2.9
Discovering What Apps Were on the Device

Figure 2.10
Discovering the Apps That Were Downloaded to the Phone
Clicking on the link “Google Play App” reveals information about Google Play (Figure 2.11):

- The owner of the Google Play app is Google Inc.
- The Google Play app is an online marketplace
- The types of data the Google Play app collects are described in the table called “Data Contexts”
- References for information about the Google Play app are also provided.

The investigator is interested in the downloaded apps that might have been on the missing phone, presumed to belong to the victim, so she clicks on “Downloaded Apps” under “Data Contexts.” This reveals the information MIKE has about Google Play (Figure 2.12). This information includes

- who the data about installed apps are sent to—in this case, Google
- how the data about installed apps are stored—remotely (at Google) or locally (on the device)
- resolution—how often the data about installed apps are stored with Google (with each app installation)

---

Figure 2.11
The Google Play App

---

<table>
<thead>
<tr>
<th>Owner</th>
<th>Google Inc.</th>
<th>OS Versions</th>
<th>Android</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activities</td>
<td>Online Marketplace</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Libraries</td>
<td>Google Cloud Messaging, Google In-App Billing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revenue Model</td>
<td>In-app advertising, In-app purchases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Installed User Base</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Website</td>
<td>play.google.com</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Types of data Google Play collects**

- Context Name | Transmitted To | Data Types | Resolution | Retention |
- Address | Google Inc. | Static | Life of account |
- Book Previews | Google Inc. | Books | Usage based | Life of account |
- Device Identifiers | Google Inc. | Mobile Network Identifiers | Static | Life of account |
- Device Information | Google Inc. | Hardware Details, Timestamp | Static | Life of account |
- Downloaded Apps | Google Inc. | App Name, Timestamp | Usage based | Life of account |
- Email Address | Google Inc. | Email Address | Static | Life of account |
- Real Name | Google Inc. | Real Name | Static | Life of account |

**References**

retention—how long Google retains the data about installed apps (the life of the account)
• references and other data.

The description of the data Google Play collects is automatically associated with information about the relevant law for obtaining access to it. In this case, the “Legal Information” table indicates that a 2703(d) order may be required to obtain information about the apps on the victim's device from Google.

Clicking on 2703(d) reminds the investigator about what is needed to obtain such an order (Figure 2.13). In case of a 2703(d) order, “specific and articulable facts . . . reasonable grounds . . . are relevant and material to an ongoing criminal investigation.” In addition, links are provided to other relevant MIKE pages and legal references. Such pages could easily include links to the specific forms Google requires when served with an Electronic Communications Privacy Act (ECPA) 2703(d) order and how such orders should be conveyed to Google.9

Armed with this information and the appropriate Google forms and Google contact information (which MIKE could also provide), the investigator then requests information from Google. The request can be narrow, perhaps based on a MAC address and/or Android_ID. Here, MIKE may have helped the investigator minimize her data request. An officer with no knowledge about what data Google saves might have requested “all information related to person X” from Google—and Google would rightly protest such a broad request.10 MIKE has helped the investigator learn what is available, and tailor her request to Google to include

---

9 Some commercial entities want to make such information available to law enforcement, rather than receiving an ad hoc order and then having to ask for revisions. This cycle both is expensive for the commercial entity and slows the response to law enforcement.

10 Courts have recently rebuked those making blanket requests for data from such services as Google (Apuzzo, 2014). MIKE will help an investigator limit the information requested from Google.
just the information relevant to her investigation with confidence that she will not be missing something important.

In our example, however, the investigator seeks the following information for the specific Android_ID:

- the associated Gmail address
- the name, address, and phone number of the account associated with the Gmail address
- additional unique identifiers (e.g., International Mobile Equipment Identity numbers) for future use in the investigation
- a list of downloaded apps associated with the Gmail account.

Google provides the investigator the Gmail address, but the real name and physical address are apparently false. A phone number is also provided, but calling it indicates that the user is out of the service area. Google also provides information that shows the following apps have been downloaded to the victim’s phone:

- Snapchat
- Instagram
• Google Calendar
• Foursquare
• SquareCash.

The investigator now has an email account associated with “a” phone, but needs to confirm that it is in fact the victim’s phone. Matching a picture of the victim with the email account would provide further confirmation. The investigator knows that both Snapchat and Instagram are photo apps. Alternately, she could query MIKE for all photo apps and see a list of apps that store photos, then could compare that list with the one of downloaded apps from Google Play to find the potentially relevant apps.

MIKE can now be used to learn more about the relevant photo apps installed on the phone. A MIKE query about the Snapchat app (Figure 2.14) indicates that Snapchat cloud storage is only transient. Because the investigator does not have the phone, Snapchat is unlikely to be able to provide the needed information.

However, by querying MIKE about Instagram, the investigator can discover that Instagram does retain photos for the life of the account and identifies users by their (now known) email address (Figure 2.15). MIKE also indicates that legal requirements to access this data may differ depending on whether the photos have been stored for 180 days or more. A 2703(d) court order is required if recent photos are the ones of interest to the investigator.

Figure 2.14
The Snapchat App

![Cloud storage photos is transient](image)

Snapchat

<table>
<thead>
<tr>
<th>Data Contexts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context Name</td>
</tr>
<tr>
<td>Account Information</td>
</tr>
<tr>
<td>Contact Info</td>
</tr>
<tr>
<td>Device Information</td>
</tr>
<tr>
<td>Local Photos</td>
</tr>
<tr>
<td>Messages</td>
</tr>
<tr>
<td>Photos</td>
</tr>
</tbody>
</table>

References

2. [http://www.snapchat.com/privacy](http://www.snapchat.com/privacy)
By consulting the “Facts About Instagram, LLC” available at the bottom of MIKE’s Instagram, LLC, page (Figure 2.16), the investigator learns that Facebook owns Instagram. She notes that this means further information might be available from Facebook.
Using the email address to request information from Instagram, the investigator is provided a “selfie” photograph, which shows the victim. The picture is geotagged (latitude and longitude) and depicts a family barbeque. The longitude and latitude point to a residence in another state. She calls the victim’s family, which makes a positive identification.

Having identified the victim, the investigator now wishes to determine where he was, what he was doing, and what financial transactions he might have made in the few days before he was killed.

Following the same approach as for discovering apps, the investigator returns to MIKE’s homepage and follows the link to query function “Query All Mobile Apps.” First, knowing that the victim had an Android phone, she specifies Android OS. Next, she begins to type “location” into the “Data” search box; autocomplete then offers a drop-down listing several choices. She chooses the Global Positioning System (GPS), which is likely to be more precise and accurate than Cell Site Location, and runs the query. This produces the list in Figure 2.17.

The GPS location query returns multiple apps, one of which is Foursquare (which is also on the Google Play list of apps installed on the phone). Proceeding to the Foursquare page, MIKE reveals two GPS location data contexts of interest (Figure 2.18). One data context is “Check-

\[\text{Figure 2.17} \]
\textbf{Querying All Mobile Apps to Discover Those that Collect Location Information}

<table>
<thead>
<tr>
<th>Mobile App</th>
<th>Owner</th>
<th>Activities</th>
<th>OS Versions</th>
<th>Libraries</th>
<th>Revenue Models</th>
<th>User Base</th>
<th>Privacy Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angry Birds</td>
<td>Rovio Inc. CP</td>
<td>Gaming</td>
<td>Android</td>
<td>Flurry Analytics, InMobi, Jumptap</td>
<td>Paid app</td>
<td>1,700,000,000</td>
<td>true</td>
</tr>
<tr>
<td>Badoo</td>
<td>Badoo Trading Ltd.</td>
<td>Social network operation</td>
<td>iOS</td>
<td>Facebook SDK</td>
<td>Freemium</td>
<td></td>
<td>true</td>
</tr>
<tr>
<td>Brightest LED</td>
<td>Surfex Technology</td>
<td>Social network operation</td>
<td>Windows Phone</td>
<td>Facebook SDK</td>
<td>Freemium</td>
<td></td>
<td>true</td>
</tr>
<tr>
<td>Facebook App</td>
<td>Facebook, Inc.</td>
<td>Social network operation</td>
<td>iOS</td>
<td>Facebook SDK</td>
<td>Freemium</td>
<td>1,100,000,000</td>
<td>true</td>
</tr>
<tr>
<td>Foursquare</td>
<td>Foursquare Labs</td>
<td>Social network operation</td>
<td>iOS</td>
<td>Facebook SDK</td>
<td>Freemium</td>
<td></td>
<td>true</td>
</tr>
<tr>
<td>Geocod</td>
<td>Enl</td>
<td>Location sharing</td>
<td>Android</td>
<td>Location SDK</td>
<td>Freemium</td>
<td></td>
<td>true</td>
</tr>
<tr>
<td>Google Latitude</td>
<td>Google Inc.</td>
<td>Location sharing</td>
<td>iOS</td>
<td>Google Latitude API</td>
<td>Freemium</td>
<td></td>
<td>true</td>
</tr>
<tr>
<td>Google Maps</td>
<td>Google Inc.</td>
<td>Navigation Services</td>
<td>Android</td>
<td>Freemium</td>
<td>Freemium</td>
<td></td>
<td>true</td>
</tr>
</tbody>
</table>

\[\text{RAND RR1482-2.17} \]

\[\text{Note that the investigator discovers that Instagram will release more information about photos than is currently indicated in MIKE. As will be described in the curation section, the investigator could easily add this information to MIKE to assist future users.} \]
In,” the other is “Radar.” Check-In is a GPS position collected whenever the user checks into a location. Radar collects GPS positions on a periodic basis, whether or not the user checks in.12

The investigator chooses Radar because it records locations at intervals and is not dependent on usage. MIKE indicates that it records locations every minute. By clicking on “Radar,” the investigator learns more about this data context and the rules for accessing it (Figure 2.19). This page informs the investigator that a 2703(d) court order is required, based on the user’s email address, to obtain location information for the week before the victim was found. Again, MIKE has helped the investigator narrow her request for information.

The time-stamped location data from Foursquare confirm that the individual stayed at the hostel. Furthermore, the final location record is from the alley in which the victim was found. The investigator postulates that the perpetrator turned off the phone immediately after killing the victim, thus establishing a precise time of death.

Next, she wishes to discover whom, if anyone, the victim was seeing at the approximate time of death. Since this phone is an Android phone known to Google Play (see Figure 2.11), Google Calendar is the obvious choice. Clicking on the Google Calendar link (in Figure 2.11)

---

12 The MIKE prototype requires the user to click on “Radar” and “Check-In” to learn about them. An implementation of MIKE should consider presenting the information about them as a table.
takes her to MIKE’s Google Calendar page, which shows that this app not only keeps calendar events, but may also have contacts for the events (Figure 2.20).

Information from Google in response to a second 2703(d) court order provides a Yahoo email address for an individual who was to meet the victim at a casino shortly before he was murdered. Given the tenuous tie to this email address, the investigator wonders what legal authorities might be utilized to request information from Yahoo. She is after basic subscriber information—name, phone number, address, etc.

She can answer this question by going to MIKE’s main page search bar and searching for Yahoo. She finds a page for the Yahoo corporate entity and notices that this entity has an associated data context called “User Information.” Because some corporate entities collect information that is shared across multiple mobile apps, MIKE allows the data contexts to be associated with corporate entities, as well as mobile apps. She clicks on this data context and is brought to a page describing the information Yahoo collects about its users. This page (Figure 2.21) reminds her that only an administrative subpoena is required to access basic subscriber information.

Yahoo provides the name (this time real) of the owner of the email account—an individual who stayed at the hostel at the same time as the victim. The hostel register provides a phone number. The individual (now a person of interest [POI]) is located out of state and denies meeting the victim the day of the murder. The phone number provides information about the POI’s carrier, which in turn provides unique phone identifiers.

Seeking more information, the investigator now turns to the victim’s financial transactions. The SquareCash app was also listed by Google Play (see Figure 2.11). Visiting the SquareCash/Transaction page of MIKE, the investigator finds she can obtain the victim’s
bank account number through SquareCash, allowing her to identify the victim's bank and contact it for additional information (Figure 2.22).

The data from the bank reveal that the victim had emailed funds to the POI’s email address on a regular schedule, stopping the month before the murder. However, SquareCash
refuses to provide details on the recipient’s (POI’s) account, claiming that there are no reasonable grounds to believe that the information is relevant because the victim had also sent money to several other individuals. The court concurs, arguing that a Google Calendar notation that is refuted by the POI is deemed to be insufficient justification.

The investigator now needs to establish that the POI did in fact meet with the victim by establishing the POI’s location at the time of the murder. However, she does not know what apps may have been on the POI’s phone. To pursue this line of attack, she returns to MIKE’s homepage and again uses MIKE’s query wizard to query for all mobile apps that collect GPS location.13 This query (Figure 2.23) turns up Geoloqi, which apparently produces third-party libraries for multiple apps and devices. The investigator has recently read that many app developers use Geoloqi to implement location-related functions and decides there is a good chance that the POI has an app on their phone that might use Geoloqi.

By clicking on “Geoloqi,” MIKE shows that there are multiple ways to identify an account on Geoloqi, including MAC addresses, Unique Device Identifiers (provided by the POI’s phone carrier), and email address (Figure 2.24).

The investigator now believes that, if she crafts a very narrow request, she may be able to convince a judge to issue the required order to Geoloqi. She therefore asks Geoloqi to determine whether the POI’s phone was in the alley within 30 minutes of the murder. No other data are requested. Geoloqi confirms that the POI was at the precise location where the victim was found, at the precise time of death. Based on this information, she then requests an order to obtain the full content of the POI’s email. The email that is obtained indicates that the POI threatened the victim if he did not resume payments. An arrest follows.

Figure 2.22
SquareCash/Transaction Page

13 For a discussion of how to do this, see text accompanying Figure 2.21.
Summary

This highly contrived example illustrates how MIKE can help a law enforcement officer utilize the data resident in the mobile ecosystem to pursue an investigation. The investigator still follows time-tested strategies—follow the people, follow the relationships, follow the money—but uses new techniques that are based on a detailed description of the mobile ecosystem and surveillance law to identify the alternatives for doing so. The additional step required here, that of linking the digital persona to a physical persona, was accomplished early on through photographs. When one investigative path was blocked (the banks would not provide the requested account information), MIKE helped locate another route (the third-party geolocation library). In this case, the investigator was familiar with the basic strategies to be used in a “digital investigation,” but as this example shows, MIKE could clearly be used for training purposes.

This case unfolded as a serial investigation with scraps of data building up over time, and MIKE helped the investigator narrow her requests for information consistent with what was known at each step of the investigation. This not only protects privacy but makes the entire process simpler and more efficient for all parties involved—the courts, law enforcement, and the commercial enterprises.

All the data MIKE used in this illustration, and the illustrations that follow, were derived from previous analyses (independent of any particular investigation) of the commercial enti-
ties and the data they collect and were made available to the investigator in this hypothetical. As others identify interesting apps and services and analyze the data they collect and the means of accessing the data, the alternatives presented to an investigator become richer and more likely to include apps associated with POIs in a criminal investigation. This is possible because MIKE provides a framework that allows investigators, and others, to incrementally share information with one another. Today, this is typically done through email distribution lists (or Search.org) and is limited to basic information about commercial points of contact for law enforcement investigators.

An Awareness and Training Example
The illustrative use of MIKE in an investigation suggests that it can also be used for raising awareness and training. Law enforcement professionals who had a chance to use the MIKE prototype pointed this out to us.

Generally, unless a law enforcement professional specializes in electronic surveillance, he or she is likely to have only a superficial awareness of how mobile devices, apps, and services are implemented and of the plethora of commercial entities that implement them. In thinking about electronic surveillance, the initial impulse is to do what is familiar: seek the needed information from a mobile carrier. Our illustration shows that, with information about the mobile ecosystem, richer investigative strategies present themselves.

MIKE can help raise awareness. The tool can be queried to display all the commercial entities that collect the types of data of potential interest to an investigator. Each of these commercial entities will have a page that describes the entity, the resolution at which it collects the type of data, and length of time it retains the data. The page also shows relationships between companies, e.g., ownership, which can be used to grasp how commercial entities may share

![Image of MIKE data context]

- **Email Address**
- **UDID and MAC address**

**Figure 2.24**
*Information About Geoloqi*

<table>
<thead>
<tr>
<th>Data Contexts</th>
<th>[edit]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context Name</td>
<td>Transmitted To</td>
</tr>
<tr>
<td>Account Information</td>
<td>Esri</td>
</tr>
<tr>
<td>Device ID</td>
<td>Esri</td>
</tr>
<tr>
<td>Location</td>
<td>Esri</td>
</tr>
</tbody>
</table>

**References**

- **Category**: Mobile App
information. This can help an investigator understand that any given type of data is collected by the many commercial entities implementing devices, apps, and services, not just mobile carriers, and develop a sense of the nature of these commercial entities.

Investigators must also understand the law and jurisprudence that define how commercial data can be obtained. Responses to requests from those who lack a good grasp of this material may be delayed because the commercial entity or the courts find them to be overly broad or improperly specified. Here, MIKE could be used in two ways for training purposes. First, trainees could be presented with a hypothetical problem, as in the illustrative case described earlier, and be asked to use MIKE to explore different ways to conduct the investigation using electronic surveillance. Second, trainees could also use MIKE to understand the legal requirements of the alternatives they can pursue, as well as the costs and known delays of specific entities fulfilling requests for surveillance data. This would help investigators make better-informed choices about how to conduct electronic surveillance when cost or timeliness of the response is a constraint.

Policy Use Case

This case demonstrates how policymakers could use MIKE to explore technical and legal issues related to the mobile app ecosystem and compare the regulatory approaches various jurisdictions have taken:

• Criminal investigations like the one described in the investigative example in the previous section have drawn the public’s attention to the information collected by and stored in the mobile app ecosystem. As a result, policymakers are considering passing a law to provide additional protections for location information.
• A legislative analyst uses MIKE to research policy issues related to location information. By clicking the link for “View all Policy Issues” on MIKE’s front page, she is brought to a page with a list of all policy issues currently captured by MIKE. From this list, she identifies the page entitled “legal protections for location information” as relevant to her concerns.
• This page displays statutes and case law relevant to location information. The analyst is able to use this page to investigate the various forms of legal protections available to location information.
• Once she has examined the law, the analyst then uses MIKE to investigate the technical issues related to location information, including the types of location information mobile apps can store and the types of mobile apps that might store this information.
• As a result, she is able to explore issues that she might not otherwise be aware of and better target her subsequent research.

Background

The recent murder of a tourist near a well-known casino and the subsequent police investigation have captured the notice of both the media and the public. In particular, law enforcement use of location data gathered from Foursquare has become the subject of much public scrutiny. While the media praised law enforcement’s use of innovative methods to identify the unknown victim, concerns have arisen about when and how location data may be used.
As both the public and the press become increasingly concerned about geolocation, a state legislator asks one of his staffers to learn more about geolocation and privacy, to figure out how to best address the issue. This staffer is unfamiliar with the privacy concerns arising from location data, but a major state newspaper published a multipart exposé on the types of location data mobile phones collect and how both public and private actors may use such information. The articles discuss the investigation described earlier at length, and the staffer notices that law enforcement was able to get location data in three different ways: geotagged photos, Four-square, and Geoloqi. She decides to use MIKE to learn more. Since she knows she will have to make a recommendation to the legislator, she decides to focus on three main concerns that any new law will have to address: what should be regulated, who should be regulated, and what legal protections the new law should create.

The Analysis

On entering MIKE, the analyst first notices the quick link on the homepage entitled “View All Policy Issues” (Figure 2.25). Since she knows she is investigating a particular policy concern, she clicks on this link. The front page of MIKE also contains quick links that allow the user to view all mobile apps or statutes within MIKE. If the staffer wanted, she could use these links to investigate specific statues or mobile apps to discover how they might relate to the policy issue she is researching.

This link takes her to the page shown in Figure 2.26. Since the analyst is interested in determining what legal protections should apply, she clicks on the link to the page entitled “legal protections for location information.” If the analyst does not see the policy issue she is looking for, she can add it to this page by clicking the link to the policy issue form at the bottom of the page. A newly added policy issue would be incorporated into the suggested policy issues that appear when a user is creating or updating legal information in MIKE.14 The user can then specify that a particular law implicates the policy issue, if appropriate. However, laws already within MIKE that implicate the policy issue do not automatically show up on the policy issue’s page unless they have been tagged appropriately.15

The link to the page entitled “legal protections for location information” leads to the page shown in Figure 2.27. The first table on the page summarizes the relevant statutory law. The second summarizes the relevant case law. The analyst uses this page to identify existing legal rules from various jurisdictions and take note of the data types each rule covers. She realizes that different types of information are governed in different ways, and makes a note of this for later.

The analyst first decides to look at the applicable federal law. Clicking on the ECPA link brings her to a page describing how ECPA regulates law enforcement access to basic subscriber information (Figure 2.28). She notes the type of information protected by ECPA in this instance (e.g., a street address) and realizes that this is different from the type of loca-

14 The ability to add a new idea or issue is an example of where crowdsourcing benefits from an editor or moderator. Stakeholders hold different interpretations of policy issues. Agreement will not always be possible, but having an editor or moderator will help ensure that the discussion is orderly and includes all relevant points of view. This person will need to be neutral and knowledgeable about policy.

15 It would certainly be possible to create a taxonomy that automatically connected legal information to policy issues based on certain characteristics. While the development of this feature is outside the scope of our current work, it would be an option for improving MIKE in the future.
Figure 2.25
Locating Policy Issues Addressed by MIKE

Welcome to MIKE

 MIKE: The Mobile Information and Knowledge Ecosystem.
 There are currently 49 mobile apps and 43 corporations.
 New to MIKE? Try visiting the Tutorials page.

Adding New Entries

Want to create a new page or edit an existing one? Enter a name and select a form to get started.

Quick Links

- View all Mobile Apps
- View all Statutes
- View all Policy Issues
- View all Corporations
- View all Activities
- View all Knowledge Types
- View all Products

Query Wizards

- Query All Data Contexts
- Query All Mobile Applications

Consult the User's Guide for information on using the wiki software.

Figure 2.26
Discovering Policy Issues

Category: Policy Issue

This category uses the form Policy Issue.

The following 8 pages are in this category, out of 8 total.

- 180 day rule
- Cell phone as location tracker
- Container doctrine
- Legal protections on location information
- Interception of communication
- Minimization of Intercepts
- Privacy in the workplace
- Policy issue of interest

Link to form to add new policy issue

Clicking on this link will display a list of all policy issues contained within MIKE
Figure 2.27
Legal Protections for Location Information

Legal protections on location information

<table>
<thead>
<tr>
<th>Jurisdiction</th>
<th>Citation</th>
<th>Data Type</th>
<th>Legal Document Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>18 USC Sec. 2703</td>
<td>Phone Number, Real Name, Mobile Network Identifiers, Street Address, Billing Information, Hardware Details</td>
<td>ECPA/Administrative subpoena</td>
</tr>
<tr>
<td>United States</td>
<td>18 USC Sec. 2703</td>
<td>App Name, Call Site Location, URL, GPS Location, Calendar Entries, To do list, Friends</td>
<td>ECPA/2703(c) court order</td>
</tr>
<tr>
<td>Texas State Jurisdiction</td>
<td>Texas Code Crim. Proc. art. 18.21 Sec. 14</td>
<td>GPS Location</td>
<td>Texas State Code of Criminal Procedure/Order for installation and use of mobile tracking device</td>
</tr>
</tbody>
</table>

Related Case Law  

<table>
<thead>
<tr>
<th>Jurisdiction</th>
<th>Court</th>
<th>Citation</th>
<th>Statutes Interpreted</th>
<th>Data Type</th>
<th>Law Enforcement Situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the Matter of the Application For an Order Directing a Provider of Electronic Communication Services to Disclose Records to the Government</td>
<td>Third Circuit Court of Appeals</td>
<td>626 F.3d 304</td>
<td>ECPA</td>
<td>Cell Site Location</td>
<td>Third party situation</td>
</tr>
<tr>
<td>People v. Weaver</td>
<td>New York State Jurisdiction</td>
<td>New York Court of Appeals</td>
<td>906 N.E.2d 1195</td>
<td>New York State Const. Art. 1 Sec. 12</td>
<td>GPS Location</td>
</tr>
<tr>
<td>US v. Jones</td>
<td>United States Jurisdiction</td>
<td>Supreme Court of the United States</td>
<td>132 S. Ct. 946</td>
<td>Federal Fourth Amendment</td>
<td>GPS Location</td>
</tr>
</tbody>
</table>

Reference

Category: Policy Issue

Administrative subpoena” links to an explanation of the appropriate legal document and the process for obtaining such a subpoena (Figure 2.29). She notices that this page includes links to two documents outside MIKE that describe the process in more detail and decides to access these documents for more information. She is particularly interested in the justification law enforcement must provide to access the information.
After conducting additional research on ECPA using traditional methods for legal research, the analyst decides to investigate how other jurisdictions have protected location information. She goes back to the policy issue page for location information (Figure 2.27), then clicks on the page for the Texas Code of Criminal Procedure (shown in Figure 2.30). She notices that this statute applies to law enforcement use of vehicle trackers, not location information from mobile phones, but decides that this information may still be helpful in determining what kind of protections have been provided to GPS location data.

To determine what kinds of protection this law provides, the analyst then clicks on the link in Figure 2.30 for the required legal document. This brings her to Figure 2.31, which explains the requirements for obtaining an order for installation of a mobile tracking device under Texas State Law. She is particularly interested in the relevant legal standard because she thinks policymakers may be interested in using similar legal standards to protect location information in her jurisdiction.

The analyst notices that GPS location provides much more information about an individual’s location than an address and that this information is correspondingly protected by a much stricter legal standard. She then returns to the policy issue page (Figure 2.26) and decides to look at recent case law providing protection for location information. She then discovers there is a recent Supreme Court case discussing legal protections for location information and decides to investigate further. This page (Figure 2.32) provides her with the citation.

---

16 It would be useful to specify the jurisdiction of interest. This could be used as a filter to focus the query—in this case, only one state of many is of interest to the analyst. This is a feature that an implementation of MIKE should consider.
for the case and information about the background and legal holding (in the "Contents" box). This helps her understand the significance of the case and provides the information necessary for her to investigate further if she wishes.

The analyst then decides that she should learn more about the technical issues related to location information. Using the search bar located at the top of every page, she searches for information about location. One result (Figure 2.33) seems to relate to location information in general, so she clicks on that link to learn more.
The page corresponding to this link is shown in Figure 2.34. That page lists the apps known to collect information location and provides a link to a page describing each app (Figure 2.35).

The location data type page lists all apps that gather some type of location information (Figure 2.35). Looking at this page, the analyst determines that the mobile app ecosystem can collect many different types of location information, including GPS location, Wi-Fi location, Internet Protocol (IP) address, street address, and even latitude and longitude. She is also surprised that some of the apps that collect location information seem to have no use for it—e.g., Angry Birds. The staffer then realizes that it might be necessary to regulate how private corporations, not just law enforcement, use information. For each app, she also investigates the types of encryption used, the retention length, and resolution.

The analyst recalls the article she read that described the criminal investigation in which law enforcement used three different ways to obtain location information (geotagged photos, Foursquare, and Geoloqi). Therefore, the analyst decides to look at these data sources in particular. She starts with Foursquare, because she recognizes it as an app she has on her own phone. After clicking on the link, she is brought to a page describing the Foursquare app and the information it collects (Figure 2.36).

She then realizes that this one app collects a host of information. Moreover, she notices that it does not treat all information the same way: While all the information is maintained for the life of the account, the resolution indicates that some of it is collected more frequently than others. Since MIKE provides her with a link to Foursquare’s privacy policy, she decides to click on it and save the page to return to later. She then notices that Foursquare collects GPS location in several different ways and decides to learn more about GPS location data. She clicks on the link and is directed back to the page containing a brief description of the information type

---

17 This article discussed the investigation described earlier, under “Investigative Use Case.”
and a list of mobile apps that collect or use this type of data (Figure 2.34). She can repeat this process for any type of information she thinks might be interesting from a policy perspective.

Once she has explored GPS location information, she decides to learn more about Geoloqi because she is unfamiliar with third-party libraries and wants to learn more. She types the name into the search bar at the top of the screen and is directed to the page in Figure 2.37.

This page provides a short explanation of what Geoloqi is and what it does. This confirms the analysts’ previous decision that it might be important to regulate the way that private entities are sharing information within the mobile app ecosystem. Furthermore, she realizes that
this regulation would need to consider not only mobile apps but also companies that provide third-party libraries, which are invisible to the consumer but handle individual information. The staffer clicks on the link to Geoloqi’s website to learn more. The staffer notices a new type of location information listed (cell site location) and decides to investigate further. Clicking on the appropriate link redirects her to the page in Figure 2.38.

From this description, she notices that cell site location data not only seem less specific in general but that the level of specificity of the data varies. However, Geoloqi is capable of providing fairly detailed information in certain circumstances. She also notices that many fewer apps appear to collect this information, suggesting that it might not be as widely used in the mobile app ecosystem. She decides to add this to the list of information types to protect. Having explored many of the issues related to legal protections of location data, the analyst decides to report her findings to the legislator and discuss possible next steps.
Summary

As this example has shown, MIKE can greatly assist policy researchers and analysts in examining issues related to the mobile app ecosystem. One of MIKE’s greatest strengths is as a tool for exploratory analysis of policy issues; it enables policy researchers to discover interactively how legal rules and technical issues relate to each other to shape policy issues. It also allows researchers to identify previously “unknown unknowns”: considerations that must be addressed for effective decisionmaking but that are not apparent from the policy problem. Because MIKE can provide extensive citations and links to supplemental material, it is an ideal starting point for policy research. MIKE allows users to engage traditional research tools, such as LexisNexis and Google Scholar, more efficiently and effectively. An analyst can also add references to other sources to MIKE as she discovers them in her research.

In the example above, the analyst used MIKE to answer three questions needed to craft legislation providing protections for location information: What kind of location information should be protected? Who should be regulated? What legal standard should apply? She was
able to explore the wide variety of location information collected and shared within the mobile app ecosystem. She also realized that the applicable legal standards other policymakers had developed seem to vary depending on the type of location information collected: The more specific the information collected, the more protections applied. Finally, although she started with the intention of creating policy to regulate law enforcement’s use of this data, she also realized that it might also be necessary to regulate the way information is shared among actors in the mobile app ecosystem. She is therefore in a much better position to make a recommendation about potential policy solutions.

If the content of MIKE is developed through crowdsourcing, the tool could be used to explore uncertainty and differing opinions in the law. For example, if a privacy advocate adds case law suggesting a broader interpretation of the Fourth Amendment, an assistant district attorney could be motivated to add information about the case that suggests it should only be applied in a limited fashion or could add new case law suggesting an alternative interpretation. In the future, it might be possible to develop MIKE’s infrastructure to accommodate further discussion of legal conflict. The resulting conversations could be useful to policy analysts who want to understand the ambiguities in current law.

One can also envision using MIKE for a more-quantitative analysis of policy questions. For example, consider a policy proposal to eliminate the 180-day rule while imposing a minimum retention period for commercially collected data. An obvious question would be to ask how many commercial firms would this change affect. MIKE could be used to find all the commercial firms housing data subject to the 180-day rule and to examine their data retention policies. If MIKE were well populated, it would be possible to the estimate the number of firms that would be affected by a particular retention rule.
### Figure 2.35
#### Location Category

Data about somebody or something's physical position. Examples include GPS Location data, home addresses, home zip code, and current metropolitan area. Location can be precise or imprecise in both geographic and temporal dimensions.

<table>
<thead>
<tr>
<th>Data Context Name</th>
<th>Data Type</th>
<th>Encryption</th>
<th>Resolution</th>
<th>Retention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angry Birds</td>
<td>GPS Location, Wi-Fi Location, IP Address</td>
<td>Unspecified</td>
<td>Unspecified</td>
<td>Unspecified</td>
</tr>
<tr>
<td>Badoo</td>
<td>GPS Location, Photo, Video</td>
<td>Usage-based</td>
<td>Life of account</td>
<td></td>
</tr>
<tr>
<td>Badoo</td>
<td>GPS Location, Photo, Video</td>
<td>Usage-based</td>
<td>Life of account</td>
<td></td>
</tr>
<tr>
<td>Brightest LED Flashlight</td>
<td>GPS Location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dropbox</td>
<td>Dropbox/Account Information, Real Name, Email Address, Street Address, Phone Number, Group Membership, Office Holder, Latitude and Longitude, Street Address</td>
<td>Entity level encryption, No encryption</td>
<td>Usage-based, No encryption</td>
<td>Life of account, Perpetual</td>
</tr>
<tr>
<td>Dropbox</td>
<td>Dropbox/Files</td>
<td>Entity level encryption, No encryption</td>
<td>Usage-based</td>
<td>Life of account</td>
</tr>
<tr>
<td>Facebook App</td>
<td>Facebook App/CheckIn, Post Friends, Group Membership, Office Holder, Latitude and Longitude, Street Address</td>
<td>No encryption, Entity level encryption</td>
<td>Usage-based</td>
<td>Perpetual</td>
</tr>
<tr>
<td>Facebook App</td>
<td>Facebook App/Events</td>
<td>No encryption</td>
<td>Usage-based</td>
<td>Perpetual</td>
</tr>
<tr>
<td>Facebook App</td>
<td>Facebook App/Groups</td>
<td>No encryption</td>
<td>Static</td>
<td>Perpetual</td>
</tr>
<tr>
<td>Facebook App</td>
<td>Facebook App/Photos</td>
<td>No encryption</td>
<td>Usage-based</td>
<td>Perpetual</td>
</tr>
<tr>
<td>Facebook App</td>
<td>Facebook Poll/Photo, GPS Location</td>
<td>No encryption</td>
<td>Usage-based</td>
<td>Perpetual</td>
</tr>
<tr>
<td>Facebook Poll</td>
<td>Facebook Poll/Video</td>
<td>No encryption</td>
<td>Usage-based</td>
<td>Perpetual</td>
</tr>
<tr>
<td>Fake Corporation</td>
<td>Fake Corporation/GPS Location</td>
<td>No encryption</td>
<td>Usage-based</td>
<td>Perpetual</td>
</tr>
<tr>
<td>Find My Friends</td>
<td>Find My Friends/Location History, GPS Location, Timestamp</td>
<td>No encryption</td>
<td>Usage-based</td>
<td>Perpetual</td>
</tr>
<tr>
<td>Foursquare</td>
<td>Foursquare/Check-In, GPS Location, Photo</td>
<td>Usage-based</td>
<td>Life of account</td>
<td></td>
</tr>
<tr>
<td>Foursquare</td>
<td>Foursquare/Photos</td>
<td>Usage-based</td>
<td>Life of account</td>
<td></td>
</tr>
<tr>
<td>Foursquare</td>
<td>Foursquare/Radar</td>
<td>Usage-based</td>
<td>Life of account</td>
<td></td>
</tr>
<tr>
<td>Foursquare</td>
<td>Foursquare/Specials</td>
<td>Usage-based</td>
<td>Life of account</td>
<td></td>
</tr>
<tr>
<td>Geolocli</td>
<td>Geolocli/Account Information</td>
<td>No encryption</td>
<td>Usage-based</td>
<td>Unspecified</td>
</tr>
<tr>
<td>Geolocli</td>
<td>Geolocli/Latitude, Longitude, Cell Site Location</td>
<td>No encryption</td>
<td>Life of account</td>
<td></td>
</tr>
</tbody>
</table>

*This category has the default form: Data Type*
Commercial Use Case

This example is specifically constructed to illustrate how a commercial entity might use MIKE to respond most effectively to law enforcement requests while respecting the privacy interests of its customers.

This example is based on a hypothetical law enforcement investigation and a fictional, small app developer’s use of MIKE to determine how best to respond when facing a request for assistance in an investigation. The example demonstrates how MIKE might be used to determine the commercial firm’s options, based on law and current practices within the community. The example features exploration of the following:

- existing practices for cost recovery
- existing corporate policies for proving basic information, metadata, and content, including law enforcement request templates
- client notification practices.
The example is highly stylized to illustrate a selection of MIKE’s features and does not presume that the indicated pathways are the only (or even the best) way to obtain the data the firm’s attorney is seeking in responding to the request.

Background
LG is the chief (and only attorney) working for a startup called Protected Gardens, which has developed an app called WaldGardn. She has been hired principally to manage the company’s patent portfolio, although she did craft the Protected Gardens privacy policy. WaldGardn
creates secure communities of interest. Within a “Gardn,” all participants freely share information—emails, text messages, pictures, location information, and interests. There is even a payment exchange system so members can easily provide loans and transfer funds to each other. WaldGardn’s privacy statement promises that the information is fully encrypted and that none of a Gardn’s information is available to anyone outside it, with the standard caveats for responding to law enforcement requests. Currently, the app has 2 million users and 10,000 Gardns. Its business model is based on providing advertisers highly characterized clusters of k-anonymized customers with common interests and means.¹⁸

The state police have information that one or more Gardns are being used to facilitate the operations of a large interstate cargo theft ring. Law enforcement officers use MIKE to find a summary page describing WaldGardn (Figure 2.39), which tells them that WaldGardn would in fact meet all the criminal organization’s needs. By looking at the data context pages, officers also use MIKE to determine that the data are encrypted on WaldGardn’s servers. Additionally, the officers use MIKE to research the Protected Gardens corporation, finding that LG is listed as its chief counsel.¹⁹

LG receives a call from the director of the state’s cargo theft task force informing her that WaldGardn will shortly be receiving a court order requesting that WaldGardn perform a search of “all” its records to determine whether there is a strong correlation between the locations of members of one or more Gardns and the times and places of a dozen hijackings over the past 18 months. If such a correlation is found, the task force will request “all available information” about members of the specific Gardn(s). The investigator wishes to know how LG would best like to receive the initial request.

LG first talks to her chief technical officer, who informs her that, in principle, the search could be done; identifying common characteristics of anonymous users is a core competency of the company. However, it will take a team of two of the company’s best engineers a week or two to construct the necessary software—this at a time when they are preparing for a major software release.²⁰ It might be able to outsource the task if funds were made available. The chief executive officer (CEO) tells LG he wishes to cooperate with law enforcement—to an extent—but wants to protect his subscribers, WaldGardn’s reputation, and his milestones. He asks if they can limit the scope of the request and be reimbursed for the work required. He also wants to know what he can tell his customers.

**The Analysis**

LG assumes that the guidance from her CEO means that staff members are to help if they can (i.e., do not fight the order) but not to bankrupt the company or harm innocent customers. After a suggestion from a colleague, she turns to MIKE (and is immediately intrigued to discover that MIKE already knows about WaldGardn) to investigate how she can answer the investigator and still meet her company’s needs. She starts by attempting to determine how

---

¹⁸ In k-anonymization, the attributes of each user are suppressed or generalized until the user appears identical to at least k – 1 other users. At this point, the database is said to be k-anonymous. MIKE could include such terms as k-anonymized within a notes section for WaldGardn, which would then appear if another user needed to research the concept.

¹⁹ We described how to use MIKE to find information about encryption and points of contact earlier, under “Investigative Use Case.”

²⁰ We recognize that the issue of law enforcement ordering a corporation to modify its software is controversial. This is a hypothetical example that assumes the company is willing to be responsive to law enforcement’s request for help.
other, larger, and more mature companies are typically reimbursed for the effort required to respond to law enforcement. At MIKE’s main screen, she types “cost” into the search bar. Several relevant entries show up (Figure 2.40).

Clicking on Facebook and Yahoo (see Figures 2.41 and 2.42), LG quickly determines that companies typically charge by the number of accounts for which information is requested, by the length of time for which information is requested, by the number of locations searched (e.g., cell sites), or by the actual time spent retrieving the information.

Given the broad nature of the request (search all 10,000 Gardns) and a not unreasonable price of $30 per Gardn, the total cost would have been $300,000. But LG realizes that this would not be acceptable to law enforcement and therefore takes her technical officer’s estimate of about 1/2 person-month to conduct the search, for a cost of $15,000. She rules out turning the entire database (and encryption keys) over to law enforcement because that would be a betrayal of the customers’ trust. She makes a note to inform her CEO that the company might contest the original request as being too broad. Because the original request will reveal only the names of Gardns and no customer specific information, LG feels that the company would
The state investigator swallows hard, agrees to pay WaldGardn the $15,000, and two weeks later is rewarded with the names of two Gardns whose (as yet unidentified) members’ locations correlate highly with the times and places of several hijackings. LG, anticipating the next step, turns to MIKE to determine what types of legal documents she will require if she is to respond to the inevitable next request. She assumes that the cargo theft task force will want the following:

1. the names of Gardn members and their locations at the times and places of the past hijackings
2. email content
3. the content of texts (Short Message Service [SMS])
4. contact lists
5. financial transactions
6. potentially, real-time texts and locations.

Her next step is another call to the chief technical officer to find out exactly what the technical limitations are on the information WaldGardn can provide. She discovers the following:

Figure 2.40
Results of MIKE Search for “Cost”

<table>
<thead>
<tr>
<th>Search results</th>
<th>“cost” entered into search bar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>Facebook appears in search results</td>
</tr>
<tr>
<td>Verizon</td>
<td>Verizon appears in search results</td>
</tr>
<tr>
<td>WhatsApp</td>
<td></td>
</tr>
<tr>
<td>Pen register</td>
<td></td>
</tr>
<tr>
<td>Instagram, LLC</td>
<td></td>
</tr>
<tr>
<td>Yahoo</td>
<td>Yahoo appears in search results</td>
</tr>
</tbody>
</table>
• The location is encrypted, but WaldGardn holds the keys and saves the information for two years.
• The email is encrypted, but WaldGardn holds the keys and saves the emails for two years.
• The SMS is user end-to-end encrypted and not available to WaldGardn.

Financial data are limited to account IDs, financial institution, credit card numbers, and records of all transactions among the Gardns’ users. Passwords for financial institutions passwords are user encrypted and not available.

Facing the potential task of researching the legal requirements for each of these categories separately (or engaging outside counsel at considerable expense), LG decides to see whether MIKE can point her to policies, procedures, and guidelines other corporations (with larger legal teams) use to assist with this research. She hopes that this will also help her answer her CEO’s questions about notification. She notes that MIKE has automatically provided suggestions regarding the legal documents that may be required for law enforcement to obtain messages from WaldGardn (see Figure 2.43).

LG also decides to return to the corporation pages that initially provided cost information to see what they say about the legal requirements for obtaining information. She notices various references to law enforcement guidelines in the pages’ reference sections (see Figure 2.41).
She enters “guidelines” into MIKE’s search bar and is rewarded with several corporate guidelines for law enforcement (Figure 2.44).

In her own legal research, LG uses policies gleaned from guidelines other corporations have issued. After using both MIKE and traditional legal resources to analyze the legal issues, she decides to tell the task force that basic subscriber data will be made available under a subpoena, at $30 per WaldGardn. In one set of guidelines, she has found an appropriate “group” template, which, with minor modifications, can be adapted to WaldGardn. She then informs the investigators of the following:

1. A 2703(d) court order is required for specific customer metadata, including contacts, at $5 per customer.
2. Given both the statutes and case law regarding law enforcement access to content information, a search warrant is required for all content, whether 180 days old or not—$30 per subscriber.
3. A search warrant is required for location information—$30 per subscriber for each 60-day period.
4. SMS content is end-to-end encrypted; WaldGardn does not hold encryption keys and will not aid in decryption.
5. Real-time data will require a wiretap order.
She also tentatively decides that the only available financial information is transactional and will therefore also require a 2703(d) court order. However, she makes a note to use MIKE to find other apps (and relevant contact information) that might acquire financial information and contact their legal staffs to further discuss the issue.

She sends her requirements to the investigator preemptively, hoping that the costs, although reasonable, will deter overly broad requests. In keeping with policies she has read, she also informs the investigators that, unless she receives court orders not to disclose the request to her customers, she will notify all members of the specified Gardns of the requests from the task force. The task force director is happy to receive detailed guidance from WaldGardn and, in fact, adds to MIKE what he has learned about dealing with the organization. MIKE’s curators confirm the accuracy of the new data.21

Feeling she has this all under control, LG returns to the important patent work. She resolves that, at a future date, she will revisit WaldGardn’s privacy policy and compose her own guidelines that can then be incorporated into WaldGardn’s data in MIKE.

---

21 This guide does not cover the options for operating MIKE in any depth. Here, we have assumed the role of a curator in the operation of MIKE and have assumed that the curator can verify the accuracy of new information.
Summary

This example has illustrated how MIKE can support the commercial sector of the mobile ecosystem by providing guidance for how to respond to court orders and reducing the time devoted to serving law enforcement requests. MIKE can expedite a request for information relevant to an investigation even when law enforcement knows (or strongly suspects) where the needed data are located. A corporate attorney about to receive her first court order and unfamiliar with Fourth Amendment legal processes is quickly able to determine best practices within the industry and adapt them to her own situation.

MIKE was able to provide guidance on what would constitute an appropriate request for cost recovery and thereby minimize the economic impact on LG’s small start-up. It also suggested which type of order was most appropriate for each category of information that her company possessed. References to guidelines that other members of the mobile ecosystem had produced were easy to discover and use to provide further guidance. As a result, the company attorney and the law enforcement organization were able to rapidly reach an agreement on what data could be provided, what court orders would be needed to justify releasing the information, and how much it all would cost. A law enforcement organization was able to avoid delays by formatting its requests to specific guidelines from the company, thereby avoiding multiple time-consuming exchanges.
Finally, both organizations were able to update MIKE easily, so that future requests on WaldGardn from law enforcement organizations or to other WaldGardn attorneys can be handled efficiently.
CHAPTER THREE
Curator’s Guide to MIKE

The technology, corporate entities, and laws involved in the mobile app ecosystem change rapidly. Consequently, it is important to keep the information in MIKE up to date. MIKE was designed to simplify the process of adding and updating information. Information can be added either by curators specifically tasked with keeping MIKE current or by regular users who find new data about the mobile app ecosystem.

This chapter describes how to maintain and update the information in MIKE. We first explain how to add information to MIKE and to change existing information. We also briefly mention how to alter the structure of MIKE—for example, by adjusting the entries for mobile apps to create a placeholder for a new type of information. Then, we review methods of locating information for inclusion in MIKE.

While this chapter is designed to stand alone, we assume that the reader is familiar with what MIKE is and how to use it. For more information these topics, see Chapter Two.

Curating and Updating MIKE

Entering Data and Maintaining the Map
MIKE is flexible in how data can be entered. A user can focus on entering legal statutes, cases, apps, data types, firms, or some other category of information previously discussed. In most cases, forms have been set up to help the user enter the information correctly.\(^1\) If a user tries to create a relationship to a page that does not yet exist, MIKE shows it as a “red link,” which indicates to the user that the page for the concept has yet to be created. This behavior is inherited from the MediaWiki framework (MediaWiki, 2014). Clicking on the link will take the user to the form used to populate the necessary information. Note that, even though links are red, they can still be used like any other link. Searches will still find them, and the same link created elsewhere will link to the correct page once it has been created.

While information can be entered in any order, it is typically easiest to start with the highest-level entity and work down. This means creating a page for an app’s owner before creating one for the app and creating an app’s page before defining pages for the data it captures. For instance, to enter data for a new mobile app in MIKE, first search for the parent firm in MIKE. If it does not exist, create that page first. On the firm’s page, there will be an “Add Mobile App” button that can be used to create the page for the mobile app with default

\(^1\) When MIKE does not currently have a form to help the user add specific information, it would be possible for someone with knowledge of Semantic MediaWiki to add new forms.
information. This ensures that the firm’s name is used correctly in the mobile app page. Once the app is created, the data contexts can be filled in, with any new data types appearing as red links. Any additional apps the firm owns can be created in the same way by going back to the firm’s page in MIKE and using the “Add Mobile App” button.

Several features of the forms used to create new entries in MIKE may be helpful when adding information (see Figure 3.1). First, there is a blue circle with a question mark next to most items in the form; hovering the mouse over these blue circles causes a pop-up box to appear with additional information about the item. This box is meant to provide guidance for individuals who are unsure about what content is appropriate for each item.

Second, for MIKE to properly make connections between different types of entries, it is important for terminology to be consistent. To simplify this effort, the forms used to create new entries in MIKE can generate lists of suggested inputs that are already in use. As Figure 3.2 shows, when the user starts typing, a list of previously used inputs will appear in a drop-down

Figure 3.1
Obtaining Information About Entries on Forms

Edit Corporation: Facebook, Inc.

General Information

Facebook Inc. runs the popular online social networking service Facebook. Facebook Inc. was founded in February 2004 by Mark Zuckerberg while at Harvard University. Initially, the website’s membership was limited to Harvard students, but was expanded gradually over the years to include other colleges and eventually to everyone (over 13 years of age).

Law Enforcement Guidelines Facebook: https://acobmoth.x.fbdn.net/hphotos-xf1/1/t39.2178-4/851585_54424278309435_854724444_n.pdf
Law Enforcement Guidelines Facebook: https://www.facebook.com/safety/groups/law/guidelines/

Website: http://www.facebook.com/

Law Enforcement Email:

Law Enforcement Email:

Activities: 

Subsidiary Of: 

Response Time: 

Response Cost:

- We may seek reimbursement for costs in resp

Business Addresses

Add Another Address

Data Contexts

Add as many “Data Contexts” below as you need. You only need to specify the name, which can be anything, but should probably reflect the underlying data type. For instance, you could call one “ZIP Code Data”. After you create this page for the Corporation, you will then have the opportunity to specify the details for each Data Context.

Data Contexts

Add Data Context

Summary:

This is a minor edit
Watch this page
Save page  Show preview  Show changes  Cancel

RAND RR1482-3.1
The user can either select an input that corresponds to the information he or she is trying to add or continue typing to create a new input. If the user does not know how to describe the information, pressing the down arrow key will bring up a list of all previously used entries.

Additionally, every data entry form has a section called “General Information.” This section can contain any information the user wants and uses the MediaWiki formatting tools. In addition, references can be specified using  to create properly formatted links that will appear as superscripts that link to nicely formatted hyperlinks at the bottom of the page.

Finally, since MIKE is built on top of MediaWiki, all changes are logged and can easily be rolled back, so mistakes are easy to correct, and information is never overwritten or lost. MIKE is very robust, and it is difficult to undermine MIKE’s basic functionality by changing

---

2 This list is populated based on all members of the category corresponding to the input. For example, the drop-down list shown in Figure 3.2 contains all members of the category “Activities.” New entries added to a category will also appear on the drop-down list. Not all blanks have drop-down lists, but it is possible to add them. However, creating new drop-downs requires some knowledge of Semantic MediaWiki.
the information within MIKE. However, if such damage were to occur, it would be possible to use Semantic MediaWiki’s built-in functionality to revert to the last functioning version of MIKE. Edits are attributed to the user making them.

The rollback function is necessary but not sufficient for quality assurance. An implementation of MIKE will require concepts of operation. One of these concepts is the entry, refinement, and vetting of content. Entering and vetting content might involve crowdsourcing, for instance. This could, however, lead to “editing wars” of the kind Wikipedia has experienced. To mitigate this, an implementation might include a moderator or editor, who would be responsible for adjudicating differences of opinion and vetting the content. Alternatively, an implementation could rely on specified curators tasked with entering and vetting data. Both the monitor or editor and the curator would need to appear to the stakeholders using the implementation as impartial.

MIKE anticipates and is intended to capture differences of opinion about how to interpret ambiguous situations. The law is particularly problematic in this regard. Law enforcement professions and privacy advocates often disagree about policies and can cite different court decisions to support their points of view. It is important to capture these differences in interpretation.

We recognize that an effective implementation of MIKE needs to maintain the highest quality of information. Accuracy is paramount, but completeness is also important. MIKE is intended to be a body of reference material supporting law enforcement, privacy advocate, and commercial firms alike. We believe it should not provide legal advice or function as a case-management system. Legal advice should provide an additional quality assurance check on the use of information in an implementation of MIKE before using the information to guide a surveillance effort or for a prosecution.

Other important functional requirements for an implementation of MIKE include assuring access to and the confidentiality and integrity of MIKE’s data, portion controls based on user roles, user vetting and access control, timely access to a help desk, trust in the sources and in the vetting of MIKE content, and confidence in the ownership and governance of an implementation. Note that these are both organizational and functional requirements.

We will first describe the process of entering an app. However, in practice, it is usually easier to start with the app’s owner.

**Describing an App in MIKE**

Figure 3.3 shows the form used to create and update mobile apps in MIKE.

To update an existing mobile app, the user simply clicks on the “Edit with form” tab at the top of the app’s page. Forms for new mobile apps can be generated using a link prominently displayed on MIKE’s homepage.

Once the editing form has been opened or created, there are places to enter both structured and unstructured information. Unstructured information can be added via a box called “General Information,” which allows free-form text entry. Here, users can enter any information they think is necessary, including links to other sources.

---

3 *Structured* information can be thought of as the answer to a specific question, provided in a predetermined form. *Unstructured* information can be thought of as any discussion the user wishes to contribute, provided as free-form text. The unstructured information can include hyperlinks to websites inside or outside MIKE. The use of drop-down menus helps standardize inputs but limits the granularity of entries. An implementation of MIKE should carefully consider this trade-off.
Figure 3.3
Updating the Facebook Mobile App

Edit Mobile App: Facebook App

General Information

Facebook is the world's most popular social networking site. The amount of data it collects is huge. The data contexts listed below are not all-inclusive, but they are a good sample of the types of data Facebook collects.

Facebook's new Graph Search allows extensive searching of their social network across a wide variety of personal and public information.

Data Contexts

Add as many "Data Contexts" below as you need. You only need to specify the name, which can be anything, but should probably reflect the underlying data type. For instance, you could call one "ZIP Code Data." After you create this page for the Mobile Application, you will then have the opportunity to specify the details for each Data Context.

Data Context Name: Friends
Data Context Name: Likes
Data Context Name: Checkins
Data Context Name: Comments
Data Context Name: Messages
Data Context Name: Events
Data Context Name: Accounts
Data Context Name: Groups
Data Context Name: Logins
Data Context Name: Photos
Data Context Name: Shares
Data Context Name: Searches
Data Context Name: User Apps

Summary:

- This is a minor edit
- Watch this page
- Save page
- Show changes
- Cancel

Structured Information

Edit with Form Tab

Unstructured Information

Click button to remove data contexts

Click double-sided arrows to reorder data contexts
Structured information is added via a series of labeled boxes just below the “General Information” box. Many of these are self-explanatory, with places to enter the name of the owner, the operating systems the app supports, links to the website and privacy policy, an estimate of the installed user base, and any revenue sources. “Libraries” is for tracking what third-party libraries an app includes in its installation. Third-party libraries provide functionality and will often collect information. For example, Dropbox has a library that helps developers integrate Dropbox’s file sharing and syncing services into their apps.

Below the structured information is a place to list the data contexts associated with the app. Recall that a data context includes the data captured by an app and other important information, such as the presence of encryption, how often the data are saved, and where they are saved. The user can associate a new data context with a mobile app by adding an entry to this list. Data contexts can be deleted by clicking the “Remove” button to the right of the data context name and can be reordered using the double-headed arrows on the far right. After a new data context has been created and the changes saved, the data context name will appear as a red link in the mobile app’s page. Clicking on this red link brings users to a page where they can enter information about the data context, as described in the next subsection.

Finally, at the bottom of the form are buttons that can be used to save, preview, or cancel the user’s edits. Because MIKE does not save information entered in forms automatically, the user must save changes manually. If the user is not sure whether he or she has added information correctly, the preview function can be used to view changes before they are saved.

These forms are a shortcut for adding information to MIKE. A user wanting to add information that these forms do not address can use the edit tab to further customize the mobile app’s page, as shown in Figure 3.4. However, any information added in this way must be formatted correctly using Semantic MediaWiki syntax. Consequently, we do not recommend adding information this way unless the user is already familiar with this syntax. Again, changes to these pages will not be saved automatically, and so users must use the buttons at the bottom of the page to save their work.

**Describing a Data Context**

Figure 3.5 shows a form for updating the Facebook App/Accounts data context in MIKE. Data contexts are always associated with a particular mobile app. Data contexts are therefore named using two pieces of information separated by a slash mark: the name of the app (in the example, “Facebook App”) and the name of the data the app collects (“Accounts”). The form to update a data context is accessed in different ways, depending on whether the user is adding a new data context or updating an existing one. A user wanting to add a new data context must do so on the associated mobile app’s page (as described earlier). Once added, that data context will be presented as a red link on the mobile app’s page in MIKE; clicking on that red link will bring the user to the form shown in Figure 3.5. A user wanting to make changes to an existing data context can access the form shown below by clicking on the “Edit with form” tab on the data context’s page.

The form used to edit a data context provides places to add both unstructured information and several types of structured information:

- **Transmitted to** is for listing the organizations that receive the data associated with the data context. In this case, it is only Facebook, but multiple organizations can receive the same data context.
Figure 3.4
Updating the Facebook Mobile App Without Using a Form

![Image of Facebook Mobile App editing page]

**Edit Tab**

**Unformatted Semantic MediaWiki Information**

---

Figure 3.5
Updating the Facebook Apps/Accounts Data Context

![Image of Facebook Apps/Accounts data context editing form]

**Edit with Form Tab**

**Data Context Name**

**Unstructured Information**

**Structured Information**

**Datatype**

**Drop Down Menu**

**URLs**

---

RAND RR1482-3.5
• **Datatype** allows the user to specify the type (or types) of data this data context describes. In theory, a new data context could be entered for every single data type. However, because that would be tedious, MIKE allows users to specify multiple data types for a single data context. The rules for how to group data types into data contexts are not hard and fast. The most important concern is that the other information in the context apply equally to each data type listed. For instance, it would not be appropriate to put two data types in the same context that had different types of encryption applied to them (rather, it would be important to create two separate data contexts, each pertaining to data types with particular levels of encryption). Also, as we will discuss later, MIKE bases tables providing legal information based partly on the type of data described by the data context. Therefore, the user should consider creating a new data context if he or she knows that the data types collected are protected by different legal requirements. A user who later discovers that two datatypes in the same legal context are protected by different legal provisions can simply create a new data context for one of the data types.

• **Storage Mode** is for describing how the data are physically stored. Current options include local (for data saved only on the device) and remote (for data stored on a server somewhere). Future options may include cloud storage or even specific cloud storage services.

• **Encryption** is for describing how the data are encrypted. If no information is available on whether the data are encrypted, this row should be left blank. A user who knows what kind of encryption is used but is unsure how to describe it can press the down arrow to create a drop-down menu of previously used options.

• **Resolution** allows the user to select from several options to describe how frequently the data are saved. Note that, unlike with previous rows, the user must here select from different options on a drop-down menu rather than enter text. The options are
  – **Usage based** is the most likely option and simply means that the data are saved when the user performs some action. Uploading a photo is one example.
  – **Static** means the data are saved once and only infrequently updated. An account name would be one example.
  – **Most recent use** means that all data are discarded except the most recent value. An example would be a location-tracking app that saves only the current location.
  – **At intervals** means that the data are polled at fixed intervals while the app is active. New resolution types can be added as technology changes.
  – There is also a blank option, which implies that the resolution is not known.

• **Retention** is for specifying the deletion policy for the data context. The user is presented with several options on the drop-down menu:
  – **Life of account** means that the data are kept for some fixed amount of time after the account becomes inactive or until the user asks for it to be deleted.
  – **Fixed time span** describes data that are kept for some fixed amount of time and then deleted.
  – **Perpetual** describes data that have no deletion policy—the organization has no plans to ever delete the data.

---

4 It might be useful to quantify retention policies, whenever possible. This will not always be possible, e.g., a retention policy of "life of the account" is not generally quantifiable. For prototyping purposes, we focused on qualitatively different policies. An implementation of MIKE should consider including quantitative information when it is available.
– **Transient** is for data that only reside a very short while, such as a message waiting for a user to come online before being deleted after delivery.
– **Blank** is for when the user does not know the retention policies.

• **References** allows the user to provide URLs to webpages providing additional information about the data context.

As always, a user who wants clarification while updating the data context can hover over the blue circle with the question mark to open a pop-up box with additional information. The user must manually save any changes before exiting the page.

The user describes the technical aspects of the data context in detail but provides no information about the legal provisions implied by the particular data context. Instead, MIKE populates appropriate legal information automatically based on the parameters entered in the data context. Figure 3.6 shows a completed data context page. The legal information is in a box on the right. This legal information is not manually linked to a data context, but rather dynamically applied based on the information encoded into the rule and how it matches the data context. For instance, in this particular case, MIKE thinks an administrative subpoena is required to access the data in Facebook’s “Account” data context. This is because MIKE has a legal context (described later) for ECPA that says an administrative subpoena is needed to get phone numbers (among other data types) from firms engaging in the activity of social network
operation (among other activities), and these descriptions fit the Facebook App/Accounts data context. It is important to note that this is not legal advice but rather information provided to investigators or policymakers highlighting rules that might apply.

**Describing a Data Type**

MIKE relies heavily on data types for providing search results and linking laws to data contexts. Because data types are entities, each has its own page in MIKE. The form for updating the “Real Name” data type is shown in Figure 3.7. The information describing each data type is currently very simple. A place is provided to describe the data type, and the user can select from several categories to categorized the type of information.

However, once the data type’s page has been created, additional detail will be generated based on the information within MIKE. As Figure 3.8 shows, in addition to the description of the data type, each data type’s page lists all apps and corporations that collect this type of data. For example, because the Facebook App described above collects the user’s real name, this app appears on the “Real Name” data type page. While this list will often consist only of apps, a corporation will sometimes have access to data it does not obtain from any particular app. This can occur when firms share data, or firms buy data from a data brokerage.

**Describing a Firm in the Map**

Continuing with the prior example, Facebook, Inc. is the owner of the Facebook app. It is important to make the name of the app different from the name of the parent company because these are very distinct entities in MIKE. Figure 3.9 depicts the form used to enter information about a firm.

As usual, there are places for both structured and unstructured information. The structured information includes both details about how law enforcement can obtain information from the firm and specific details about the firm itself. “Activities” will generate a drop-down list of suggested inputs if the user presses the down arrow on the keyboard.

---

5 *Real name* refers to how an app characterizes the data type. Of course, that name may be fake and just an identifier. That identifier may also be used elsewhere to help link investigative threads.
Several entries on this form can be used to describe how law enforcement can obtain information from the firm. For instance, if a firm has a dedicated phone number or website for law enforcement officers seeking user information, the appropriate phone number or URL can be added here. Users can also add information about how long it takes the firm to respond to a law enforcement request and the cost of the response. These boxes can be populated either from information from the company or through the experiences of law enforcement officers who have sought information from the company.

The user can provide additional information about the firm itself. For instance, the user can enter information about the activities the firm directly engages in. Activities that are the result of owning or operating a product should not be listed here. For instance, if a firm owns a social networking site, list “Social network operation” in the page for the social network, not on the page for the firm. Activities on the firm page are for things the firm does without any digital product; for instance, “Data Brokerage.” It would not hurt to list the activities implied by all its products, but doing so would involve entering data multiple times, and MIKE automatically assigns activities to a firm that are entered for any of its apps.

Additionally, the user can enter information about any corporations the firm is a subsidiary of. As before, it is best to enter information from MIKE from the top down: First enter the parent company, then enter any subsidiaries. However, MIKE is flexible enough that, if needed, a firm can later be updated to specify information about the parent company.

As with mobile apps, corporations can have data contexts that describe how information is being collected. A firm-level data context is meant to capture the data the firm collects that is not collected by a particular product. In this case, MIKE has nothing for Facebook. Using one of MIKE’s search wizards will reveal all the data types a firm collects—the data contexts of both the firm and its apps. Figure 3.10 shows the completed page for Facebook, Inc.
MIKE populates additional information about the firm based on other entries already in MIKE. For example, even though the user did not specify the mobile apps Facebook operates in the form in Figure 3.9, a table describing these apps automatically appeared on the page in Figure 3.10. This table will continue to update as more apps are added to MIKE. To add a new mobile app owned by this firm, the user enters the information into the tool bar under the table to create a page for the mobile app. Once the mobile app's page has been created, the form depicted in Figure 3.3 can be used to enter information about the app.

The firm's page also provides a list of the subsidiaries of the firm. This list is automatically generated based on the information provided in the subsidiary's page. In the example shown in Figure 3.10, Instagram, LLC is listed as a subsidiary of Facebook, Inc. For this to occur, the user who edited the page for Instagram, LLC must have described it as a subsidiary of Facebook, Inc.
Describing a Law in the Map

Describing a Statute

It is generally preferable to add legal information to MIKE by beginning with a particular statute or case, and then adding information about the terminology, contexts, and legal processes implicated by that statute. First adding general information about each statute allows the user to provide important background, as well as create a nexus for the individual rules within the statute. We begin by describing the process for adding information about a statute.

Figure 3.11 shows the form used to create an ECPA entry.

This form begins with a place to add unstructured information. Here, the user can provide a short, general description of the statute. If the statute has a preface or purpose section that describes the statute, that information could be used. Below the unstructured information are boxes in which the user can enter structured information describing the statute, including the jurisdiction affected by the statute, references to the statutory text, and appropriate legal
This allows the user to refer to general information about the jurisdiction or reference the statute as needed.

There is also a box for listing relevant terms defined in the statute. We will explain the process for describing legal definitions in the next subsection. But for now, note that “Relevant Definitions” is generally populated by referencing a code section defining terms for the statute as a whole (e.g., 18 U.S.C. §2510). A page is created for each statute that lists the definitions in it, and each definition links to a separate page pertaining to that definition. The user can then add a description of how the term is defined to the definition page.

Finally, the user can enter the different legal contexts. Each of these corresponds to a particular rule created by the statute. Because many privacy statutes articulate different rules for
different types of information, legal contexts will frequently correspond to different information types. However, if the statute articulates different rules for a particular type of information depending on the circumstances, it may be necessary to create more than one legal context pertaining to that information type. Users can add as many legal contexts as they wish. Legal contexts can be removed using the buttons on the right and reordered using the arrows on the far right. The process for describing legal contexts will be described next.

Describing Legal Definitions

Figure 3.12 shows the form used to update a page corresponding to a legal definition. In this particular instance, the term defined is *intercept*.

First, note that the form depicted on this page has entries pertaining to many legal definitions—not just a single definition found in a particular statute. This is done so that the completed page will allow the user to compare how the definition of a term varies across jurisdictions and statutes. As we will show later, all these definitions will be presented in a single, organized chart.

To add a new entry to this page, the user begins by clicking the button at the bottom of the page entitled “Add Legal Definition.” This creates a new blank entry in which the user can add a definition from a new statute. Once this blank entry has been created, the user should enter the text of the definition provided by the statute, the relevant jurisdiction, the citation, and (if possible) a link to the full text of the statute. The user must save these changes before proceeding.

A screenshot of the completed entry for “Intercept” is shown in Figure 3.13. This page allows the user to easily cross-reference definitions from different jurisdictions to analyze similarities and differences.

Describing Legal Contexts

A single statute may govern many different types of information collected by many different types of entities. Consequently, when entering a statute into MIKE, it is important to deconstruct the statute into an individual set of rules, each of which governs a particular type of information that law enforcement officials can obtain and a particular type of entity they can obtain it from. Correctly defining legal contexts is crucial for ensuring that MIKE connects the appropriate legal information with the appropriate technical information. Figure 3.14 shows the form used to update the legal context related to ECPA/Interception of Communication.

Legal contexts relate to statutes much the same way that data contexts relate to mobile apps. Legal contexts are always related to a particular statute, and the title of each legal context includes both the title of the statute and the type of information being obtained by law enforcement, separated by a slash mark.

The user has the option to enter both structured and unstructured information about the data context. The structured information provides a description of the legal context and how it is applied. The following are included as structured information:

- **Data type** refers to the type of information governed by this particular legal context. Wherever possible, the same terminology should be used for defining data types in legal contexts as is used for defining data types in data contexts, to ensure that the legal information populates correctly. To assist with this process, the user can press the down arrow to generate a drop-down menu of potential entries.
• **Jurisdictions** refers to the places governed by the statute underlying this particular legal context. In this example, both “United States federal jurisdiction” and “United States jurisdiction” are included as relevant jurisdictions. These terms describe two different jurisdictions: “United States federal jurisdiction” pertains only to federal actors, while “United States jurisdiction” pertains to any state or federal actors in the United States.

• **Activity** describes the activity conducted by the entity from whom law enforcement seeks information. Because different laws apply to information collected by different categories
of entities (i.e., entities that provide electronic communications services), specifying the appropriate entity ensures that legal information is correctly populated within MIKE.

- **Data location** describes the place where the data governed by the legal context are being held. In this example, the data are held remotely by the commercial firm; they could also be stored locally on the device.

- **Law enforcement situation** refers to the activity law enforcement is undertaking by seeking the information. As we will discuss later, under “Describing Law Enforcement Situations,” specifying the law enforcement situation helps MIKE connect related statutory and case law.

- **Policy issue discussed** allows specification of any policy issues that the user thinks are implicated by this particular legal context. This enables MIKE to create pages capturing all statutes and cases related to a particular policy issue, to facilitate policy analysis. While the law enforcement situation and the policy issue discussed are the same in this particular example, this is not always the case.

- **Legal code** provides a statutory reference for legal context.

- **Legal references** provides a link to a website containing the text of the statute.

- **Required legal document** describes the legal document law enforcement must seek to obtain the information covered by this particular legal context. Once the legal context page is complete, the entry in this row will turn into a link, which the user can follow to...
find more information about how to obtain this legal document. MIKE’s page for legal process documents will be described in more detail in the next subsection.

In addition to the structured information, the page also contains a place for the user to enter unstructured information. Whenever possible, the statutory text authorizing collection of information by law enforcement should be entered as free text to facilitate research and analysis. All information except the statutory text is stored semantically, meaning once the page has been saved this information will become links. Users can follow these links to obtain additional material whenever they wish.

Describing the Process for Obtaining Legal Documents
The pages describing legal documents provide information about how a law enforcement officer can seek the information governed by a particular legal context. Because legal process documents always relate to a particular statute, their name is derived from the name of the

---

6 It should be noted that multiple statutes may require law enforcement to obtain the same legal document before seeking information under different circumstances. Therefore, MIKE could have created centralized pages for legal process documents that would allow the user to compare how the legal document is used in different contexts (similar to the pages created for legal definitions). We elected not to create centralized pages for legal process documents to ensure that information about each document is presented as simply and clearly as possible. However, future updates to MIKE could explore developing pages that cross reference how particular legal documents are used in different contexts.
statute and the name of the document being described (separated by a slash mark). Figure 3.15 is the form for updating the entry for “Court Order Authorizing Intercept.”

Each form provides multiple places where the user should enter information about the legal document, how law enforcement should request this document, and the circumstances under which this request should be granted:

- **Jurisdictions** refers to the jurisdictions in which the statute associated with the legal documents applies.
- **Decisionmaker** refers to the person or entity that determines whether the legal process document should be granted.\(^7\)
- **Relevant legal standard** refers to the proof the government must provide for the document in question to be granted.
- **Legal code** refers to a citation to the relevant statute.
- **Legal references** provides a link to a webpage with the full text of the statute (if available).
- **Requirements for app** describes what the government must demonstrate when requesting the legal document. Whenever possible, this section should incorporate the text of the statute to facilitate further research and analysis.

\(^7\) In this particular box, there is no drop-down menu with potential options for decisionmaker. While it would be possible to create a drop-down menu in future iterations of MIKE, this would require creating a category describing potential decisionmakers and updating the form used to create legal context pages.
• **Process for granting document** describes what the person or entity presented with an app for the document must do (including what findings that person or entity must make) for the document to be granted. Whenever possible, this section should incorporate the text of the statute to facilitate further research and analysis.

• **Contents of document** describes what should be included in the document. Whenever possible, this section should incorporate the text of the statute to facilitate further research and analysis.

• **Other information** provides any other information related to requesting or granting the document.

Describing Law Enforcement Situations
MIKE also contains pages that describe the laws that might apply in particular law enforcement situations. These pages connect statutory and case law, and are primarily meant to facilitate legal research and policy analysis. Figure 3.16 shows the form used to update the law enforcement situation page for interception of communication.

This form only provides space for the user to enter a description of the law enforcement situation and links to webpages outside MIKE that pertain to that situation. However, once the user saves any changes, the completed page referring to each law enforcement situation also includes tables summarizing the relevant statutory and case law (Figure 3.17). These tables are created and updated automatically based on the statutes and cases described in MIKE. Although the legal information presented is from all jurisdictions included in MIKE, this information can be sorted based by jurisdiction by clicking on the arrow next to the jurisdiction heading. Users wanting to obtain more information on other laws implicated by the law enforcement context can click on the links associated with a particular case or statute. Users can click on any entry in the tables to obtain more information about a particular case or statute.

Figure 3.16
Updating Law Enforcement Situation for Interception of Communication
Describing Case Law

Generally, MIKE should include only the highest court decision for each case. For example, if a case was decided by a federal circuit court and then decided again by the U.S. Supreme Court on appeal, MIKE should include only the Supreme Court decision. Figure 3.18 is the form used to describe *United States v Jones*.

---

8 The prototype of MIKE does not have a mechanism to enforce this. The prototype assumed this was the responsibility of a legal curator and that this would be a feasible way to operate an implementation of MIKE. We believe that adding an enforcement mechanism to MIKE would require extending it to recognize the court hierarchy to select the highest court decision. An implementation of MIKE might consider doing this.
Each form provides places for the user to enter information about the jurisdiction of the deciding court, legal citation, date, holding, background, applicable circumstances, and types of electronic surveillance implicated. In addition, the entry for each case includes a field that allows the user to specify the law enforcement situations and policy issues implicated by the case. The contents of these fields are used to create pages presenting all the laws relevant to a particular scenario or policy issue (as discussed earlier).

While much of the information used to describe case law is the same as that used to describe statutes and legal contexts, some aspects are unique. In particular, users should include the following information when adding a case to MIKE:

- **Court** refers to the court that issued the ruling on the particular case being added to MIKE. A drop-down menu is available to assist in adding the relevant court.
- **Constitutional provisions interpreted** refers to the constitutional articles and amendments discussed by this particular case. If the Constitution is not referenced in the court’s opinion, this should be left blank.
• Statutes Interpreted refers to the statutory law that the court discusses in this particular case. If no statute is referenced in the court’s opinion, this should be left blank.
• Date refers to the date on which the opinion in this particular case was issued.
• Background refers to the context of the particular case and the issue presented to the court. For example, it would be appropriate to include information about the technology at issue and actions law enforcement took to obtain evidence in this case and any factors that might limit application of the holding of the case in the future.
• Primary holding refers to the court’s decision in this particular case. If possible, this should describe not only the outcome of the case but also the rationale underlying this decision.
• Surveillance type refers to a description of the specific methods that law enforcement used to gather information in this case.

Wherever possible, the information provided should include direct quotations from the case to facilitate research and analysis.

Editing the Structure of the Map
In addition to editing the content of MIKE, it is also possible to edit its structure (assuming the user has sufficient privileges in the system). Editing the structure would allow the user to create a permanent page that shows the results of some query, to add new types of structured information to corporations, or to specify a new type of relationship. MIKE is built on top of MediaWiki and makes use of the Semantic MediaWiki plug-in. The structure can be edited within MIKE. There is no need to edit the underlying source code. However, it is necessary to understand how MediaWiki templates work and how queries are specified in Semantic MediaWiki. While a thorough description of MediaWiki is outside the scope of this discussion, several resources available online provide further information about MediaWiki.

Changes to the underlying templates can create problems with existing pages if not done carefully. The addition of new relationships can usually be done such that existing pages simply display a blank value, which can be filled in when a user updates the page. Because all changes to MIKE are logged, it is possible to revert back to a previous functioning version of MIKE if changes to the structure of MIKE negatively affect MIKE’s functionality.

Instructional Videos
The MIKE prototype includes a number of instructional videos to help new users understand the system. These include a broad overview of what MIKE is, a demonstration of how to enter a new mobile app into MIKE, and instructions on how to search MIKE. These can be found from a link on the main page.

Populating and Developing MIKE
MIKE is populated with several types of data. These include descriptions of commercial enterprises and their relationships, data types collected by commercial enterprises, and the law and

9 For the prototype, MIKE was populated using publicly available information describing about 40 apps for mobile devices for various device operating systems. For the legal information, the law and jurisprudence of six states, including the overarching federal laws applicable in these states, were included, as well as the District of Columbia.
its relationship to these data types. The subsections that follow describe how we gathered the data used in the prototype of MIKE, some alternative ideas for gathering and refining more data, and issues that may need to be addressed as MIKE is developed.

**Sources of Information on Mobile Apps**

Obtaining the data to enter into MIKE can be difficult. Mobile app developers are not always explicit about what data they collect, how they store and protect the data, and what their data retention policies are. To facilitate communication, it helps to categorize information types. There are many possible ways to do this. To populate the MIKE prototype, we developed one useful way by considering the transmission viewpoint (Figure 3.19):

Figure 3.19 shows the following levels:

- **Level 1** data are those a user enters into an app or provides access permissions to. This is stored locally (on the device) and is usually transparent to user. An example would be a photograph taken with the device.
- **Level 2** data are sent remotely to a server controlled by the app (or mobile website) owner. Users are likely aware of transfer but may require technical expertise to examine exactly what is being transmitted. Checking in with Foursquare is one example.
- **Level 3** data are sent from the phone to a third party (via an application programming interface [API], for instance). Users are less likely to be aware of this data channel and, as with level 2, they would need technical expertise to examine the contents. An example would be Foursquare sharing user data with another company.
• **Level 4** data are sent from the app server to some other server. Users are unlikely to be aware of this data transfer, and there is no way for them to know that it occurs (aside from a possible mention in the firm’s privacy policy). The transfer is one to one between servers and does not involve the mobile device. An example of this is a bilateral agreement between two firms to buy and sell data about users.

• **Level 5** data are sent to an aggregator or exchange. Users are almost certainly unaware of this, and the exchange is one to many. Any number of companies may have access to the data after they have been placed in an exchange. Foursquare selling data to Axciom, a well-known exchange, would be at this level.

Data types that are more opaque to end users are also more opaque to researchers attempting to understand how data are collected and shared in the mobile ecosystem. To understand the sources that do exist, it is useful to imagine what an ideal data source would look like, then compare it to what is actually available.

An ideal data source would contain a complete listing of all the types of data that a particular app or mobile website could *potentially* save, along with information about whether the data are *actually* saved.10 This is a subtle but important distinction that we will make clear later through an example. In contrast, all actual data sources contain a mixture of potential saved data and actual saved data. For instance, all Android apps are required to declare what permissions they require to run in an “AndroidManifest.xml” file. These files can be automatically scanned and the data aggregated to provide a listing of possible data types each app can access. It would even be possible to import the data directly into MIKE, where it could be adjusted, added to, and searched. The problem is that permissions allow a device to access particular hardware features (such as the camera) or software features (such as the calendar), but this does not mean the app actually accesses those features; even if it does, it may not save any relevant data. A wealth of data can also be created by apps that require no permissions, such as personal notes and to-do lists.

Another, less-common data source is a forensic analysis that monitors an app in progress and keeps a record of any information saved to the phone or transmitted to a remote server. While this ensures the results of the analysis include only data that are actually saved, it does not ensure that all possible saved information is captured. The app may only save or transmit data under particular circumstances, which the analyst does not investigate.11

Tables 3.1, 3.2, and 3.3 identify data sources that have been found for the first three of the levels of data depicted in Figure 3.19. The “Effort per App” column in each table uses a qualitative scale to indicate how difficult it is to extract the information about an app using the particular source. The “Technical Difficulty” column is for the programming skill necessary to use the data source.

The “Scope” and “Coverage” columns describe the total number of apps that can be analyzed and the completeness of information obtained for any particular app, respectively. All assessments are based on the authors’ qualitative judgment. *Scope* specifically refers to how many apps the data collection method would apply to. A *full* scope means that the process can be used to analyze any app. *Partial* and *limited* indicate progressively smaller subsets of apps.

---

10 Indeed, decisionmakers may find this distinction useful as they craft policies governing mobile app privacy.

11 It may be possible to automate this process, which would make it a much more feasible for developing information for MIKE (“The Truth About Smartphone Apps . . .,” 2015).
**Table 3.1**
**Level 1 Information Sources**

<table>
<thead>
<tr>
<th>Source</th>
<th>Description</th>
<th>Effort per App</th>
<th>Technical Difficulty</th>
<th>Scope</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>App permissions</td>
<td>Apps are granted permission to access particular types of data. However, these are only a subset of the types of data the app may store, and there is no guarantee that the app actually uses the data.</td>
<td>Medium</td>
<td>Medium</td>
<td>Full</td>
<td>Possible</td>
</tr>
<tr>
<td>App description</td>
<td>Data other than those granted via permission are often listed as features of an app, like storing user-generated content.</td>
<td>Medium</td>
<td>Low</td>
<td>Partial</td>
<td>Subset</td>
</tr>
<tr>
<td>App permission data sets</td>
<td>Some organizations have already collected app permission data (Sverdlove and Cilley, 2012; Hoffman, 2012).</td>
<td>Low</td>
<td>Low (with access)</td>
<td>Partial</td>
<td>Possible</td>
</tr>
<tr>
<td>Privacy policy</td>
<td>Apps often provide a privacy policy. However, these are often vague.</td>
<td>Medium</td>
<td>Low</td>
<td>Partial</td>
<td>Subset</td>
</tr>
<tr>
<td>Phone investigation</td>
<td>A forensic examination could reveal all data accessed, saved, and transmitted by an app.</td>
<td>High</td>
<td>High</td>
<td>Full</td>
<td>Complete</td>
</tr>
</tbody>
</table>

**Table 3.2**
**Level 2 Information Sources**

<table>
<thead>
<tr>
<th>Source</th>
<th>Description</th>
<th>Effort per App</th>
<th>Technical Difficulty</th>
<th>Scope</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Privacy policy</td>
<td>The privacy policy may include information about what data are stored remotely.</td>
<td>Medium</td>
<td>Low</td>
<td>Partial</td>
<td>Subset</td>
</tr>
<tr>
<td>App description</td>
<td>Some apps will not work without the ability to save certain data externally (anything that has synchronizes multiple devices).</td>
<td>Medium</td>
<td>Low</td>
<td>Full</td>
<td>Subset</td>
</tr>
<tr>
<td>Forensics</td>
<td>Monitor an app using software to capture all external data transmissions. TaintDroid (Enck et al., 2010) is an example.</td>
<td>High</td>
<td>High</td>
<td>Full</td>
<td>Complete</td>
</tr>
<tr>
<td>Investigative reports</td>
<td>Reports have shown which data are transmitted by particular apps (see, for example, “What They Know—Mobile,” 2013).</td>
<td>Low</td>
<td>Low</td>
<td>Limited</td>
<td>Subset</td>
</tr>
</tbody>
</table>

*Coverage* refers to how much of the data being collected by an app could be observed using the process. *Possible* refers to finding out that an app could (but does not necessarily) save information; *subset* indicates sure knowledge that an app does save a particular set of data (but not knowing whether it saves anything else), and *complete* indicates sure knowledge of all data saved.

Currently, no resources have been found for gathering information about level 4 and 5 information sources. These sources are new and opaque.
Similarly, identifying the appropriate legal data to enter into MIKE can be difficult. Both federal and state governments have been active in regulating law enforcement surveillance of mobile devices, apps, and services. Legislators have made policy by enacting statutes, and courts have made policy by interpreting state and federal constitutions. As the “supreme law of the land,” federal law prevails over contradictory state law. Therefore, state-level law enforcement must consider not only the statutes and case law of their particular states but also the federal statutes and case law relevant to their jurisdictions. It was therefore crucial that MIKE include both state and federal law to completely describe the legal protections available for information within the mobile app ecosystem.

As part of our preliminary research effort, we selected six pilot states for a thorough analysis. After identifying the pilot states, we sought the relevant statutes and case law on the state and federal levels. The relevant law was identified through a literature review, search of proprietary legal research databases (including Lexis), and a review of published indexes of privacy laws. Despite our use of proprietary sources to conduct research, we obtained the documents used through public source materials wherever possible so that the finished product would be useful to researchers without access to these databases.

After identifying the relevant laws, we developed a framework for analyzing them in a consistent, reproducible, and scalable way. Drawing from the social science methodological literature, we employed systematic content analysis to create such a framework. In the remainder of this subsection, we discuss in greater detail both how we identified the laws that MIKE should include and how we analyzed the laws we identified.

12 Because we partnered with law enforcement officers in Indiana and Nevada, these states were included so that we could obtain vital feedback on MIKE’s usefulness and accuracy. New York and California were also included because laws established in these states are often influential in other jurisdictions. We also identified Ohio and Texas as useful pilot states, since recent case law from these jurisdictions suggests the adoption of interesting policy positions related to electronic surveillance.
**Identifying Laws to Include in MIKE**

While the necessary documents are in the public domain and generally easily available, it is not always clear whether MIKE should include a particular law. Courts often approach novel technology by analogy with technology regulated by existing case law; therefore, a case that appears to be entirely unrelated to mobile technology may be very important in the development of new rules. However, it would be impossible to include all cases related to law enforcement searches and seizures, and even if it could be done, the resulting tool would be hopelessly cluttered and difficult to use.

Instead, we decided to prioritize the case law in MIKE according to its relevance to the mobile app ecosystem. Initially, we only included cases that directly involved or affected mobile technology. However, as future users expand MIKE, a broader scope of information can be included. We suggest that this expansion should proceed as described by the Venn diagram in Figure 3.20.

MIKE is currently designed to include laws that directly concern mobile technology—e.g., a case in which the court determines whether law enforcement can search the contents of a cell phone subject to an arrest. We recommend that future users next focus on laws that

---

**Figure 3.20**

A Strategy for Deciding What Legal Information to Include in MIKE
regulate the data types implicated by mobile technology, such as location or social network information—e.g., a case concerning whether location information is inherently deserving of Fourth Amendment protection. Next, we recommend expanding the scope of MIKE to include all cases that involve technological search and seizure—e.g., a case determining whether law enforcement has exceeded the scope of its warrant during the search of a computer. Finally, users may consider including cases dealing with search and seizure—e.g., foundational cases describing the scope of individual protections against government searches—although this is probably too ambitious and may overwhelm every user of MIKE except for policy analysts.

Broadening the scope of legal information has two advantages. First, MIKE will become more useful to its intended user base as additional laws are included, since this will provide richer and more nuanced information. Second, including foundational and background information allows a broader audience to use MIKE. However, care must be taken to present this background information in such a way that it does not overwhelm the information specific to law enforcement searches of the mobile app ecosystem. Partitioning the case law within MIKE and allowing users to select the breadth of information they want presented may accomplish this.

**Analytic Framework for the Law**

After identifying an appropriate set of laws, it was necessary to analyze them so that the information they contain can be incorporated within MIKE. We sought a framework for legal analysis that was consistent, replicable, and scalable. This would ensure that new and changing legal rules could continue to be added to MIKE as it was developed and expanded. If the framework is consistent, similar legal protections will be entered the same way across different laws and different jurisdictions. If the framework is replicable, two different people will be able to code the same legal rules the same way. If the framework is scalable, the process will lead to interesting and illuminating results, even if research is conducted only in particular jurisdictions. A review of the literature revealed a well-established, widely respected social science methodology that fulfilled all these criteria: systematic content analysis.

**Description of Systematic Content Analysis**

Systematic content analysis is a method for “classifying textual material, reducing it to more relevant, manageable bits of data,” thereby developing “a set of procedures to make valid inferences from text” (Weber, 1990). This method makes it possible to cull information from documents in a reproducible, consistent way. Systematic content analysis has previously been used to analyze legal decisions, allowing researchers “to find some order and logic in a body of case law that, by conventional analysis, appears chaotic or haphazard” (Hall and Wright, 2008). As Fourth Amendment case law has frequently been described as inconsistent and disorganized, a framework derived from systematic content analysis seems particularly well suited to provide the type of consistent and replicable investigation needed for MIKE’s development.

Systematic content analysis is superficially similar to traditional legal analysis. Both methods require a researcher to identify relevant documents, read them carefully, and draw inferences about their content. However, systematic content also requires the researcher to develop coding procedures prior to analyzing the documents. These coding procedures are rules for assessing the meaning of the laws: Rather than reading cases individually to determine their substance and importance, the researcher evaluates them based on a predetermined set of decision criteria. If these decision criteria are well developed, two different people evaluating the
same law should generate the same classification. Furthermore, provided the cases are properly selected, this procedure can be performed on any subset of laws. Consequently, the results of systematic content analysis should be consistent, reproducible, and scalable—precisely the characteristics needed to enter legal information in MIKE.

Description of Framework
We created a three-step procedure, based on systematic content analysis, for analyzing and entering laws into MIKE. The first step was to determine which laws should be entered in MIKE. This required both jurisdictional and subject-matter scoping. Jurisdictional scoping required the selection of appropriate pilot states and the identification of the state and federal laws applicable in these pilot states, as discussed earlier. In addition, because case law is often applied to emerging technology by analogy, we developed criteria for determining whether a case was related to the mobile app ecosystem closely enough to warrant inclusion in MIKE. For the initial development phase, we limited the case law included to cases that directly involved mobile technology. However, as was discussed previously, we also developed criteria for expanding MIKE to include cases that more broadly implicate mobile technology.

The next step was to create a set of relevant dimensions: aspects of each law that we wanted to capture and include in MIKE. Relevant dimensions are the options for entering information. These are intended to constrain how data are entered and reduce variations. These dimensions ranged from simple background information intended to allow the law to be properly situated in the semantic network to detailed questions intended to extract intricate information related to the application of the law. After identifying these dimensions, we developed specific coding criteria for determining the outcome for each dimension. These criteria required specific references to the laws wherever possible, to help the user conduct research and analysis.

Finally, we entered the laws into MIKE using the coding criteria developed. The fields designated for each dimension were filled with the appropriate responses. This included developing pages for each legal rule and associated pages for the legal documents referenced, law enforcement context implicated, and the policy issue affected. These associated pages tie together information contained in many different laws, to allow the user to research and compare laws across jurisdictions.

Addressing Conflicting Interpretation of the Law
Although electronic surveillance law is evolving rapidly, there will always be areas of ambiguity where application of the law is unclear. The goal of MIKE is to present objective information wherever possible to assist the user with further legal research. This is why the user is encouraged to add quotations from the text of a statute or case rather than summarize it. This leaves interpretation to the reader rather than to the person entering information. However, where areas of conflicting interpretation cannot be avoided, MIKE’s goal is to identify and manage the conflict, not resolve it. When the process of systematic content analysis described earlier can be applied to yield conflicting valid interpretations of the current state of the law, MIKE should be developed in such a way that it presents both alternatives.

Variations could be further reduced by involving an impartial curator in an operational version of MIKE. However, there is genuine uncertainty in the law. MIKE is intended to capture disagreements, and a curator would need to ensure that meaningful disagreements were incorporated. This feature is important when stakeholder groups disagree about policy, e.g., law enforcement and privacy advocates.
Some aspects of the legal information MIKE requires are relatively unambiguous and uncontroversial. In particular, it is unlikely that there will be many disagreements about which court issued an opinion in a particular case. Other aspects may be ambiguous. For example, it is likely that prosecutors and privacy advocates may describe the holding of certain cases differently. If possible, the preference is to add a quote from the case in which the court directly stated its holding (“We hold that . . .”). However, many opinions do not include such language. In this instance, users should be encouraged to add conflicting statements of the holding (providing supporting language directly from the case) but not to erase conflicting interpretations. Consequently, an entry in MIKE describing a case could eventually present several possible interpretations of the holding, each supported with quotes from the case. It is likely that these interpretations will be similar but vary in scope: A district attorney might take a more constrained view of a holding expanding Fourth Amendment rights, while a privacy advocate might take a broader one. A user reading MIKE’s entry for the case could then see competing descriptions of the holding, which would help provide insight into the range of potential interpretations. By viewing the history of the page, the user could also observe who added which interpretations.

Observations About Adding Information to MIKE
Data about the mobile app ecosystem can be difficult to obtain. Fully populating MIKE will require researchers with expertise in technology, commerce, and law to identify and exploit potential sources of information. Particularly as MIKE is being initially developed, having experts (or crowds of experts) add information will help create a base of knowledge that can be expanded in the future.

Additionally, MIKE is meant to allow investigators and others with hands-on experience with surveillance of the mobile app ecosystem to add additional information as they discover it. In particular, some types of information MIKE collects are best added based on the experiences of those involved with surveillance of the mobile app ecosystem, for example, the length of time a company takes to fulfill a request. MIKE’s philosophy is to capture data about electronic surveillance, either by systematic analysis or in the course of an investigation, and to retain and organize the data in a tool that makes it easy for others to use, refine, and share. This approach should increase the value of the tool as it is used and as its content is refined by stakeholders.

Potential Issues in the Development and Implementation of MIKE
One of the challenges an implementation of MIKE faces is how to ensure that its content is filled out in a conceptually uncluttered way. It is unlikely that all users of an implementation of MIKE would enter data. In practice, a small community of law enforcement professionals who are electronic surveillance specialists typically support others in an agency. Electronic surveillance specialists are the individuals most likely to add data. Most law enforcement professionals will be expected to be consumers of MIKE’s content. The organized privacy advocacy community is small and reasonably well aligned. It is unlikely that an implementation of MIKE would fail because of scaling problems with the number of users contributing content to it (as opposed to using it).

We assumed that an implementation of MIKE could use some combination of crowdsourcing and moderation to assure coherence and avoid editing wars, in which stakeholder groups that hold different opinions about ambiguous information, such as law. The MIKE
prototype relied on wiki extensions that enable drop-down lists of standard terms to choose from when adding a term in a template field. When new terms need to be introduced, they are added to the drop-down list.

While the operational details of an implementation of MIKE are outside the scope of this report, any implementation must consider and establish roles and responsibilities in the operation of MIKE. Introducing a moderator or editor for crowdsourced information is analogous to Wikipedia’s approach. An implementation—built using the same software that Wikipedia uses—could be extended to use Wikipedia’s moderation tools (e.g., blogs to dispute the accuracy or completeness of content). The moderator or editor needs to be neutral so that valid alternative interpretations of ambiguous data by stakeholder groups are tolerated and documented. The moderator or editor can also look for different terms being used to describe similar concepts and can revise them using a single conceptual model and terminology. Alternatively, a curator could be designated to gather and vet all data captured by an implementation. Again, the curator needs to be impartial and attuned to finding and documenting valid alternative interpretations. In this case, the curator defines the conceptual model, and users of the implementation would need to learn that model.

Security and trust are very real concerns for an implementation of MIKE. For example, a law enforcement-only implementation would need to vet users, have a user-authentication mechanism, and ensure access to and the confidentiality and integrity of the data captured in the implementation. To support users of multiple stakeholder groups, an implementation could take one of two approaches. It could extract subsets of data into separate instances supporting individual stakeholder groups. The challenge would be to maintain consistency among multiple views of the data. Alternatively, an implementation could use portion marking–type controls and role-based user access to limit the data stakeholder groups can see or change.

To design the data extracts or apply portion controls, an implementation would need to study the trust issues among stakeholders. Generally, stakeholder groups have overlapping interests in the content of an implementation of MIKE. For example, law enforcement and privacy advocates both want to know the details of the information collected by commercial firms, but for different reasons. However, they do not fully trust one another and are only willing to share portions of the data in an implementation with another stakeholder. Also, stakeholder groups do not uniformly trust a particular organization that might operate an implementation of MIKE. Thus, the design of an implementation depends not only on the structure of MIKE but also the design and choice of operators, curators, etc. This analysis is beyond the scope of this document, but we believe is doable.
Architectural Goals and Attributes

To adequately model the complex and rapidly evolving technical, commercial, and legal landscapes of the mobile ecosystem as it relates to electronic surveillance, we set out to create a mapping tool that would have the following architectural attributes:

1. **Expressive.** MIKE needed to convey complex relationships among commercial firms that supply mobile devices, communications, apps, and services to understand what information is stored within the devices or by the commercial entities and how law enforcement can use that information. MIKE also needed to convey the equally complex arrangement of federal, state and local laws, and the jurisprudence that governs access to that information. Therefore, we needed a tool with rich semantics.

2. **Navigable.** MIKE is intended to be used by a wide variety of individuals with varying levels of technical ability and interests—law enforcement, commercial entities, and policy advocates or makers. Therefore, our goals were to not restrict how users navigate MIKE and to create relationships for searching MIKE that support a variety of purposes, e.g., investigation and policy analysis. We needed a tool with a familiar look and feel in which user action determines the information appropriate to view next.

3. **Extensible.** Given the rapid evolution of mobile phone technology and related law, as well as an ever-increasing number of commercial entities involved in the mobile ecosystem, any tool that maps the relationships between them will be obsolete the day it is declared “complete.” Therefore, it was imperative that MIKE have the capability to add new information, new entities, and new relationships with minimal burden on maintenance personnel and no disruption to users.

4. **Auditable edit.** While we wanted MIKE to be easy to update and extend, we are also aware that updates can damage information either by mistake or malicious action. At minimum, we needed MIKE to support
   a. version tracking with the ability to restore content from prior versions (i.e., rollback)
   b. clear delineation of verified and unverified content
   c. the ability to develop reliability ratings for contributors and block access to those who are determined to be malicious.

The software we used to implement MIKE meets the first of these three needs. The second and third requirements are associated with the content of MIKE and its structure. The prototype does not fully implement them, but we recognize the need to do so in a deployed system.
With the above architectural attributes in mind, the first section of this chapter reviews prior work that informs and inspires our approach to modeling the technical, commercial, and legal landscapes. The third section presents a brief overview of the architecture of MIKE and discuss how this architecture supports the required attributes.

Prior Approaches

Prior Depictions of the Technical and Commercial Landscapes

One type of mobile ecosystem map is a high-level graphic meant to illustrate one specific aspect of mobile technology. Most are static depictions. Usually, a graphic is provided with no explanation of how firms were assigned to particular categories. Since technology and commercial entities are evolving quickly, any static depiction is likely to be brittle. While static maps may be directionally correct, they are not easily extended in response to changes in the technical, commercial, and legal concepts they describe and are ill-suited to the aims of this project. However, some static depictions provide compelling views that inform the structure of our ecosystem-mapping tool, as described in the following paragraphs.

Srinivasan (2012) describes the mobile ecosystem as a set of “walled gardens.” Each major platform provider (Amazon, Google, Apple, Microsoft, and others) has a hardware platform to which the provider adds an operating system and critical apps, such as geographic maps, email, and contact management. These providers also supply proprietary “stores” from which consumers can purchase additional third-party apps that the platform provider authorizes. Such app stores are designed to keep consumers locked into a particular “garden” (Apple’s iTunes and Google Play are two examples). Mozilla’s idea that carefully designed and integrated vertical chains exist in an otherwise unplanned (e.g., organic) system deserves special note. Ideally, this is a desirable output of a fully characterized ecosystem-mapping tool: Firms could be arranged into groupings according to the relationships encoded in the tool, and these groups would likely match the walled gardens Mozilla describes. Using these relationships, it would also be possible to estimate the relative strength of the various walled gardens and identify key players that participate in multiple gardens.

Another useful static depiction describes the supporting services available to mobile app developers. These services supply features and scalability for integration into other, higher-level apps, saving development time and effort and giving the set of higher apps that use a specific support service provider a common look, feel, and operating pattern. These back-end service providers allow even two-person development teams to bring new apps to market quickly. For example, Kinvey Inc. operates a what it calls a Backend as a Service platform and depicts the available supporting services within the mobile ecosystem (Rice, 2013, diagrams this).

No discussion of ecosystem depictions would be complete without mention of LUMA Partners’s various “LUMAscapes.” LUMA Partners (2015) has created a series of these types of illustrations. In the particular image cited, firms that participate in the distribution of mobile advertising are grouped together and arranged from “Marketer” to “Consumer” and color coded. The impression one gets from this depiction (and all their others) is the sheer number and complexity of the firms involved in the ecosystem. Indeed, an appreciation of the confusing state of the industry may be the ultimate goal of these illustrations. These depictions allow us to discover new and emerging market segments and representative firms to further investigate.
From the academic world, one deserves special note (Basole, 2009a; Basole, 2009b). Basole aggregated two main data sets of relationships between firms and assigned them to 14 different exogenous categories. He then analyzed the strength of relationships between categories and showed the evolution of the mobile platform, network, and device manufacturer ecosystem over time. Basole’s approach is closest to the type of analysis we envision for our ecosystem model. However, his depiction does not show the types of information collected by these firms, how that information is stored, or the procedures and law governing law enforcement access to that information. To understand the last item, we next discuss prior explorations of the legal landscape related to the mobile ecosystem and electronic surveillance.

Prior Exploration of the Legal Landscape
In contrast with the broad pictures that existing depictions of the mobile ecosystem offer, legal research in this area has focused primarily on developing specific doctrines or answering specific legal questions. Given the complex and confusing jurisprudence surrounding law enforcement searches, it is not surprising that many prominent legal scholars have proposed clarifying frameworks. For example, Orin Kerr has suggested that the Supreme Court has actually employed four different approaches to interpreting the expectation of privacy protected by the Fourth Amendment, each of which provides advantages in certain contexts. Other scholars have branched out beyond the Fourth Amendment to explain privacy rights more broadly. Solove (2006) developed a “taxonomy of privacy” to analyze and characterize harms to individual privacy from a wide spectrum of possible sources. While these and similar writings may illuminate the jurisprudence underlying surveillance of the mobile devices, apps, and services, their conceptual approach does not provide the detailed information necessary to relate specific data, devices, apps, and services to specific laws.

While the analyses discussed above have focused on providing broad theoretical bases for Fourth Amendment jurisprudence, other scholars have attempted to show how the law should be applied in certain narrowly defined circumstances. For example, Gershowitz (2008) has provided a thorough discussion of the permissible scope of smartphone searches incident to arrest. Some scholars have developed comprehensive analyses of particular legal doctrines across many different jurisdictions. Henderson (2006) performed such an analysis for the third-party doctrine. While these comprehensive reviews provide systematic explorations of specific legal issues, they are not intended to scrutinize the interaction between different legal doctrines.

Our approach is intended to provide guide to legal issues directly related to law enforcement searches of the mobile devices, apps, and services across a selection of state and fed-

---

1 The four approaches Kerr, 2007, describes are as follows: (1) the probabilistic model, (2) the private facts model, (3) the positive law model, and (4) the policy model. In the probabilistic model, the supreme court is most likely to recognize a reasonable expectation of privacy as the likelihood of public knowledge of the information in question becomes less likely. A higher probability of public knowledge results in a lower probability of recognizing a reasonable expectation of privacy. In the private facts model, the court considers whether the information in question is particularly sensitive or intimate and therefore worthy of constitutional protection irrespective of how it was collected. In the positive law model, the court relies on sources of law other than the Fourth Amendment (such as the law of trespass) to determine whether law enforcement acted in an unexpected way to collect the information in question. If law enforcement undertakes an action that would have been illegal for a member of the public, it is likely that it has violated a reasonable expectation of privacy in collecting the information. Finally, in the policy model, the court considers the effects of law enforcement practices on the broader community, rather than on the suspect. If the particular law enforcement action is the type of conduct that should be regulated by the Fourth Amendment, the court would find that it violated a reasonable expectation of privacy.
eral jurisdictions. As such, it is most closely related to the comprehensive, multijurisdictional reviews legal scholars undertake. However, rather than focus on one particular legal issue, we describe the set of legal issues that directly regulate the surveillance of mobile devices, apps, and services. This allows us to understand variations in the legal protections afforded to users of mobile devices both within jurisdictions and across jurisdictions.

**MIKE’s Approach**

MIKE aspires to improve on these approaches to depicting the mobile ecosystem. MIKE’s depiction of the mobile ecosystem is dynamic. Static depiction is a key limitation of many of the prior approaches, whether they are graphic or text based. Static depictions are brittle and need to be restructured as information is added. MIKE’s wiki-based approach overcomes this limitation.

MIKE defines conceptual relationships, as do most prior approaches. In MIKE, these relationships are also structural. Its use of semantic relationships in a wiki-based system enables more-powerful searches and inferences about the relationships between concepts.

MIKE integrates three views: technology, commerce, and law. Prior work typically depicts a subset of these views.

The prototype’s purely text and hyperlink approach forgoes graphic representations of information and relationships. Graphics provide powerful ways to understand information and relationships, and it should be feasible to extract and depict relationships captured in the prototype of MIKE graphically. This is a feature that an implementation of MIKE should consider adding.

**Architectural Overview of MIKE**

**Introduction**

Software and system architects often use ontologies to formally name and define entities and the relationships between them within a particular domain of interest. Ontologies organize complex information into a format machines can use to “reason” about the entities and their relationships and present that information to humans in a way that is easier for them to understand and is relevant to the context of a given search. This hides much of the complexity. A wiki is a website, implemented by a well-accepted software framework, that allows users to modify its content and structure directly from a standard web browser. The wiki software that runs such sites automatically tracks user changes and often has tools to protect against both inadvertent and malicious damage to the content and structure of the site. A tool that allowed us to define and navigate an ontology and provided wiki features, such as user editing, change tracking, and information protection, would meet all our architectural goals. Fortunately, the movement toward a semantic web infrastructure over the last several years has produced several “semantic wiki software” packages with just these attributes.

In selecting a software package for development of the MIKE prototype, we preferred a mature package with active support and good user acceptance. Therefore, we based MIKE on the Semantic MediaWiki extension of the free, open-source MediaWiki software used to power Wikipedia and many other popular websites (MediaWiki, 2014). While the basic MediaWiki software supplies the standard wiki features for reliable user editing, Semantic MediaWiki augments it with support for semantic web technologies to encode the relationships between
entities in our ecosystem, to search these relationships, to infer new knowledge from these relationships, and to display relevant contextual information on user-editable pages.

A brief overview of semantic web technologies may help with understanding how MIKE works. The current World Wide Web can be thought of as a web of documents (e.g., webpages) housed in servers around the globe. The connections between these documents are specified in the form of hyperlinks. While a hyperlink allows a user to navigate to a site where specific information can be found, it tells us nothing else about the relationship between documents and/or webpages. The goal of the semantic web is to allow users to define additional relationships between data objects (e.g., documents and/or webpages) to create a knowledge structure that allows computers to extract useful information automatically. Fundamentally, semantic web technologies work by linking data entities that contain content (a document or webpage) with relationships. Formally, each semantic statement links a subject to an object with a predicate (the relationship). Figure 4.1 provides an example.

The standard hyperlink between data contents simply indicates that something “has a link to” something else and provides navigation to the latter. No other real information can be inferred without examining the content of the data objects themselves. Semantic links carry more information. First, the data objects can represent any real-world entity, such as people, places, or ideas, not simply webpages and documents. Second, the semantic link allows us to specify that some entity has a specific type of relationship to another entity.

To illustrate this distinction, consider Wikipedia. One of its pages is a list of approximately 300 U.S. cities by population. Learning about all cities in Texas requires searching the table for all occurrences of “Texas.” A semantic wiki would define a relationship between city and state, allowing a single search for “Texas” that returns a table of only the cities in Texas.

In MIKE, entities are defined for (1) the types of collected data within the mobile ecosystem that are relevant to electronic surveillance, (2) the commercial entities that collect the data, (3) the laws that cover both the collected data and the firms that collect the data, etc. The semantic links in MIKE relate these entities. For instance, MIKE contains an object representing the “Gmail” mobile app, with a “has owner” relationship linking it to the object representing the firm “Google Inc.” This indicates that Google owns the Gmail app. The Google Inc. page in MIKE displays all the apps for which Google is the designated owner. This list of apps is not hard coded into MIKE but is dynamically populated when the page is brought up from a database query that returns the list of all objects with a “has owner” property set to

Figure 4.1
Difference Between Normal Hyperlinks and Semantic Links

![Diagram showing the difference between normal hyperlinks and semantic links.](RAND-R81482-4.1)
“Google Inc.” Adding a new app to the list is as simple as creating a new app object with the appropriate “has owner” relationship and refreshing the Google Inc. page. Similarly, if a product is acquired by a new owner, a simple change to the “has owner” property of that product is all that is required for it to be displayed on the new owner’s product list and deleted from the prior owner’s product list.

Much more complex queries based on the relationships stored in MIKE are possible. For instance, one can generate a display of all the companies that have access to email addresses and phone numbers (possibly from different apps) or of all firms that share GPS-based location data with third-party libraries. Thus, law enforcement professionals could use MIKE to discover all the commercial entities that may collect and retain an email address and the apps that they provide. In an investigation, this list of apps can be compared to a list of apps known to be on a victim’s (or suspect’s) phone to identify potentially relevant sources of electronic surveillance data about an email address.

Once entered, these semantic links can be used to infer new information. For instance, if firms A and B share data and if firm B shares data with firm C, the potential sharing of data between firms A and C can be automatically inferred and need not be manually entered into MIKE. Using the same inference capability within MIKE, legal statutes can be defined such that they apply in specific circumstances and can then display automatically in appropriate contexts. This means that there is no need to explicitly state which laws apply to which data coming from which app and that no single person needs to define all three. Instead, the objects that define the law, each app, and each data type (i.e., the context) can be defined separately, and MIKE automatically provides a link between an app and the governing law based on the data type used within the app.

The basic WikiMedia software package also provides a template capability. Pages within MIKE are based on templates that are simply another class of object that users create and modify to simplify entering information in a consistent way. Change tracking makes it relatively easy to roll back to prior versions of information and ensures that data are not “lost” and that mistakes can be reversed. This encourages users to update both the content and format of displayed information, keeping current with user needs.

Conceptual Elements of MIKE

Entities and Their Relationships

We have populated the MIKE prototype with the entities and relationships depicted in Figure 4.2 and further described in Table 4.1. For clarity, not all entities and relationships are shown, only those that describe the primary objects of the mobile ecosystem knowledge structure embedded within MIKE. We recommend exploring the MIKE prototype to see

---

2 Actual sharing of data may be different. For example, firm A may place restrictions on how firm B can further share data firm A provides. The potential sharing relationships are intended to guide an investigation search for sources of relevant data. However, as more details about sharing emerge, e.g., through a subpoena, it would be useful to capture this information to inform others that the data are, in fact, not shared. MIKE is a proof of concept, and this refinement is something that should be consider by an implementation of MIKE.

3 The notation for the entity-relationship diagram used in Figure 4.4 is described in Stewart, 2008. Rectangles represent classes of entity objects, and lines represent the relationships between them. Each relationship is labeled with a number referencing an entry in Table 4.1. The “crows feet” at the end of a link denote a one (or zero)—to-many relationship; a bar indicates a zero or one relationship. For instance, a firm can own multiple mobile apps, but a mobile app will only have a single owner.
examples of the entities and relationships. Our choice of entities and relationships is the foundation of MIKE and is one of the primary contributions of this research project. Our examples and working sessions with stakeholders using MIKE suggest that the foundation is sound. As MIKE evolves, as the need for a different entity or relationship becomes apparent, it can be added easily using the Semantic MediaWiki software.

A computer can read an entity-relationship diagram both forward and backward. On the leftmost side of the diagram in Figure 4.4, relationship 1 denotes that “multiple firms can be subsidiaries of a firm” and that “a firm can have multiple subsidiary firms.” Relationship 2 can be read as both that “a firm is located in multiple jurisdictions” and that “a jurisdiction governs multiple firms.” If we designate “Google Inc.” as a firm, the computer can reason that Google may have multiple subsidiaries and that both Google and each of its subsidiaries may be located in multiple jurisdictions. If the computer reads further, it will find (relationship 19) that a statute can be applicable in multiple jurisdictions and that a jurisdiction applies multiple statutes. If asked to display statutes affecting law enforcement’s ability to obtain user information from Google, the computer can easily produce and display a list based on the jurisdictions applicable to both Google and its subsidiaries. Using more of the map, it can winnow that list based on the legal context of data within a particular mobile app or third-party library owned by Google.

MIKE works by linking entities with relationships. While many of the entities shown above represent concrete things, like mobile apps or firms, other entities represent abstract but still real concepts, such as data types and legal process documents (such as a warrant). Other entities can be yet more abstract and refer to a collection of things. An important example of such an entity is the data context. In general, it is not enough to say that an app saves a par-
More information, such as where the data are saved, how often, and the presence of encryption, is needed if the data are to be accessed by and useful to law enforcement. This additional information, along with the type of data, is the data's context. The data type and the storage method, in turn, define the legal context of the data. By treating both the data context and the legal context bundles of information as entities in their own right, we can better understand the relationships between the data and its context.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Is subsidiary of</td>
<td>Firms can be owned by another firm.</td>
</tr>
<tr>
<td>2 Is located in</td>
<td>Firms can have locations in many jurisdictions.</td>
</tr>
<tr>
<td>3 Is engaged in activity</td>
<td>Firms can engage in activities directly or via one of their mobile apps.</td>
</tr>
<tr>
<td>4 Owns</td>
<td>Firms can own mobile apps and third-party libraries.</td>
</tr>
<tr>
<td>5 Has data context</td>
<td>Firms can control data contexts directly or through their mobile apps. Apps can have data contexts that specify the type of data collected and how it is secured, shared, and retained. Data an app collects can be transmitted to the app's owner but need not be.</td>
</tr>
<tr>
<td>6 Is transmitted to</td>
<td>The data within a data context can be transmitted to one or more firms or third-party libraries. This includes the firm that controls the data context (as specified by “Has data context”) and, potentially, other firms.</td>
</tr>
<tr>
<td>7 Incorporates library</td>
<td>Indicates that an app comes bundled with a third-party library. The same library may extract different data, depending on the permissions of the host application, so particular data contexts still need to specify the library as receiving the data.</td>
</tr>
<tr>
<td>8 Has encryption</td>
<td>Specifies what kind of data safeguards a data context has in place. This could include different forms of encryption or anonymization.</td>
</tr>
<tr>
<td>9 Has data</td>
<td>Specifies the type(s) of data stored in a data context.</td>
</tr>
<tr>
<td>10 Has resolution type</td>
<td>Specifies how often data are saved. This could be once (e.g., only at account setup), at regulator intervals, only the most recent use, or only when the user performs an action.</td>
</tr>
<tr>
<td>11 Has retention type</td>
<td>Specifies when data are deleted after collection. This could be never, a fixed time span, for the life of the account, or only transiently.</td>
</tr>
<tr>
<td>12 Has storage mode</td>
<td>Specifies where the data are located. Currently, the only options are local or remote, but future extensions could be more specific (such as specifying what cloud storage service is used)</td>
</tr>
<tr>
<td>13 Has applicable data type</td>
<td>Specifies what types of data are protected by the statute.</td>
</tr>
<tr>
<td>14 Has applicable storage location</td>
<td>Specifies whether the statute applies to data stored locally or remotely.</td>
</tr>
<tr>
<td>15 Has required legal document</td>
<td>Indicates what legal document is required to lawfully gain access to the data protected by the statute.</td>
</tr>
<tr>
<td>16 Has applicable situation</td>
<td>Specifies what legal situations this statute applies to (stop and frisk, search incident to an arrest, etc.).</td>
</tr>
<tr>
<td>17 Has relevant legal definition</td>
<td>Specifies that the statute contains a legal definition. A legal definition may be defined by the statute or elsewhere.</td>
</tr>
<tr>
<td>18 Has applicable activity</td>
<td>Specifies what activities a firm or app must be engaged in for the statute to protect the data within a data context.</td>
</tr>
<tr>
<td>19 Has applicable jurisdiction</td>
<td>Specifies the jurisdiction where the statute is valid.</td>
</tr>
</tbody>
</table>
can make appropriate links to other entities via relationships and allow MIKE to appropriately reason about the statutes that apply within in a particular data and legal context.

Our focus is on criminal investigations, not intelligence gathering. State and local police are ill equipped to deal with encryption. Typically, it is only important to know whether encryption is being used. If it is, police will generally not go to the effort of seeking warrant for something they cannot read. This shaped our unnuanced description of encryption (present or absent).

As the recent conflict between the Federal Bureau of Investigation and Apple Inc. over an encrypted iPhone illustrates, decryption requires a more-nuanced approach. Here, knowing something about the location and protection of encryption keys was central to any approach to decrypting the information sought. While murder was involved, this is primarily a national security case, with the bureau taking a leading role in the investigation. The technical sophistication of state and local law enforcement should improve with time, and it would be important to extend MIKE to capture details like this in a future implementation. The framework used to build MIKE is intended to support such a refinement. A future implementer of MIKE should consider this extension.

Our storage model draws a sharp distinction between local and remote storage of content. In practice, data stored remotely may also be stored locally. Our storage model was developed to support a proof of principle. It is unnuanced and should be expanded in any implementation of MIKE to allow indication of whether the same data could be found both locally and remotely. The framework used to build MIKE is intended to support such a refinement. A future implementer of MIKE should consider this extension.

Entities Hierarchies Within MIKE

Entities are grouped into different hierarchies that allow the semantic wiki software to store and search the information in MIKE more efficiently. Our current MIKE prototype uses the following groupings:

- The **knowledge hierarchy** includes data types.
- The **activities hierarchy** includes the types of services firms perform in the mobile ecosystem.
- The **products hierarchy** includes physical or digital things traded in the ecosystem.
- The **organization hierarchy** includes social groups.
- The **context** and **concept hierarchies** include concepts that link important information together.

We now describe each of these hierarchies.

**Knowledge Hierarchy**

The knowledge hierarchy organizes the data types in MIKE into seven categories:

- **Communication** data relate in some way to correspondence between individuals or an individual and a wider audience. Email, text messages, and Facebook posts are examples of communication data types.

---

4 Hierarchies in data management tools allow inheritance of properties between parent and child entities and play much the same role as class structures in object-oriented programming.
• **Financial** data are related to commercial transactions or wealth management. Examples include bank account information, credit scores, and purchase histories.
• **Interest** data can be mined to reveal subjects an individual is or has actively attempted to learn more about or devoted attention to. Examples include hobbies, browsing histories, and Facebook-style “likes.”
• **Location** data describe someone’s or something’s physical position. Examples include GPS location data, home addresses, home ZIP codes, and current metropolitan areas. Locations can be precise or imprecise in both geographic and temporal dimensions.
• **Personal identifiers** can be mined to reveal, and perhaps steal, the identity of a specific person. Some identifiers are anonymized, such as an encrypted version of a phone’s MAC address, while others are easily linked to an individual, such as a Social Security number.
• **Relationship** data are related to the social connections between individuals. Friends, call logs, and contacts are examples of data types that reveal a connection between two people.
• **Other knowledge** encompasses data types that do not fit well into one of the above categories. An example is health information. If this category becomes unwieldy, the inherent flexibility of MIKE’s architecture allows adding new knowledge categories to refine the contents of this category.

Any particular data type can be tagged as a member of multiple categories. For instance, a photo can reveal relationship information, but also location information (either via the background or embedded in the image’s Exchangeable Image File [EXIF] data). Calendar data usually reveal some interest of an individual (such as a yoga appointment) but could also be used to infer past or future locations and relationships with any invited guests. Thus, MIKE will include calendar apps when asked to list apps that reveal interests, locations, relationships.

**Activities Hierarchy**

This hierarchy organizes the activities that firms engage in. The purpose of the activity categorization is to link firms engaged in specific activities to legal statutes containing rules applicable to that activity. Sometimes, that linkage is direct, as in the applicability of the Communications Assistance for Law Enforcement Act to firms providing infrastructure for cellular services. At other times, the linkage is filtered by data type—for instance, the Health Insurance Portability and Accountability Act applies only to firms engaged in data-centric activities for health-related data.

Activities are divided into the following categories:

• **Advertising** covers activities meant to inform or persuade a target audience. Examples include serving advertisements and operating a real-time ad exchange.
• **Data-centric** activities involve the accumulation, exchange, or analysis of data. This activity covers firms that provide backup services or web traffic analytics or that operate a data brokerage.

---

5 EXIF is a standard format for storing information in digital photography image files. Digital cameras use the EXIF annotation to record information about the conditions under which the camera was operated (shutter speed, exposure, if a flash was used, etc.) and the date and time the image was taken. Many cameras embedded in mobile phones also store GPS information to record the location where an image was taken.

6 The Communications Assistance for Law Enforcement Act is a wiretapping law passed by the U.S. Congress in 1994.

7 The Health Insurance Portability and Accountability Act was passed by the U.S. Congress in 1996.
• *Developer support* includes the provision of developer tools, such as APIs, software development kits, and libraries. Rather than serving end users, firms engaging in this activity provide services to developers.

• *Infrastructure* is the provision of digital transport or storage services to other entities. Cellular service providers engage in this activity.

• *Mobile services* are the development and provision of mobile apps or websites, e.g., the Google Play store. The focus is on services to end users. This is the broadest type of activity.

Just as with the knowledge hierarchy, a firm can engage in multiple activities. For instance, Google engages in all the activity categories.

**Product Hierarchy**

The product hierarchy organizes the goods that are produced or consumed by organizational entities within the mobile ecosystem. Some goods are physical, like a cellular phone, others are executables, like an app or operating system, while still others are purely abstract, such as statutes or legal process documents. There are three categories for goods:

• *Digital goods* is the largest category and includes mobile operating systems, third-party libraries, third-party app program interfaces, and the mobile apps.

• *Legal goods* organizes federal and state-specific statutes and legal process documents.

• *Physical goods* organizes the physical platforms, such as the cell towers and the makes and models of phones.

**Organizational Hierarchy**

The organizational hierarchy is the least developed of the classes and is used to organize the social groups, or organizational entities that produce goods and services. Currently, the organizations are categorized as either firms or courts. In the future, we anticipate that other categories may be added, such as consortiums or legislative bodies. Consortiums are an increasingly popular organizational structure engaged in developer support activities within the mobile ecosystem.

**Context and Concept Hierarchies**

Although currently implemented separately, we recommend that these two hierarchies be refactored into a single concept hierarchy within MIKE. There are three categories of context in MIKE: a data context, a legal context, and the law enforcement situation, each of which describe conditions for understanding critical aspects of collected data items. Currently, the data context is categorized under the context hierarchy and the legal context and law enforcement situation are categorized as legal concepts under the concept hierarchy. We recommend

---

8 We are using a narrow definition of *product* to refer specifically to goods, not the broader economic definition of *product* as information, goods, and services. Within MIKE, information products are categorized under the knowledge hierarchy and service products are categorized under the activities hierarchy.

9 The prototype of MIKE was implemented using separate hierarchies. This recommendation is based on feedback from multiple stakeholders who had the opportunity to use the MIKE prototype.

10 The *law enforcement situation* refers to the circumstances surrounding law enforcement’s request for information; “search incident to an arrest” is an example of a law enforcement situation.
that, going forward, the data context be organized under the technical concepts of the concept hierarchy. The final hierarchy would then have three categories:

- **Technical concepts** currently includes the data context, describing the conditions under which a specific data item is collected and/or stored, but can be expanded to include additional concepts related to understanding technical issues.
- **Legal concepts** includes both the legal context defining the conditions under which data can be legally accessed and the law enforcement situation defining the typical situations under which law enforcement would find that information to be of use.
- **Business concepts** currently includes one the revenue model of the business but can be expanded to include other business related concepts in the future.

All collected data items are assumed to have both a data context and a legal context, with some conditions overlapping the two contexts.

**Miscellaneous**

Since MIKE is implemented using open-source software supporting web apps, it also allows use of the standard search capabilities incorporated in these tools. The hierarchies we just described provide a conceptual structure for organizing information. Information can also be tagged and, as with any tagging, benefits from using a standard lexicon. As is discussed elsewhere in this report, MIKE uses drop-down menus in fields that serve as tags. Drop-down menus, when available for a field, are displayed by typing a down arrow in the field. The combination of the two ideas enables qualified searches. A qualified search is launched by typing “concept: tag” in the search field, e.g., “context: location.”
APPENDIX
Stakeholder Reactions to the Wiki

As part of the development of MIKE, we consulted with many different stakeholders to obtain feedback and criticism. We selected these stakeholders opportunistically, based on their knowledge about electronic surveillance of mobile technology, and they should not be considered typical or representative. We talked with representatives of three major stakeholder groups: law enforcement agencies, industry groups, and policy advocates. We asked these representatives to identify current problems with how electronic surveillance of the mobile app ecosystem is conducted, what characteristics they would consider important for a solution to these problems, and their reactions to MIKE. The results of these conversations can be found in the tables at the end of this appendix. We used the results of the stakeholder interviews to guide the development of MIKE. Our goal was to ensure that MIKE would address as many concerns about electronic surveillance as possible, and would be easy for a broad range of stakeholders to use. These conversations will also guide our future work as we analyze potential avenues for transforming MIKE from a prototype to a usable (and used) piece of software.

In our discussions, we found that all stakeholder groups were concerned about the complexity of the mobile app ecosystem and about the significant costs associated with navigating both the information within the mobile app ecosystem and the legal framework governing access to this information. Law enforcement entities were particularly concerned about the multiplicity of communication pathways now available to suspects. Rather than relying on telephones or pagers alone, criminal organizations now have a host of messaging apps available, and each app may have different encryption mechanisms or may be produced by a company that is more or less hospitable to law enforcement requests. Industry stakeholders were concerned that the many sources of information and various rules governing access to this information may lead to inefficient requests for information, thus requiring them to expend unnecessary time and money searching for information. Policy advocates were interested in a way to demonstrate the complexity of the mobile app ecosystem and associated legal framework to policymakers, to facilitate better policy decisions and create clearer rules about when and how law enforcement can access the information within the mobile app ecosystem.

Both law enforcement agencies and industry representatives suggested that development of best practices for requesting information would improve requests for information within the mobile app ecosystem. While law enforcement officers do share their experiences informally, there is no formal mechanism for law enforcement officers to share their experiences obtaining particular types of information from particular companies. Representatives from law enforcement agencies discussed the establishment of a shared knowledge base as one of their highest priorities, describing it as a “real and imminent” need. Industry representatives viewed a shared knowledge base as both a way to ensure that they receive well-tailored information requests
and a way to ensure that responses to requests meet industry standards. Such an information-sharing mechanism would also help educate new start-ups on the customary way to respond to law enforcement requests. This is a particularly pressing need as law enforcement officers seek information from new sectors of the mobile app ecosystem. While telephone companies and operating system developers are probably accustomed to law enforcement requests, mobile app developers may not have had these experiences before. The development of best practices would therefore aid both law enforcement agencies and industry representatives.

All stakeholder groups we consulted viewed MIKE as a potentially useful way to explore the mobile app ecosystem, help them understand the laws regulating law enforcement access, and share information about best practices for requesting information. MIKE could be used to help law enforcement officers identify new sources of information and would ensure that law enforcement officers received consistent information, thus minimizing the burden on the companies providing the information. To that end, representatives from all stakeholder groups discussed the need for MIKE to be expanded to include new applications and jurisdictions, updated to reflect the development of new laws, and curated to ensure accuracy. However, not all groups agreed on how to prioritize expansions and updates to MIKE. Law enforcement officers were primarily concerned with ensuring that popular apps and apps that could be particularly useful to criminal organizations were well represented. In contrast, policy advocates wanted a broad range of apps represented, rather than many examples of apps that fundamentally perform the same functions. Their perspective was that a range of apps would better allow them to use MIKE to demonstrate the complexity and variance of the mobile app ecosystem to policymakers.

While information sharing was viewed as an incredibly important tool for improving electronic surveillance, both law enforcement agencies and industry groups were deeply concerned about who the information would be shared with. If information about the best practices for requesting and obtaining information from law enforcement were to be available to the general public, criminal organizations and individuals might adapt their communication habits to render these best practices obsolete and ineffective. Additionally, potential users of knowledge-sharing mechanisms might be dissuaded from providing information if that information were available to the general public because this might limit the effectiveness of their usual investigative techniques.

These tensions shape how the content of an implementation of MIKE would be entered and how MIKE would be used. Managing these tensions requires characterizing the shared interests of stakeholder groups and the sources of mistrust between them. It also requires understanding the choices that can be made about how to operate and implement MIKE and how stakeholder groups would respond to alternative ways to implement MIKE. Such analysis is beyond the scope of the document.

However, it might be possible to provide some public access to information about law enforcement searches of the mobile app ecosystem while still respecting the concerns of law enforcement and industry representatives. One industry stakeholder suggested that a version of MIKE without proprietary law enforcement information could be made available to the public. MIKE could therefore be used as a tool for policymaking and public education without threatening law enforcement’s ability to conduct effective investigations. A limited, public-access version of MIKE would seem to fulfill the needs of policy advocates. From their perspective, even limited access to MIKE would provide a helpful tool to demonstrate policy issues arising from the mobile app ecosystem.
The results of our conversations with stakeholders are significant not only for what was said but also for what was not said. We discovered that stakeholders did not often explicitly disagree or present contradictory viewpoints. Instead, organizations involved in electronic surveillance largely voiced similar concerns, regardless of whether individual was from law enforcement, the telecommunications industry, or the policy community. Most people expressed that the important policy determination was not whether society should value privacy or public safety more highly but rather how to simultaneously provide meaningful privacy protections and facilitate law enforcement investigations. While stakeholder groups may disagree about the optimal trade-off between privacy and public safety, all agree that a compromise is necessary. While this is certainly not the end of policymaking in the mobile app ecosystem, it is certainly an auspicious beginning.
<table>
<thead>
<tr>
<th>Concern</th>
<th>Law Enforcement Stakeholders</th>
<th>Industry Stakeholders</th>
<th>Policy Stakeholders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practical difficulties in</td>
<td>• Because suspects have multiple ways to communicate, it is not clear in advance what communication methods law enforcement should target.</td>
<td>• Industry representatives discussed the fact that they have small teams dealing with high volumes of law enforcement requests.</td>
<td>• Policy groups discussed their concerns that standards for accessing information were complex, and traditional legal definitions may not be adequate in this context.</td>
</tr>
<tr>
<td>obtaining information</td>
<td>• Electronic surveillance can be expensive and burdensome, particularly for smaller agencies.</td>
<td>• Start-up firms need to be educated in how to respond to law enforcement requests.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Encryption may be an issue, depending on the type of information being sought.</td>
<td>• Inexperienced law enforcement officers formulate ill-composed orders that lead to excessive back-and-forth and inefficiency.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Some corporate entities are resistant to law enforcement information requests.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complexity regarding availability of information</td>
<td>• Law enforcement stakeholders discussed the different types of information available to them (e.g., real-time versus historical data) and the different technical and legal barriers to obtaining these types of information.</td>
<td>• When given a template to assist in making an information request, law enforcement officers tend to ask for all available information types, creating a burden to the entities that hold the information.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• A particular type of content may be “scattered” all over the various applications on a smartphone, making it difficult to determine where it is stored.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inter-jurisdictional issues</td>
<td>• Law enforcement expressed their concerns regarding the ease of transporting mobile devices across state lines.</td>
<td>• Industry groups with an international presence noted that different countries may have conflicting requirements concerning data retention and storage.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Privacy rights</td>
<td>• While the technological tools available to law enforcement may give them broad access to mobile devices, some representatives expressed their concern that these searches are done in a way that respects constitutional rights.</td>
<td>• Industry groups are primarily concerned with protecting their customers from hackers, i.e., protecting personally identifiable information.</td>
<td>• Policy groups discussed their interest in having clear standards for obtaining information, to protect privacy rights.</td>
</tr>
</tbody>
</table>
### Table A.2
#### Necessary Components of the Tool

<table>
<thead>
<tr>
<th>Component</th>
<th>Law Enforcement Stakeholders</th>
<th>Industry Stakeholders</th>
<th>Policy Stakeholders</th>
</tr>
</thead>
</table>
| Understanding what data are available and from where | - Members of law enforcement expressed their interest in knowing where they can get information most quickly.  
- The goal of a tool should be to help law enforcement officers find out what the suspect is doing, rather than understand where data are located. | - Some industry representatives described how law enforcement might benefit from being able to identify additional sources of information, including mobile apps. | - Policymakers need to be able to understand and appreciate the complexity of the mobile app ecosystem. |
| Protections for individual privacy | - Law enforcement officers expressed their interest in having clear rules regarding protections for individual privacy to be able to operate within the rules.  
- Offers also expressed concern about interactions between state and federal laws. | | - Privacy protections should be in a clear, simple, and certain.  
- The goal is to balance privacy protections against law enforcement access to information, not to prohibit law enforcement access entirely. |
| Sharing knowledge and best practices | - Setting up a shared knowledge base was described as a “real and imminent” need.  
- Law enforcement officers were interested in being able to contact other officers who had obtained similar information in the past.  
- The goal is to make sure that officers do not have to reinvent the wheel every time they need information, particularly for smaller agencies or agencies that request information less frequently. | | | 
| Transparency and confidentiality concerns | - Any tool that requires users to create accounts should include an option to keep contact information private.  
- Some law enforcement officers expressed their concerns about allowing non-law enforcement personnel to access a tool created for law enforcement. | | |
<table>
<thead>
<tr>
<th>Reaction to</th>
<th>Law Enforcement Stakeholders</th>
<th>Industry Stakeholders</th>
<th>Policy Stakeholders</th>
</tr>
</thead>
</table>
| **Perceived usefulness and likelihood of use** | • MIKE would help law enforcement officers explore new avenues for obtaining information.  
• One law enforcement officer stated that “there are times when it could be extremely valuable.”  
• Law enforcement officers described themselves as very likely to use MIKE, provided that it contained useful information.  
• One law enforcement officer particularly liked the fact that there were multiple ways to get the same information, making it more user friendly.  
• One officer described MIKE as a “pretty darn good” project and expressed his interest in obtaining MIKE for his office’s use. | • They described MIKE as being potentially valuable because it would provide law enforcement with consistent information.  
• MIKE might clarify the law and help law enforcement officers formulate requests efficiently, ensuring that law enforcement and commercial entities “speak the same language.” | • MIKE was perceived as useful in informing policy.  
• MIKE may be particularly useful as a tool for demonstrating the complexity of the problem space to policymakers. |
| **Controlling access to MIKE** | • Law enforcement officers expressed their willingness to share information within the law enforcement community, although they were very concerned about the possibility of this information being shared outside the law enforcement community. | • One industry representative suggested maintaining different versions of MIKE—one for law enforcement agencies and one for the general public.  
• Concerns were raised that any tool might assist cyber criminals if it were in the public domain. | |
| **Criticisms and suggested improvements** | • They would like to see improved filtering capabilities to make it easy to focus on relevant information.  
• One law enforcement officer mentioned that “drop-down” menus could facilitate uniform data entry. | • Industry representatives noted that MIKE would have to be curated and kept up to date to maintain its usefulness. | |
| **Suggestions for expansion and maintenance** | • A curator would probably be required.  
• Investigators will probably not have time to enter additional data.  
• Representatives from larger law enforcement agencies suggested that their technical analysts might be able to update information. | • Policy issues are not immediately apparent from data contained in MIKE, leadership needs to identify relevant policy issues. | • MIKE does not need to contain many apps to be useful as a policy tool. However, it would be more useful if it contained representative examples of many different types of apps (rather than many examples of apps that do similar things). |
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>application programming interface</td>
</tr>
<tr>
<td>app</td>
<td>application</td>
</tr>
<tr>
<td>ECPA</td>
<td>Electronic Communications Privacy Act</td>
</tr>
<tr>
<td>EXIF</td>
<td>Exchangeable Image Format</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>HTML</td>
<td>Hypertext Markup Language</td>
</tr>
<tr>
<td>ID</td>
<td>identification</td>
</tr>
<tr>
<td>IEEE</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>iOS</td>
<td>Apple’s operating system for mobile devices</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>MAC</td>
<td>media access control</td>
</tr>
<tr>
<td>MIKE</td>
<td>Mobile Information Knowledge Ecosystem</td>
</tr>
<tr>
<td>POI</td>
<td>person of interest</td>
</tr>
<tr>
<td>POP</td>
<td>Post Office Protocol</td>
</tr>
<tr>
<td>SMS</td>
<td>Short Message Service</td>
</tr>
<tr>
<td>URL</td>
<td>Uniform Resource Locator</td>
</tr>
</tbody>
</table>
Glossary

ad exchange service Firm that allows websites to “sell” advertising space to the highest bidder in real time. The seller sends information regarding its site, where the user is located, or other information known about the users (perhaps from their recent browsing history). Within milliseconds, advertisers respond with a “bid” to fill that space. The ad from the highest bidder is then displayed to the user.

anonymization A process of irreversibly severing collected information from the identity of the person who contributed the information. The goal of anonymization is to prevent any future reidentification, even by the data collector, under any conditions. Encrypting identifying information and then destroying the key is one method of anonymization.

application (app) Software installed on a device that provides the user with a specific capability (calendar, browser, text messaging service, interactive games, social network access, etc.). Apps can be started, stopped, and configured for use. Most apps require access to remote services to deliver the promised capability, although some provide stand-alone capabilities.

application programming interface (API) An interface that enables an app to exchange data with other software, usually to obtain services. APIs can be highly proprietary and specialized or public and generalized. The later are often referred to as “open” or “standard” APIs. The more generalized an API is, the more data it will need to exchange, perhaps increasing the likelihood that applications and services will exchange identifying information.

business ecosystem A term of art based on the metaphor of a biological ecosystem. A business ecosystem arises when firms provide products or services that are mutually dependent on products or services provided by others, often without formal contracts, business agreements, or architectural standards. The ecosystem’s operation is constrained by legislation; regulation; economic forces; and a mutual need for compatible interfaces, norms, and standards of behavior.
call detail records  Metadata describing the source and destination of a phone call and its time and duration. This information is needed to place and bill for the call.

cell site  Physical location of an array of antennas that is part of a mobile phone and data network.

content data  The data that are meaningful to the two endpoints (source and destination) in a network transmission, e.g., the conversation taking place during a telephone call.

data brokerage  An entity that assembles comprehensive dossiers on individuals based on their online habits, social media posts, U.S. Census records, Department of Motor Vehicles, real property records, buying habits, etc. The brokerage then sells that information to other brokerages or businesses.

deep packet inspection  Extension of packet inspection to include examination of both identifying information and content data contained in protocol messages that implement the application layer of the internet.

device operating system  Software that provides the basic computing, communication, and storage services of a mobile device, e.g., Google’s Android, Apple’s iOS, and Microsoft’s Windows Phone.

dynamically extensive  The ability to add entities and relationships to a product without invalidating existing data and relationships.

electronic surveillance  Any law enforcement use of information transmitted by or stored within the mobile ecosystem, independent of how or by whom the information was collected. This includes collection and recording initiated by law enforcement, such as a wiretap, and law enforcement access to commercially collected and recorded information, such as location information or contact lists. This definition is broader than the one used in the rapidly evolving legal code.

encryption  Encoding a message or content in a way that allows only authorized parties to read it, using an algorithm and secrets known only to the authorized parties. This encoding transforms messages and content into a form unreadable to those who do not share the secrets.

envelope data  Metadata describing a telephone call (similar to information included on a paper envelope) that includes both caller and callee phone numbers; time of day; duration of call; the ID of the device used; and, sometimes, the geographic locations of the caller and callee.
Exchangeable Image Format (EXIF)  
A standard format for storing information in digital photography image files. Digital cameras use the EXIF annotation to record information about the conditions under which the camera was operated (shutter speed, exposure, if a flash was used, etc.) and the date and time the image was taken. Many cameras embedded in mobile phones also store GPS information to record the location where an image was taken.

executable  
A file that contains a set of instructions that a computer can run (execute) to perform a set of tasks. The instructions are written in a language easy for machines, but not humans, to read (often called machine language or assembly language). Executable files are often generated from source code written in a programming language humans can understand more easily.

geotag  
A piece of data embedded in a digital media file to indicate geographical information about the subject, usually latitude and longitude.

internet end point  
A computer, mobile device, or other technology (e.g., a printer) that acts as the source or destination of information transmitted via the internet.

internet protocol suite  
The rules that govern the transmission of packets of information on the internet. These rules are implemented via messaging protocols acting at the different layers of communication: application, transport, and network. A fourth layer—the link layer—is outside the internet protocols and specific to the physical connectivity (Ethernet, Wi-Fi, optical fiber, cellular). This layering has allowed the rapid convergence of smartphone and tablet capabilities. Of special import to law enforcement, the protocol messages contain identifying information (usually called a header) that is added and/or stripped at each layer. Each layer’s information may therefore be accessible from a different entity within the mobile ecosystem.

map (verb)  
The act of connecting and displaying interrelationships between entities in a multidimensional space. MIKE maps the following entities to each other:
- commercially collected data about mobile users that may be useful to law enforcement
- the commercial entities that have access to that data
- procedures that allow law enforcement to obtain legal access to that data
- law governing access to the data.
mapping tool Software that enables interrelationships to be defined and displayed dynamically and interactively in response to user inputs.

media access control (MAC) address A unique identifier assigned to network interfaces for use in the physical link layer of internet-based communication. Typically assigned at the time of manufacture, the address can be overridden by a local network administrator. If not overridden, the MAC address can be used to track the geographic movement of the device. To provide consumers greater privacy, Apple began using random, locally administered MAC addresses when its phones search for nearby networks (this feature was introduced with iOS 8).

metadata Data that describe other data, usually information to facilitate finding and working with particular pieces of data. When used in this document, the term generally refers to data that describe how, when, and by whom a particular set of data was used or collected.

mobile carrier Provides the “last mile” transport of cell phone calls, signals, and data via radio frequency to a mobile wireless device. The mobile carrier may also control other aspects of the network.

mobile ecosystem The firms (and their interconnections both physical and financial) that offer products and services to users of mobile devices. This includes device manufacturers, telecommunications providers, application developers, and the network and service providers that transmit, store, and process information. The ecosystem’s operation is constrained by legislation, regulation, economic forces, and a mutual need for compatible interfaces, norms and standards of behavior. (Within the context of MIKE, we consider the actual goods and services themselves to be part of the mobile ecosystem.)

onion routing (e.g., The Onion Router) A technique to make communication over the internet resistant to traffic analysis and eavesdropping, using successive layers of encryption to hide not only the content of the communication but also information identifying the source and destination of the communication.

open source A software licensing and distribution method in which users receive not only the executable software but also the source code. While licensing terms vary, users (including commercial entities) are allowed in many instances to modify the source code and incorporate it into their own products. License terms may restrict the redistribution of these derivative products.
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>packet inspection</td>
<td>A general term referring to examination of identifying information contained in the protocol messages that implement the network layer of the internet.</td>
</tr>
<tr>
<td>personally identifiable information</td>
<td>Data that can be used to establish the identity of an individual. National Institute of Standards and Technology Special Publication 800-122 (McCallister, Grance, and Scarfone, 2010) defines personally identifiable information as “any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual’s identity, such as name, social security number, date and place of birth, mother’s maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information.”</td>
</tr>
<tr>
<td>Post Office Protocol (POP)</td>
<td>An internet standard protocol used to retrieve email from a remote server. It is this protocol that allows you to access your email from any device using a simple web interface.</td>
</tr>
<tr>
<td>public switched telephone network</td>
<td>The wired network used to transmit voice telephone calls and data.</td>
</tr>
<tr>
<td>router</td>
<td>As used in this document, an internet router. More generally, any device that routes data from source to destination is a router. On the internet, the data being routed are formatted as a set of Internet Protocol (IP) packets, each containing a small amount of content data with a header designating the source and destination of the packet. Routers provide traffic direction functions, typically forwarding the packet from one router to another within the network until it reaches its destination. Each packet of a particular message is routed separately and may take a different path to the destination, depending on network traffic conditions.</td>
</tr>
<tr>
<td>service</td>
<td>As used in this document, applies to capabilities for communication (e.g., network access, encryption), storage (temporary or permanent), or processing of data that are not supplied by the application. These services can be resident on the mobile device or can be accessed remotely.</td>
</tr>
<tr>
<td>service provider</td>
<td>A corporation that provides a service to a mobile application, e.g., the ability to make a payment, archive data, display information on a geographic map, or determine restaurants in proximity to the phone.</td>
</tr>
</tbody>
</table>
shallow packet inspection  Extension of packet inspection to include examination of identifying information contained in protocol messages that implement the transport layer of the internet.

Short Message Service (SMS)  A service provided by mobile carriers that enables a user to send and receive short text messages addressed to a set of specific telephone numbers. The message may be stored and read later, making the communication asynchronous (i.e., the sender and receiver do not both have to be present at the same time and place for communication to occur).

smartphone  A mobile device that provides voice communication services and is capable of executing installed applications requiring computing power and (usually) internet access. Although originally designed to provide internet access through the mobile carrier, a smartphone can provide that same access through connection to a static wireless network (Wi-Fi).

software development kit  A set of software tools that aid software developers in the rapid development of applications for a specific software operating system or family of hardware devices.

stored content  Data retained for some length of time (but is not ephemeral) to implement a service (e.g., a message that is retained by an email service until deleted by a user) or that is stored by the user for archival purposes (pictures, contacts, passwords, etc.) Content may be stored on the mobile device itself, on service provider servers, or within a cloud service.

tablet  A mobile device that is capable of executing installed applications and providing internet access. A tablet generally has less computing capacity and storage than a personal computer but has more than a smartphone. Originally designed to access the internet through a static wireless connection, most can now provide that access through a mobile carrier. The distinction between tablets and smartphones is rapidly disappearing.

time-division multiplexing  A time-based method of synchronously transmitting independent voice calls between two end points in the public switched telephone network.

voice coding schemes (vocoders)  The methods used to digitally encode human speech for transmission over a network.

webpage  A document, typically written in plain text interspersed with formatting instructions written in Hypertext Markup Language (HTML). Webpages may incorporate links to other websites or pages. A web browser reads the HTML and displays the content to the user of a computer, tablet, or smartphone.
website  A set of related webpages that are usually accessed using a simple Uniform Resource Locator (URL), often referred to as a web address.

web traffic analytics  The measurement, collection, analysis, and reporting of a website’s use, usually to judge the popularity of a specific website or webpage. The data can also be used to assess and improve the effectiveness of a website.

Wi-Fi  A trademarked term that refers to the IEEE 802.11x interface standard. Like mobile phones, Wi-Fi communication uses the radio frequency spectrum, removing the need for wired connection between source and destination. The protocols and frequencies used for Wi-Fi distinguish it from the mobile phone (cellular) network. The range of Wi-Fi networks is typically short (~100 ft.)

wiki  A website that allows users to modify its content and structure directly from a standard web browser

wiki software  Software used on a wiki site to automatically track user changes, contributions, and corrections. Wiki software often has tools to protect against both inadvertent and malicious damage to the content and structure of the site.
References


Communications Assistance for Law Enforcement Act (CALEA), 1994.


Health Insurance Portability and Accountability Act (HIPAA), 1996.


Swire, Peter, and Kenesa Ahmad, “‘Going Dark’ Versus a ‘Golden Age for Surveillance,’” Center for Democracy and Technology website, November 28, 2011.


