JUSTICE

National Information Architecture

Toward National Sharing of Governmental Information

Representing Chief Information Officers of the States

NASIRE

Toward National Sharing of Governmental Information

This report is the result of a year-long partnership effort between NASIRE: Representing the Chief Information Officers of the States and the U.S. Deparment of Justice's (DOJ) Office of Justice Programs (OJP). This project was supported by Award No. 98-DD-BX-0067 awarded by the Bureau of Justice Assistance, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of NASIRE and do not necessarily reflect the official positions or policies of the Department of Justice.

NASIRE

Representing the Chief Information Officers of the States

TABLE OF CONTENTS

EXECUTIVE SUMMARY **4**

INTRODUCTION **6**

Acknowledgement 6
Associated Reports 6
Technical Terms 6
Overview 6

WHY: MOTIVATIONS FOR SHARING GOVERNMENTAL INFORMATION **7**

Motivation: Shared Information Is More Accurate 7
Motivation: Shared Information Is More Timely 7
Motivation: Shared Information Is More Complete 7
Motivation: Shared Information Is Less Expensive 8
Motivation: Answering Objections 8

WHO: ENTITIES SHARING GOVERNMENTAL INFORMATION **9**

Entities: Information Holders and Receivers 9
Entities: General Obligations of Holders and Receivers 9

WHAT: SHARED INFORMATION SCOPE AND REQUIREMENTS **10**

Scope: Dynamic Structured Information 10
Scope: Transactions and Data Collections 10
Scope: A Criminal Justice Example of Governmental information Sharing 10
Scope: Other Examples of Governmental Information Sharing 12
Requirements for Information Sharing 12
Implications of the Requirements 13

HOW: TELECOMMUNICATIONS INFRASTRUCTURE **14**

The Need For Open Internet Standards 14
Telecommunication Infrastructure Needed 14
Telecommunication Infrastructure In Place 14
Telecommunication Conclusions 15

HOW: DOCUMENT DEFINITION INFRASTRUCTURE **16**

The Need For Standard Document Definitions 16
Document Definition Infrastructure Needed 16
Document Definition Infrastructure In Place 17
Document Definition Conclusions 17

HOW: HIGH-LEVEL VIEW OF INFORMATION SHARING **18**

Regular Local Information Sharing 18
Regular Vertical Information Sharing 18
Occasional Sharing 18

WHEN: ACTION PLAN FOR THE NEXT 24 MONTHS **19**

Strategic Directions 19
General Strategy 20
Next Phase Activity Areas 20

GLOSSARY **22**

Executive Summary

In 1998, NASIRE, which represents the Chief Information Officers of the States, and the U.S. Department of Justice's (DOJ) Office of Justice Programs (OJP) began a cooperative, grant-funded effort to facilitate governmental information sharing via a national information architecture. This project is part of OJP's larger, ongoing Integration Initiative. It was supported by a one-year grant from the Bureau of Justice Assistance (BJA) to NASIRE. Additional information on this project can be found at NASIRE's web site.

Architecture, sometimes called information architecture, is the name given collectively to those characteristics of a network, operating system and/or application program which facilitate information sharing. Local architecture, state architecture and national architecture simply define the geographic scope of information sharing being considered. This report is focused on a *national governmental architecture*, that is, the nationwide sharing of information collected by governmental agencies at all levels.

Inter-agency sharing within a single locality is part of nationwide sharing; local agency sharing with its statewide information system is part of nationwide sharing; local and state branch or agency sharing with federal agencies is part of nationwide sharing. Local or state agencies sharing with companies, organizations or individuals is part of nationwide sharing. The concept is very broad. Sharing between federal agencies is excluded from our consideration.

Each instance of governmental information sharing involves two **sharing entities**. The *holder* entity is always a governmental unit; the *receiver* entity may be governmental or private (e.g., private attorney, day care center or private citizen). The holder must be willing, able and entitled to provide the information; the receiver must be able to demonstrate a need for the information, be able to receive it, and agree to abide by usage rules set by the holder. No holder is forced to share.

The **business case** for such an architecture rests on four foundations. First, shared information is *more accurate*; it is collected once and used many times, thereby avoiding the misunderstandings and keying errors associated with multiple collection. Second, shared information is *more timely*; it can often be made available instantly rather than waiting for a separate collection effort. Third, shared information is *more complete*; information from multiple sources can be assembled into a full description. Finally, shared information is *less expensive*; it costs much less to store data and send it to another user than it does to collect it again.

Our focus is on sharing **dynamic structured information**. (Organization charts and mission statements are *static* and outside our scope; wanted person reports and parole status reports are *dynamic* and inside our scope.) Structured means that the information is rigorously defined (e.g., date of birth in a standardized format) rather than free-text at the point of exchange.

Our focus is on information used in **transaction reporting** (e.g., a police department reporting an arrest to a prosecutor) and **data collections** (e.g., monthly summary reports) and intentionally excludes batch reporting (i.e., deferred transmissions of groups of transactions).

The shared information may be character-based, photographs or graphics, fingerprints, page images or facsimiles, or basically any information which can be reduced to a form that can be sent between computers via a telecommunications network.

Some sharing will be between governmental entities which have a long-term sharing **relationship** and high-volume sharing requirements tightly linking their information systems, for example a prosecutor case management system and the court case management system for the same jurisdiction. Other sharing will be between governmental entities which seldom interact, for example a prosecutor in Alabama who desires case information concerning a subject's previous prosecution in Indiana. Still other sharing will be of a one-time nature, for example a new home buyer wanting to check a sexual predator file relevant to her neighborhood. The architecture must support this *range of sharer profiles*, balancing ease of use with semantic precision and high volume for the various categories of user.

Information sharing requires several different **modalities**. Some information systems need the ability to push pre-agreed information to another governmental branch or agency (e.g., prosecutor system files charges to a court system). Other information systems need the ability to *pull* pre-agreed information (e.g., arrest booking system automatically retrieves subject's criminal history record). Some information system users need the ability to *query* outside systems, and these systems need the ability to *respond* to queries (e.g., check for stolen articles). Some information systems need the ability to *publish* notifications that their database content has changed in a way meaningful to a particular user, and the user needs a way to subscribe to such a notification service (e.g., parole officer to be notified if her client has been arrested). The national architecture must be rich enough to support all these modalities.

In order for a national architecture to make sense and be defensible, it must be based on **open standards**, and it must be attuned to the technological direction of information technology, especially the *Internet* and Internet-like solutions. The report identifies the major Internet standards applicable to this architecture.

Nationwide information sharing requires a nationwide **telecommunication infrastructure**, using existing facilities where available and providing new facilities where necessary. The telecommunication infrastructure must be based on open technical standards, for reasons of vendor-independence and ease of evolution over its life cycle. The open standards adopted should be those associated with the Internet, for reasons of broad interoperability and availability of a broad pool of technologically sophisticated labor.

Much work has already been done. Most states have begun to build statewide Internet-like networks, and the major networks used for state/local-to-federal communications have already made this transition. There is widespread agreement on the technical direction to be taken, and the task is one of implementation rather then persuasion to adopt.

Nationwide information sharing requires a nationwide **sharing vocabulary** (i.e., definition of terms); this need not lead to a common vocabulary for *internal* use in any governmental branch or agency, locality, state or federal entity. Several states have begun the task of constructing statewide sharing vocabularies, but there is much to be done in extending this work to nationwide scope.

Nationwide information sharing requires the adoption of a set of common **sharing documents** (e.g., an agreed inquiry format and response format for stolen-property checks). A few states have started the work to define statewide document sets, but there is much to be done in extending this work to nationwide scope. The open standards needed to define and enforce the content of the sharing documents have only recently been adopted, and widespread concurrence in their use for nationwide information sharing has not yet been obtained.

The nationwide sharing vocabulary and nationwide sharing document specifications do not place any limitations on the internal database structure or internal document specifications of any system participating in the nationwide sharing. Individual agencies have complete freedom of system design, limited only by whatever state and local laws and policies may apply.

There is a great deal of information sharing of the state-to-federal and local-to-federal kinds, mostly done using agency-specific vocabularies and non- standard document coding methods. The early and widespread translation of these sharing instances to the new standard-based methods will provide useful experience for IT professionals at all levels of government, as well as lower maintenance costs in the future.

Nearly all the components of a nationwide architecture are already in place one place or another, but not tied together systematically nor deployed widely. Several states have begun to analyze vocabularies used at the local and state levels with an eye toward making them compatible. Most states have some experience with Internet-like telecommunications and electronic documents, and many have experience with electronic signatures, data encryption and other technical dimensions of the national architecture. There is no dispute about the major directions in which work must be done to achieve nationwide information sharing. Nevertheless, reaching a national consensus and deploying the nationwide architecture to governmental entities large and small at all governmental levels will occupy the attention of IT professionals for the next decade. The report provides a very high level view of efforts which could be undertaken immediately and completed within two years, which would advance the cause of nationwide information sharing.

Surveys have been conducted of the states' Chief Information Officers (CIOs) who have an all-agency perspective on Information Technology, and of the states' Control Terminal Officers (CTOs) whose perspective is focused on Information Technology for law enforcement and criminal justice purposes. Substantial majorities of both groups believe that use within their areas of Internet-like technology, electronic documents, electronic signatures, Internet use and digital signatures will grow in the next five years. In all cases the CTO majority is smaller than the CIO majority in these beliefs; this may reflect merely a more conservative mind-set among CTO's, or perhaps the facts that only very recently have federal criminal justice networks changed to Internet-like technologies, and only very recently have federal requirements for law enforcement management control over information systems and networks connected to the FBI-run databases been eased.

The achievement of nationwide governmental information sharing does not depend on any future technological breakthroughs; present technology is sufficient. It does not depend on changing the hearts and minds of present practitioners; there is broad consensus on the desirability of such sharing and on the technological directions to be taken. It depends on leadership, on funding, on participation, and on patience. But what major advance does not depend on these factors?

Recommendation of the criminal justice community as the pathfinder for governmental information sharing is based on the high potential for intra-locality sharing opportunities and on the existence of extensive local-to-state and local-to-federal information sharing that can be migrated to the new technological model. Careful evaluation of pathfinder projects, coupled with active observing of the process by governmental agencies outside the criminal justice community, will permit safe and efficient national deployment of the sharing technology.

Acknowledgement

Special thanks are deserved by the members of NASIRE's 1999 Information Architecture Committee who volunteered many hours preparing this report. Those members included George Boersma, chief information officer, State of Michigan; W. Michael Hale, chief information officer, State of Georgia; Don Heiman, chief information technology officer, State of Kansas; Steve Kolodney, director, Department of Information Services, State of Washington, David J. Roberts, deputy director, SEARCH; Aldona K. Valicenti, chief information officer, Commonwealth of Kentucky; Richard C. Webb, chief information officer, State of North Carolina; and Gerry Wethington, director, Information Systems Division, Missouri State Highway Patrol.

NASIRE would also like to mention the important contributions made by David Gavin, chair, CJIS-APB; John Greacen, state court administrator, State of New Mexico; and Col. Michael D. Robinson, chair, Global Criminal Justice Information Network Advisory Council (GAC).

In addition, several corporations provided important technical context for the effort by attending a one-time meeting to provide an industry point of view. The meeting was helpful, but neither the individuals nor their firms they represented have any responsibility for the content of this report. Those who attended the meeting included Rich Bongiorno, Unisys; Alan Brown, Sterling Software; Steve Deatrick, PRC Inc.; Marty Dunning, Sun Microsystems; John Featherstone, TRW; Ralph Ioimo, SAIC; Nick Pattakos, Oracle; Mauro Regio, Microsoft; and John Simone, IBM.

Robert L. Marx, senior system specialist, SEARCH, served as consultant to the committee.

The federal project monitor was José McLoughlin, program manager, Bureau of Justice Assistance.

Elizabeth Miller, association director, NASIRE, and Chris Dixon, assistant director, NASIRE, provided staff support for the project.

Associated Reports

The following reports were produced under the auspices of this project and are available at the NASIRE web site. Data from these reports is referenced in this report.

- Poll Data: Mid-Year Conference Interactive Polling Results, March 30, 1999: Data from real-time survey of 123 attendees at the conference;
- Poll Analysis: First Analysis of Opinion Data, April 1999: Analysis of data from the real-time survey;
- CIO Survey: Survey of State Chief Information Officers: Survey instrument and response summary;
- CTO Survey: Survey of State Law Enforcement Chief Terminal Officers: Survey instrument and response summary.

Technical Terms

The use of technical terms is unavoidable in a report such as this. Please consult the Glossary at the end of this report when a technical word is encountered. In a few cases we have used a term while intending a definition consistent with but more narrow than the common definition; in those cases the Glossary provides the intended narrow definition.

Overview

This report is organized into these broad sections:

WHY: Motivations for sharing governmental information;

WHO: Entities sharing governmental information;

WHAT: Shared information scope, general requirements and transfer types;

HOW: Infrastructure in place and infrastructure required;

HOW: High-level view of information sharing;

WHEN: Action plan for the next 24 months.

WHY: Motivations for Sharing Governmental Information

Sharing governmental information is not a new concept. When one fills out a multi-part paper form for a governmental agency, very often the various parts are destined for different agencies, or at least to different functional or geographical groups within a single agency. Some states, under some circumstances, require that a copy of the federal tax return be attached to the state tax return. Paper parking tickets issued by a local police department are often also sent to the cognizant court. Paper building permits are sent to the local assessor. The issue is not whether or not sharing will occur, but whether *generalized* technological methods will be developed to make information sharing easier, faster, less expensive and more prevalent.

It is in the public interest[1] for governmental agencies[2] to share information with other governmental agencies, private organizations and the general public. A few reasons which are broadly applicable are:

- Shared information is more accurate;
- Shared information is more timely;
- Shared information is more complete;
- Shared information is less expensive.

Motivation: Shared Information is More Accurate

The alternative to sharing information is for each governmental branch or agency, organization and private party with a right to and need for the information to collect the information separately and redundantly. But each time information is keyed into a database, the potential for inserting errors into the database increases. By removing redundant data collection and computerization, each governmental branch or agency can devote resources to increasing the accuracy of the information when first captured. Transfer of the information to other agencies, organizations and persons can be done essentially error-free with suitable computer technology.

Motivation: Shared Information is More Timely

Decisions made with reliance on old information are often bad decisions. The decision maker wants to know: is the person on parole *now*, in school *now*, receiving financial support *now*, licensed to drive *now*, is the job available *now*, have the goods been shipped *now*, has the accident scene been cleared *now*.

Systems that cannot deliver timely information often implement exception-processing to ameliorate the system deficiency. When one sees inquiries received by facsimile or telephone, and responded to in some non-standard method of priority processing, one can usually conclude that the regular methods of retrieving data are not meeting the timeliness requirements for proper decision making.

In very many situations, it adds no value at all to reduce processing time from six weeks to two weeks, but it often adds enormous value to information when one can provide same-minute response to *all* inquiries, and decision makers can adjust their own processes in the reasonable expectation that they will routinely get such a service level. It is from this reengineering of business processes that a major component of system efficiency is gained. Providing information too late but not as late as it used to be is not a half-solution; it is no solution at all.

A powerful form of information sharing is anticipatory sharing. Suppose that a person is a student and a welfare recipient and a parolee; a change in student status (e.g., failure to appear for classes) may affect the probation status (if being a student is a condition of probation), the change in probation status (e.g., re-incarceration) may affect welfare status (if jail inmates are not eligible for benefits), and the change in welfare benefits may change the student status (e.g., becoming eligible for reduced-cost meals). The most effective information sharing may well be for the governmental branch or agency which is recording a change in status to report that change to all other agencies that have an interest in that person, rather than each agency routinely interrogating every other agency as to whether some aspect of the subject record has changed. The result of such just-in-time information sharing goes beyond enriching a decision, to actually triggering the need for a decision. A second example: many states have laws or regulations that require a "clean" criminal history record when a person applies for certain jobs or licenses; would the public good be served if the employing or licensing agency were automatically informed of certain criminal history events while the person is employed or licensed? Such notification does not occur immediately today because the technical method implemented to accomplish it does not provide real-time updates.

Motivation: Shared Information is More Complete

Arrest information is valuable. Arrest information coupled with prosecution information, court information and correctional information is more valuable. Local criminal history information is valuable but state-level aggregations of criminal history information are more valuable and national-level aggregations even more valuable. The national criminal history system is moving toward a decentralized model which requires an efficient standardized way of requesting, sending, assembling and receiving such criminal history information.

Employment opportunity data is also geographically decentralized, but should be available on a statewide or even regional basis depending on the specific needs and mobility characteristics of an inquirer.

In many cases a single person is known simultaneously to several governmental agencies: a client of several parts of the criminal justice system while also a client of several health maintenance, income maintenance, food provision and training services of the government. It is not unreasonable to think that the public interest would be served if someone, under certain circumstances, would be able

to look at the "whole person" rather than viewing through the lens of a single service or agency.

Motivation: Shared Information is Less Expensive

For agencies, organizations and persons that receive information captured by others, the costs of data capture are avoided; the data capture costs, which are ongoing and mostly labor (and thus subject to wage inflation) are replaced by data transfer costs, which are largely one-time system development costs followed by miniscule transaction costs.

Certainly the cost savings are not evenly shared. "Sending" agencies, organizations and persons bear the data capture costs while "receiving" entities enjoy the cost savings. At the enterprise[3] level, however, the savings are substantial, and indeed nearly all agencies are both senders and receivers.

Providing a standardized way of sharing information will in the long run drive down the procurement cost of governmental information systems. Much of the custom programming required in such information systems is dedicated to the connectivity of a system with other systems. When the sharing requirements are identified and documented, clear and comprehensive specifications for information content and format are defined and tested, and technical methods for information sharing become subjected to open standards, these custom programming tasks will become a part of the general software development environment, produced once and implemented many times. For example, standard RFP language could be crafted to assure that a police information system contains all information sharing capabilities needed to connect to other agencies in a standardized way. No agency thinks it unusual that the ANSI/ASCII character set is to be used for communications, or that the agency will be charged more money for a system that uses this open standard; in the same way in the future no agency will think it strange or more expensive to specify the information-sharing specifications and standards that will then be common.

Motivation: Answering Objections

Change is always painful, and so there are always objections to any call for change. Here are a few.

Objection: We are already integrated at the county level with a consolidated system. **Response:** Consolidation provides data sharing within the county, but maybe not all possible integration with state and federal entities. Moreover, consolidation imposes the burden of maintaining the entire county system as a single system, whereas message-sharing integration methods allow the alteration or replacement of individual parts without affecting the other parts, thus permitting gradual change.

Objection: Our staff doesn't understand the document-sharing concept. **Response:** The staff is probably already supporting several document-sharing applications, like the NCIC-report and query formats, the state criminal history disposition reporting format, and the NIST fingerprint arrest reporting format. Adopting a *single* message-sharing technology for all connectivity requires less total expertise, not more.

Objection: We buy commercial software, and the vendors don't support the technology you are describing. **Response:** Vendors provide the system features that their customers require. Granted that some early adopters will end up financially supporting vendor development of the new capabilities, the final result will be a commodity-like set of software offerings which will be lower-priced for all.

Objection: We already have a large investment in legacy message-sharing techniques. **Response:** The various transmission formats imposed by the FBI, states and others do have substantial legacy implications. But if all these disparate formats used the same technology, the long-term costs of the overall national system would be lower. The key consideration is to allow for an extended period during which the state and federal level entities support both the legacy methods and the new standardized methods, so that the legacy can gradually fade out over a decade or so.

1 The Center for Technology in Government of the State University of New York at Albany has received a grant from USDOJ/BJA to develop the "business case" for such information sharing; its report will undoubtedly amplify and refine the reasons given in this section.

2 Among those in attendance at the NASIRE mid-year meeting who expressed an opinion, 92% believe that information sharing is important. Poll Analysis item 1.

3 The term enterprise is often used and seldom defined. For our purposes it may be as broad as "the body politic" or perhaps more narrow such as "all American governments" or "my state government" or "the criminal justice system".

WHO: Entities Sharing Governmental Information

Let us first introduce the term *entity*, which can be a governmental branch or agency, a private organization (e.g., a for-profit or not-for-profit corporation, or partnership) or individual person. If the entity is a governmental branch or agency it may be at the local, state or federal level.

Entities: Information Holders and Receivers

Information sharing involves the transfer of information from a *holder* entity to a *receiver* entity.

In the work reported in this report we have limited our consideration *solely* to governmental[4] entities as *holders*, and *almost exclusively* to governmental entities as *receivers*. This tight focus can be expanded to the consideration of non-governmental holders and non-governmental receivers as the work continues.

The holder-to-receiver information flows considered during our work are:

Information Holding Entity	Information Receiving Entity	Example	Volume & Complexity[5]
Local Govt.[6]	Local Govt.	Police to Prosecutor	High/High
Local Govt.	State Govt.	Police to Criminal History	High/Medium
Local Govt.	Federal Govt.	Police to NCIC	High/Low
Local Govt.	Private Org.	Police to Parade Sponsor	Low/Low
Local Govt.	Private Person	Court to Defense Attorney	Medium/Medium
State Govt.	Local Govt.	Criminal History to Prosecutor	Medium/Low
State Govt.	Federal Govt.	Prison to Bureau of Prisons	Low/Low
State Govt.	Private Org.	Criminal History to Employer	Medium/Low
State Govt.	Private Person	Parole Board to Crime Victim	Low/Low
Federal Govt.	State Govt.	File Synchronization Data	Low/Low
Federal Govt.	Local Govt.	Response to Warrant Check	Medium/Low
All[7]	All	Investigation	Medium/Low

In the work reported in this report we have limited our consideration primarily to criminal justice governmental agencies. This was done to focus our limited resources. Nearly all governmental entities and many private (organizational and individual) entities have roles as information holders and receivers. Further work will expand our horizons.

Entities: General Obligations of Holders and Receivers

There are general rules which govern information sharing:

- The *holding* entity must be *willing* to share the information; this implies willingness to share some information and unwillingness to share other information;
- The holding entity must be an *authoritative* source for the information, that is it must be the entity that first collected the information or one that can trace the information to the entity which first collected the information via paths that did not involve re-keying of the information;
- The *holding* entity must have the *right* to share the information; in some cases this may mean that the information transfer is specifically required by law or regulation, while in other cases it may mean that the transfer is not proscribed; if the holder is not the original collector of the information there may be limitations on further distribution placed by the original holder;
- The *holding* entity must be *able* to share the information, that is, to have the proper technological capabilities, or to have designated another entity with such capabilities to act in its stead;
- For some types of information, the *receiving* entity must have a demonstrable *need* for the information and the *right* to obtain the information from the holder;
- The *receiving* entity must comply with any *rules* imposed by the holder concerning use, storage or further dissemination of the information, and must pay any fees associated with the information transfer;
- The *two* entities must have a mutually agreed technological *method* for the information transfer.

4 Of those in attendance at the NASIRE mid-year meeting who expressed an opinion, 78% believe that local-local, local-state, local-federal and state-federal sharing should be included. Poll Analysis item 2.

5 High complexity implies a broad variety of information to be exchanged in many different message content models; it also implies several different exchange modalities (push, pull, publish, subscribe, query, respond). Low complexity implies only a few information content models and perhaps one or two modalities. Medium complexity is in between.

6 Local-Local means in the same locality; Local to State means in the same state.

7 All to All means any governmental entity anywhere in the nation, at any level, to any other governmental entity in the nation.

Scope: Dynamic Structured Information

We have excluded from our consideration nearly all *static information* such as mission statements and organization charts, legislation and regulation and case law publication. This is not to say that such information should not be shared, only that existing methods of distribution are probably satisfactory.

We have also excluded that portion of *free-text information* that is not converted into a structured format. Such documents are meant solely for human reading, rely on context for the meaning of content, and utilize a large natural-language vocabulary. Again the exclusion from scope does not imply that sharing of such information is unnecessary, only that it does not fit the pattern of our concerns in this effort.

Our focus is on *dynamic information* presented in the form of *structured information*[8].

Dynamic information changes with time. A person is on probation and then free of probation, under arrest or not, on trial or convicted, and the correctness of the 'answer' depends on when the 'question' is asked and answered.

Structured information (e.g., <DateBorn>19660608</DateBorn>) carries both the information content (e.g., 19660608) and the semantic content[9] which make clear that the content is a date in the ISO format so that it can be understood as the year 1966, the sixth month (June) the eighth day, and it is the date of birth of the record subject. The semantic data may be within the sharing message itself or may be pointed to; for example if the sharing message is known to be an Internal Revenue Service Form 1040, it is necessary only to show explicitly the line number (e.g., 1040-32) in order to know that the numbers which follow are the adjusted gross income in dollars.

Much of the information to be exchanged is textual, but some includes document images, fingerprint images, photographs, drawings and charts.

Scope: Transactions and Data Collections

Batch Reporting. A considerable proportion of present inter-entity information sharing is done by *batch reporting* of transaction data. A court case management system flags each case-closing function throughout the day (or week or month) and at the end of the time period the system collects all these data, transforms them into the required format for sharing, dials a telephone number and uploads the data to the specified information system. The shared data is less timely than it could be. Special programming must be conducted to determine the number of individual records in the batch, to assure that the receiving governmental branch or agency sends a response message, to assure that the response message specifically acknowledges receipt of each piece in a usable format, to retrieve and correct as necessary all pieces not acknowledged, to form another batch of these pieces, and to restart the process for the new smaller batch.

Why is the batch mode chosen? Not for programming ease or efficiency. It is chosen because telecommunications linkage is (or was when the system was designed) rare and expensive. Batch mode is not the choice for an information-rich internet-like telecommunication environment.

We have excluded batch reporting from our scope statement

Transaction Reporting. The preferred alternative for the sharing of operational data is *transaction reporting*. That is, each court disposition in the foregoing example is treated as a separate transaction and is shared as soon as it becomes available. Ideally the court disposition is captured in the courtroom when it occurs. Simple and standardized methods for inter-connection are used and if an error it encountered it is immediately known and can be corrected.

Examples of transactions drawn from the criminal justice processes include: incident reports, arrest reports, booking reports, prosecutor declinations, court filings, court dispositions, corrections intakes, corrections releases, status changes.

Transaction reporting is included in our scope.

Data Collections. Data Collection reporting is a special form of information sharing. Examples are the annual reports from all state prisons to the U.S. Bureau of Prisons, and monthly and annual NIBRS and UCR reports. They usually cover a fixed time period.

Many data collections are summaries and are presented in tabular form showing totals and subtotals in various categories. Other data collections, like the NIBRS report, provide a set of data elements for each occurrence being reported (e.g., incident or arrest or court case).

Data Collection reporting is included in our scope.

Our scope statement at this point includes all governmental information which is structured and dynamic, if it can be exchanged in transaction-reporting or Data Collection-reporting form.

Scope: A Criminal Justice Example of Governmental Information Sharing

An example of appropriate information sharing drawn from criminal justice processes is provided in this section.

1. A suspicious police officer submits a *query*[10] to the statewide warrant system and discovers from the *response* that the subject of his car stop is wanted.
2. A police officer arrests a subject, then completes and (digitally) signs an arrest document which describes the crime, arrest and arrestee, stores the document in the police information system, and finally affixes to the arrestee a bracelet containing the arrest document number.
3. The arrestee is taken to the sheriff's office to be booked. The sheriff's *information system*[11] uses the arrest document number, scanned from the subject's bracelet, to *pull* the arrest document from the police *information system*, and uses

data from that report to (partially) fill in the booking document.

4. The sheriff's *information system*, using personal-description data in the arrest report, *pulls* the local, state and national criminal history records, and, based on the information in these documents, a person who makes a security decision enters that decision into the sheriff's *information system* which assigns an appropriate cell.
5. The sheriff's *information system* uses information from the arrest and booking documents to fill in a standard press release and pushes it to the appropriate web page.
6. The sheriff's *information system* uses information from the arrest and booking documents along with the mug shot and electronic fingerprint capture submission, using live-scan or card-scan solutions, to push to the state criminal history repository the arrest report required. The state repository's *information system* will, after its own processing, push the information to the national criminal history repository maintained by the FBI.
7. The sheriff's *information system* pushes to the prosecutor's *information system* the arrest document, booking document and three criminal history records, and the prosecutor's *information system* uses some of this information to (partially) fill in a prosecution case intake document. An assistant prosecutor views all the available information and makes the decision to prosecute.
8. The assistant prosecutor decides the specific charges to be filed, and the prosecutor's *information system* prepares the charging document using statute-specific standard charging language plus information from the earlier police and sheriff's documents. The assistant prosecutor (digitally) signs the charges and the prosecutor's *information system pushes* the charges to the court's *information system*.
9. The court *information system* creates the initial records for a new case, assigns it, reviews the judge's calendar and sets a time for a preliminary hearing. It then *pushes* to the prosecutor's *information system* and the public defender's *information system* information concerning the assigned judge, courtroom, date and time. The prosecution's *information system* reviews its workload and fills in its calendar by assigning the hearing to one of its assistant prosecutors, and the public defender's *information system* similarly assigns the hearing to one of its defenders.
10. Throughout the course of the trial the prosecutor's and defender's *information systems push* digitally signed motions to the court's *information system*, the court's *information system pushes* copies of motions and notifications of hearings to the prosecutor's and defender's *information systems*, and the human parties, namely the prosecutor, defender, defendant and judge, meet and make decisions. At appropriate times, the court's *information systems pushes* notifications to witnesses who are required to attend a given hearing or trial portion, and the prosecutor's *information system pushes* notifications to victims at important milestones of the case.
11. At one point the defendant, who is free on her own recognizance, fails to appear for a court date and the judge issues and digitally signs a warrant, which the *information system pushes* to the original arresting agency for service.
12. As the trial ends the judge decides the case and passes sentence, which the court's *information system* transforms into a (digitally) signed document which it then *pushes* to the sheriff's *information system* to send the newly convicted prisoner to the state prison, *pushes* another such document including the sentence ordering the prison's *information system* to receive the prisoner, *pushes* yet another document to the defender's *information system* setting the date by which a notice of appeal must be filed, and so on.
13. Later the prison's *information system* and the parole *information system* coordinate the transfer to parole.
14. Unfortunately, while on parole the subject commits a crime and is arrested. When the booking *information system* retrieves the criminal history record via electronic fingerprint transactions, it 'notices' that the subject is on parole and that the parole agency has *subscribed* for news of any arrests during the time of parole. It pushes to the parole *information system* an arrest notification message. The parole *information system* then prepares a parole violation document and pushes it to the appropriate parole officer for consideration.

The example, up to this point, has shown only interactions among the governmental entities which are usually thought of as comprising the criminal justice system. But there are many other entities which are pulled into the criminal justice *process*. For example:

- Private defense attorneys;
- Jurors and potential jurors;
- Public and private schools (truancy, vandalism)
- Public and private half-way houses;
- Public and private diversion alternatives to incarceration;
- Fire and other Public Safety governmental components for combined operations;
- Alliances of nearby police agencies for combined operations;
- State and federal agencies for periodic reporting on local crimes, arrests, prosecution cases and outcomes, court cases and outcomes, prisoner populations, correctional supervision cases and outcomes, and many more.

Notice that in the examples an information system (really the business process which is incorporated into the information system) sometimes "knows" that information is required and *pulls* it without being asked by a person. Other times an information system "knows" that another system will need certain information and *pushes* it, without being directed to do so by a person. Still other times an entity "knows" that it would want to be informed if certain events happen to a certain person, and *subscribes* to such information, confident that if the event happens the entity which holds the subscription will *publish* a notification to the subscriber. Finally, some-

times a person launches a query and receives an appropriate *response*.

Some documents are digitally signed and thus become the official and original document (there is no underlying paper original, and no need for one). For example, the file which has traditionally been called the warrant file is in fact a file of data *extracted* from real warrants; it is the real warrant which must be retrieved and served. Maintaining a file of real digital warrants will permit substantial reengineering of business processes within police and court entities.

Does the level of information sharing described in this section require the development of an entirely new technology? It does not.

Does it require extraordinary computer power? It does not.

Would it make for increased governmental efficiency and effectiveness and a more just society? It would.

Have bits and pieces of it been implemented in various places? Yes.

Has it been fully implemented anywhere? It has not.

Scope: Other Examples of Governmental Information Sharing

The present effort has not expanded beyond the bounds of the criminal justice system. Lest some reader draw the incorrect inference that sharing is not required in other parts of government, we offer a few possibilities for non-justice use.

1. When a city streets department or state highway department *information system* 'learns' that a route will be closed, it could inform the *information systems* of public safety agencies, transit agencies, and the press.
2. Individual school *information systems* could report daily attendance figures to school district *information systems*, which could aggregate the data and send the appropriate "invoice" to the state education department *information system*.
3. Responses to criminal record checks for some job applicants could be provided directly from the criminal history *information system* to the prospective employer via Internet.
4. Approvals for certain sensitive jobs could be kept on file and the employer automatically notified of any subsequent conviction.
5. Laws against 'double-dipping' into certain public benefit programs could be enforced by checking the roles of separate aid agencies.

Requirements for Information Sharing

The basic requirement is to exchange:

(a) structured information including transaction reports and data collections containing

(b) textual, photographic, fingerprint, graphic and page-image data

(c) nationwide

(d) from governmental entities in all functional areas

(e) at all three levels of government

(f) to entities some of which are governmental, others are private organizations and still others are individual persons.

(g) Some of the receiving entities will be regular and high-volume users,

(h) others will be one-time or casual users.

(i) At the point of final use, the information may be displayed, printed or introduced directly into the receiving computer's database or application.

Within an entity, it is desirable to perform all information sharing using a single technique and employing a single set of software tools; it should not be necessary to support one expertise for local sharing, another for intrastate sharing and another for national sharing.

There should not be a requirement for a single national database structure or database data dictionary.

Requirement (a) concerning *structured data* limits the number of suitable methods; we can not use free-text[12] streams of natural language. We cannot use the "flat file" techniques[13] used to send data from spread sheets because the technique cannot express the more complex[14] data structures needed. SQL can express complex data structures, as can various "tagged-field" formats; SQL, however, requires that the requester know the database structure of the holder. The requirement to support summary reporting does not add any additional complexity to the requirement.

Requirement (b) concerning textual and non-textual information, implies that the method must provide a rich variety of data types.

Requirement (c) concerning *nationwide* usage makes it practically impossible to use anything like SQL which relies on the user to know beforehand the storage[15] data structure of the holder's database. The tagged-field contenders remain.

Requirement (d) concerning the *multifunctional* dimension of the governmental entities throws further cold water on the SQL alternative, but reinforces the notion that field tags (i.e., data element format and content descriptions) must be agreed on a very broad basis.

Requirement (e) concerning the three *levels of government* makes it unlikely that storage database uniformity can be achieved even within each individual functional area. The tagged-field concept is reinforced.

Requirement (f) bringing *private organizations and individuals* into the picture suggests that use of standardized software-based methods of sharing should be emphasized since these entities will not be keen to invest in special purpose hardware and software. It also suggests that the Internet will be a necessary part of the information sharing infrastructure.

Requirement (g) concerning high-volume regular users suggests that simplified procedures should be available for such uses.

Requirement (h) concerning *low-volume and one-time users* suggests

using absolutely open standards, relying on software readily available in the Internet domain, and limiting special-purpose software usage to small programs which can be downloaded as part of a one-time information request/response transaction.

Requirement (i) concerning the need to (perhaps) transform the received data into new definitions appropriate to the user, and to *display, print or insert*[16] the transferred data argues for the use of software techniques that can perform the necessary semantic transformations and re-formatting required.

Implications of the Requirements

Implementations designed to meet these requirements will have these characteristics:

- A focus on the sharing of structured data suitable for display, printing or insertion into application programs and databases; therefore a tagged-data syntax supported by style sheets and document-to-database software tools;
- A focus on the sharing of pre-defined combinations of data elements (documents) rather than free-form collections of data;
- A vocabulary which, to the extent feasible, can be made compatible with a broad variety of information systems in all governmental applications and also in general commerce;
- A telecommunication environment similar to the Internet.

8 Of those who attended the mid-year NASIRE meeting and expressed an opinion, 84% believe that such a model is appropriate for information sharing.

9 Semantic data expresses meaning to be ascribed to the "raw" data. The example is in XML format; the DateBorn attribute carries with it a precise meaning, edits and format for the data expressed between the start-tag and the end-tag.

10 The terms query, response, push, pull, publish and subscribe, as defined in Appendix A, are information sharing modalities; each is appropriate to a different operational setting for information sharing.

11 The italicized phrase information system is meant to denote the computerized information system operating under program control without human intervention.

12 Free-text is just a trivial form of structured data; that is, one can specify that the structure of the document one is transmitting is a structure-less structure.

13 The so-called comma-delimited files exported by many spreadsheet programs fit this category.

14 For example a criminal history record nests several sentences per charge, several charges per arrest, several arrests per subject.

15 For example, each state maintains a criminal history system collecting approximately the same data, but the database structure of such a system is almost always unique to that state; furthermore the structure may change over the life of the state system.

16 It is the data insertion requirement that forces the most stringent need for edited structured data. Many of these sharing situations are either lateral local interchanges (within a county) or are vertical interchanges (local to state or state to federal). In other sharing situations where human interpretation of the information can be assured, less formal structure may be permitted.

The Need for Open[17] Internet Standards

There are good reasons to insist on open Internet standards:

- *Open* standards maximize competitive behavior and thus minimize procurement costs;
- *Open* standards exhibit a straightforward evolution path so that the standard can survive several generations of hardware and software change.
- *Internet* standards come with a large variety of software tools which minimize development costs;
- *Open Internet* standards are understood by most IT professionals, which assures a large supply of competent workers.

Proprietary standards and so-called Industry standards are not sufficient. They do not provide the platform-independent, vendor-independent procurement suitable for a major governmental initiative. Furthermore such standards do not carry with them the assurance of thorough peer review and evolutionary growth path that open standards do.

Non-Internet standards are not positioned for the future. Andrew Grove, founder of Intel, has said that within five years all companies will be Internet companies or they will not be companies at all. Allowing for a modicum of hyperbole, this attitude nonetheless summarizes the consensus of seasoned technologists. Government entities, even non-Internet government entities, will continue to exist but will become increasingly ineffective and inefficient in the Internet-saturated world.

The list of open Internet standards appropriate for the tasks of national information sharing include at minimum:

- TCP and IP telecommunications standards;
- HTTP transmission protocol;
- XML tagged-field data structure meta-language;
- XSL and CSS style sheets for information transformations and formatting;
- LDAP to access directories for passwords and permissions;
- ISO 8859-1 for character set (a-z, 0-9, etc);
- MIME for email attachment wrapping;
- IPsec for transmission security;

Telecommunication Infrastructure Needed

A state can implement intrastate information sharing without an internet-like network, and some have done so. The advantages and economies of adopting the Internet-like model, however, are such that we expect most, perhaps all, states to have such technology in the mid-term future.

Since the long-term objective is to provide full connectivity among all governmental agencies, the intrastate requirements should be met with a single network, or with multiple networks which interconnect such that to their users they seem to be a single *intrastate network*.

Since these networks will carry sensitive information, provision for *encryption and authentication* are required. The encryption and authentication techniques used should be the same throughout a state. Information holders will determine which information requires encryption and/or authentication.

Some information requires a *digital signature* in order to meet existing legal requirements, and the choice of a digital signature method is best made at the state level. Because of the broad data flows among local, state and federal agencies, considerable care must be taken to assure that information exchange is not hindered by incompatible encryption, authentication or signature technologies.

There are at least three pieces of directory information that are essential for information sharing: a unique address for both the holder and receiver entities, the fact that an entity is in the sharing program, and what information sharing messages (documents) the two entities are willing to use. At this point it is not yet clear whether there will be many directories or one, whether they must all be identical or only support a standardized access method.

Telecommunication Infrastructure in Place

Every state has a network for information exchange among state and local *law enforcement* agencies within the state; not all law enforcement agencies, especially not all small ones, are connected to these networks. The state law enforcement networks also provide connectivity to two national networks, NLETS and NCIC, for information exchange with state and local law enforcement agencies in other states and also with federal law enforcement agencies. Special-purpose connectivity is also provided by the CJIS-WAN, a network managed by the FBI and primarily devoted to fingerprint-based traffic at this time. The NLETS, NCIC and CJIS-WAN networks are (or are in the process of becoming) TCP/IP-based.

The intrastate law enforcement networks provide connectivity among police, courts, prosecutors, supervision and correction agencies. These agencies also share in the services of NCIC through those intrastate networks. NLETS provides interstate connectivity among these same agencies. Courts, prosecutors, public defenders and correctional agencies sometimes have their own dedicated intrastate networks or are provided connectivity via a statewide governmental network, and these dedicated or general networks usually supply some gateway to/from the law enforcement networks.

The smallest agencies are often not connected to any network, sometimes being served by a nearby larger agency that will receive their messages and relay them by voice telephone.

Some intrastate networks are Internet-like (using the open Internet standards listed above) and more are becoming Internet-like.

Twenty-one states with 57% of the national population have some Internet access to own-agency databases now, and 30 states with 61% of the population believe that such sharing will increase in five years[18]. About the same number of states and proportion of the national population allow other-agency access now and expect

that use to grow[19]. Twenty states with 52% of the national population allow private-citizen access now and 33 states with 64% of the population believe such use will grow in five years[20].

Twenty states with 48% of the national population use some Internet encryption now and 40 states with 81% of the population believe that such usage will grow in five years[21]. Nine states with 36% of the national population use digital signatures now and 31 states with 65% of the population expect that usage to grow in five years[22].

Fourteen states with 37% of the national population collect information-related fees by credit card now and 27 states with 53% of the population believe that usage will grow in five years.

Telecommunication Conclusions

The survey results make it clear that the states understand the direction in which telecommunication technology is headed, and agree that state and local telecommunication infrastructures need to be aligned with the new technology. The problem is not lack of understanding; problems of funding, of scheduling, of implementing, of training users, of maintenance exist and are being worked on.

Although the direction of telecommunication technology seems quite clear at this time, and, although many technical advancements are backward-compatible, someone or some body must be responsible for maintaining nationwide governmental information sharing once it has been achieved. Within each individual state the natural candidate for this assignment is the state's Chief Information Officer. Localities can be expected to follow this lead if it is logical and firm. There is no such natural candidate for the federal level of information sharing.

Persistent problem areas include: (a) the "last mile" of connectivity, especially to small and/or remote agencies, at a reasonable cost; Internet services and some form of universal access policy are approaches to this problem; and (b) separate and overlapping networks for various user communities, which raise costs and inhibit information sharing.

Other problems areas include: QOS (quality of service) commitments, VPNs (virtual private networks), strong authentication of users (user identity confirmed by methods more foolproof than passwords), encryption and digital signatures are approaches to this problem; these technologies also apply to networks using radio frequencies (e.g., messages to vehicles).

17 Of those in attendance at the mid-year NASIRE meeting who expressed an opinion, 96% believe that adherence to open standards is of high or medium importance.

18 CTO Survey Q 94, 95.

19 CTO Survey Q 96, 97.

20 CTO Survey Q 98, 99.

21 CTO Survey Q 100, 101.

22 CTO Survey Q 102, 103.

The Need for Standard Document Definitions

There are thousands of governmental information systems. We intend for them all to be able to exchange structured information with *all* the others, subject only to their needs and willingness to share. At the same time we recognize that their needs are closely related to their geographic proximity (nearby entities exchange information more often than distant ones), and their operational proximity (entities linked by common processes exchange more often than independent ones). Also, entities which share information with the intention of populating databases require more rigorous structure than entities which share information with the intention of displaying the information to human viewers.

At the *city/county scope* it is relatively straightforward to allow information sharing. Serve all agencies through a single application program running on a single computer (the consolidation solution), or through several application programs, perhaps running on several computers, but sharing the same data dictionary (the common native language solution), or through middleware customized to the different data dictionaries involved, which converts each of them to a common language and storage in that new language (the data warehouse or datamart solution), or through middleware that operates 'on the fly' translating each message from every possible sender language to every possible receiver language (The Rosetta solution).

Each of these alternatives permits information sharing, but there is a price to be paid for each: lock-in to a certain functionality set (consolidation), or dedication to a single provider for all systems (common native language) or to major upheaval every time the database structure of any participant changes (data warehouse or Rosetta). The document exchange solution we propose in this report does not suffer from these drawbacks.

When rising to the *statewide scope* of information sharing, the consolidation solution becomes infeasible because of the sheer number of applications to be coordinated; and the Rosetta solution becomes infeasible because as the number of different dialects rises the number of any-to-any paths rise as the square of the number of dialects. The common native language solution can be made to work but a very firm hand is required to enforce uniformity and the problem of non-conforming legacy systems may last for decades; large-city systems will be especially difficult to bring into the fold. A data warehouse solution can be made to work but substantial maintenance is required with each change of the database structure of any participant. The document exchange solution we propose in this report does not suffer from these drawbacks.

Composite solutions can be made to work at the statewide level. One possibility combines a common native language solution for all the smaller participants in a data warehouse solution linking that solution with a data warehouse or Rosetta solution for the few large participants.

We believe that the only concept that scales to the *nationwide scope* is one, call it the document[23] exchange solution, in which each system speaks its native language to its own 'customers' but speaks another[24] language (call it Data Esperanto) to 'outsiders,' and can translate Data Esperanto to its own language as well.

The vocabulary and grammar of Data Esperanto need to be chosen with nationwide document exchange in mind, and expressed in accordance with the syntax rules of XML[25].

Document Definition Infrastructure Needed

The simplest and most direct method of providing national sharing of governmental information has these characteristics:

- Each entity wishing to be a holder identifies those documents that it is willing and able to *deliver*, selecting from the national library of document descriptions;
- For each such document description it prepares a set of transformations which 'map' *from* its internal database storage element formats *to* the format specified for the sharing document, and develops the ability to assemble the elements into the specified document format and to *transmit* that document;
- Each entity wishing to be a receiver identifies those messages/documents that it desires to be able to receive, selecting from the same national library of document descriptions;
- For each such document description it prepares a set of transformations[26] which 'map' *to* its internal database storage element formats *from* the format specified for the sharing document, and develops the ability to *receive* that document;
- The holder entity keeps the national directory updated concerning which document descriptions it supports.

The national sharing system becomes operational as soon as there are any two entities that register in the directory for the same sharing documents. Thereafter the nationwide rollout of information sharing is completely within the control of the willing participants.

Some states may wish to control the pace of the rollout more directly.

First, suppose a state has imposed a single data dictionary on every governmental information system at its local and state levels. Suppose further that the state provides a network or networks such that all governmental entities are connected for information sharing, but that the networks are not Internet-like. Suppose further that the state provides a single facility within the state at which translation between state-specific standard documents and nationwide-standard documents could occur. Under such circumstances would there be any need for each entity within the state to comply with a national model of information sharing? The answer is no. The "outside world" would address the single translation facility and would perhaps not even know that an intermediary had been inserted between itself and the sharing entity.

Alternatively, suppose there are two or three or ten data dictionaries in use within the state, but that the state translates among these

plus the national transfer dictionary. Now the complexity of the state translation facility is substantially increased, and it must handle not only dealings with the "outside world" but with portions of the "inside world" which do not share a single dictionary. Still, there is no requirement for each in-state entity to handle the national model.

Aside from these and similar cases, however, the requirement is that each *holding* entity which wishes to be a participant in national information sharing must be able to translate from its storage database dictionary to the national model, and each *receiving* entity must be able to translate from the national model to its storage database dictionary[27].

Document Definition Infrastructure in Place

Several states have implemented statewide court reporting to the state criminal history system using an adaptation of a fingerprint-based transmission standard[28], and at least one (Wisconsin) has used XML as the sharing syntax. The document definition in each such case has been state-specific.

For all practical purposes there is no in-place infrastructure that is precisely usable in the context of the nationwide information sharing we are focussed on. However there is a great deal of "raw material" from which such an infrastructure could be constructed rather quickly.

Many local/state-to-Federal *law enforcement documents* are already defined in tagged-field formats which could readily be converted to the specific technology foreseen for the national information sharing environment. For example, data is entered, modified and cleared from all the FBI/NCIC "hot file"[29] databases in an on-line, real-time mode by localities and states using proprietary FBI dot-delimited transactions. Electronic fingerprint and arrest data is submitted to the FBI IAFIS and III using a transaction controlled by a national standard[30] that uses tagged fields defined within the standard and accompanying specifications. National data sets for statistical interpretation of crime incident/outcome data (National Incident Based Reporting System or NIBRS) already has an MRD (machine readable data) format which is readily convertible to a tagged-field format.

Local/state-to-federal *corrections documents* defined in forms are easily mapped to standards. These documents are tabular in format and the table content is rigorously defined. Tables are easily converted to the information-sharing technology proposed in this report.

Local-to-state *criminal justice documents* are now defined in state-specific forms which probably can be generalized to achieve national applicability. These tend to be intra-functional vertical information sharing (e.g., court to court administrator, police to state-level hot files) and already have sufficient structure to allow translation to the proposed technology.

Intra-locality criminal justice information sharing requirements are not yet well defined. The exchange point study now underway will shed considerable light on these requirements and the data structures which will support them. Substantial effort will be required to produce the needed document descriptions, but the potential payoff is very large in terms of allowing interconnection of dissimilar case management systems and the reengineering of local criminal justice processes and workflow.

Information sharing involving either a holder or requester outside the criminal justice community has not been considered in this first stage of our work.

Document Definition Conclusions

The survey results show that the states understand the value of standardized documents as the unit of information exchange; in fact, most local-to-state and local/state-to-federal information sharing is already done in this way. The persistent problem is the proliferation of "non-standard standards" being used, with the result that most information systems must support several different information transfer data dictionaries and several information transfer document encapsulation methods.

The recent standardization of XML provides a robust extensible encapsulation method superior to any now in use; what is still lacking is widespread knowledge about and adoption of the new standard, both for "new-start" sharing applications and for re-work of existing applications.

The development of document specifications for sharing requires substantial work at the national level. Adopting data element specifications which reflect government-wide focus while making re-work of existing sharing documents will require expert and thoughtful work in the first stages, and then large amounts of painstaking work to complete the library of sharing documents.

Within the criminal justice and law enforcement information user communities there is long experience in sharing information. Therefore, governance of an ongoing nationwide information sharing program has solid bases upon which to build unity between the two groups. As the scope expands beyond these communities to the users of all governmental information, there is no such treasury of experience and no natural candidates for leadership of the ongoing program. It is likely that new governance channels will be required.

23 In a document exchange concept there is a common data dictionary (e.g., date of birth always has the tag "DateBorn" and the same CCYYMMDD numeric format) but the unit of exchange is not the single 'word' but an entire document (such as the arrest report document or the judicial warrant document), specified in detail.

24 If Data Esperanto is very good, then over a decade or two many system builders are likely to adopt it as the native language. This is very different from top-down coercion, of course.

25 Document Type Description (DTD) contained within the XML standard permits very precise definition of documents and their constituent terms. An even richer method involving data schemas is in the final stages of incorporation into the standard as an alternative.

26 Nearly all recently implemented governmental information systems use relational database management systems (RDMS) file structures. The latest releases of all RDMS packages include software tools to transform RDMS file structure into XML document structure, and vice versa.

27 This is not precisely true. The requesting entity may want only to view the shared information on a screen or to print it. In these cases there is no need for database-level translation. Indeed many entities may begin participation in this way, and only progress to database-level sharing later.

28 ANSI/NIST-CSL-1-1993.

29 Wants, warrants, missing persons, stolen property, etc.

30 ANSI/NIST-CSL-1-1993 Data Format for the Interchange of Fingerprint Information.

Regular Local Information Sharing

Some governmental entities, by the very nature of their work, are (or ought to be) in constant communication with affiliated entities. Police, prosecutors, jailers, courts and correctional agencies within a single county are in this category.

Because there only a few entities involved, they can agree among themselves on the telecommunication standards to be supported, the list of document specifications they will support, the security precautions to be undertaken, the modality[31] of information sharing for each of the documents, and the specific telecommunication addresses to use for sharing. Since they know the addresses and passwords needed for communication they do not need directory services.

All the participants can send and receive, display and print, screen-fill and database-fill from each other. A user neither knows nor cares where the data originated or is stored so long as the information received is complete and up to date.

Each entity sets its own priorities and runs its own system. When an agency decides to replace its present computer system it simply includes in its requirements statement the technical standards to be complied with and the list of sharing document specifications which must be supported. No adjustment to or notification of the new system is required by any of the entities with which it shares information. In fact they may never learn that a new system has been installed.

It might even be said that such entity clusters might not need to adhere to national standards or notional specifications at all, simply make up their own. Indeed some localities have done just that. The price for such localism is paid in increased development and maintenance costs, and greater difficulty in extending the information sharing activity beyond the original "circle of friends."

Regular Vertical Information Sharing

Operational data flows vertically as well as horizontally. Police, prosecutors, courts and correctional agencies report to the state criminal history system. Police report to the federal 'hot files' (e.g., warrants, stolen property). Courts report to the administrative office of courts. Local agencies retrieve information from state and federal files.

Here too there is no need for directory services since the usage is regular and relatively unchanging over time. Even when the sharing relationship is less operational and relatively infrequent, this simple interchange model is appropriate; quarterly and annual reports to state and federal supervisory bodies, and database synchronization at various levels are examples.

Each state could define its own data structure and definitions for local-to-state information sharing, and in fact in the absence of a national initiative several states have already done so and perhaps more will. The penalty for this state independence is paid in duplicatory and inconsistent specifications between the local, state and federal entities; this will only increase as the requirements for information sharing expand beyond the criminal justice community to the much wider governmental community.

Occasional Sharing

Some information sharing is triggered by a specific situation that may never recur. For example a detective is reading the national criminal history record of a person, and needs to examine the entire correctional history of that person from another state. A prosecutor nears the sentencing stage of a trial and needs to examine every pre-sentence report produced anywhere in the state concerning the defendant.

In some of these cases it will be unclear exactly what telecommunication path is be followed; in others the address of the entity holding the information may not be known or the specific document descriptions supported by the holding governmental branch or agency may not be known. In some cases, the inquirer may not even know what entity to ask concerning a specific person. A common statewide telecommunication infrastructure will remove many problems, and a common nationwide telecommunication infrastructure will remove virtually all impediments to hook-up.

For such occasional use, the requirements for a common information structure will be considerably less[32] than for the more regular interchanges which end with database population. Since human interpretation of the information is assumed, simple structure suitable for formatting a screen or printed page will be enough in many ways.

For national information sharing at anything more than the administrative message level of complexity, directory services will be required to permit the information requester to know the specific telecommunication address and path, sharing modalities and specific documents supported, passwords, authentication, signatures and encryption required to make a proper connection.

31 Push, pull, query/respond, publish/subscribe.

32 NLETS accommodates many such information exchanges. Some are quite fully structured, such as query/response of out-of-state vehicle or driver registration files, highway condition files and weather reports. Others are almost entirely free-text, such as the NLETS administrative message format, in which the message type (AM) signals that the remainder of the message is to be interpreted as free text.

NASIRE is not the only player in the effort to build a nationwide governmental information sharing framework. The issue is "ripe" and that means many initiatives have already been undertaken, others are being discussed, and many organizations are trying to shape the future at the highest levels.

NASIRE has a seat at the future-shaping table, has its own unique perspective on the role of state-level decisions in shaping that future, must and will be heard as the conversation moves from generality to specificity, from concepts to hardware and software, from criminal justice to all-government, from subsidized pathfinder projects to operational deployment which will pervade all governmental institutions.

As this document is being written there is no firmly declared and agreed national strategy. The following discussion is a contribution to the effort to construct such a strategy but does not claim to be such a strategy, certainly not in scope nor in level of detail nor in broad acceptance. It is offered as an opening voice in a national conversation concerning the governmental information environment of the not-so-distant future.

Strategic Directions

Direction: Rely on Cooperation, Not Coercion

Twenty-five states with 46% of the population report that they have the legislative authority to enforce information sharing in state-level systems, with an expectation that this will rise to 36 states and 66% of the population in five years[33]. But only five states with seven percent of the population have such authority concerning local systems, expected to rise to 13 states and nineteen percent of the population in five years[34]. *Coercion from the state level does not have a sufficient legislative backing to succeed.*

If coercion cannot achieve information sharing, can it at least be used to develop statewide data dictionary content so that information sharing would be easier? With respect to *state-level* systems, 21 states with 40% of the population have such authority now, expected to grow in five years to 30 states and 59% of the population[35]. But at the local level the corresponding numbers are three states and six percent of the population now, growing to ten states and fifteen percent of the population in five years[36]. *Data dictionary standardization for localities does not rest on a solid legislative footing.*

Federal grant administration agencies can place special conditions on grants requiring the grantee to consider information sharing, but cannot in general place an absolute requirement for information sharing on grant funds. The hortatory activities at the federal level could be reinforced at the state level; forty states with 90% of the population report that federal money flows to local agencies through state agencies which prioritize and allocate the funding, and 41 states with 74% of the population believe that these numbers will rise in the next five years[37]. *Prioritization by state entities of federal grants for local and state systems has a strategic role.*

The U.S. congress can appropriate money specifically for information sharing, and this is indeed powerful in motivating state and local entities to share information, but in general they cannot be required to take the condition-tagged money. *Special-purpose federal grants may have a role in the strategy.*

Precisely because coercion is neither feasible, desirable or effective, a strong and bullet-proof business case must be made, and the effort must be prioritized to provide early "victories" for the forces of information sharing.

Direction: Prioritize the Sharing Efforts

Information sharing among state agencies is perceived to have a high priority; 45 states with 95% of the population voice this belief and the same numbers believe the priority will grow in five years[38].

Information sharing among local agencies in the same county is perceived to have high-priority; forty states with 85% of the population voice this belief and approximately the same numbers believe that the priority will increase in the next five years[39]. The numbers for *local* agencies in *different counties* are only slightly smaller[40].

Information sharing between *local* agencies and *federal* agencies is also perceived to be a high priority; thirty-five states with 74% of the population voice this belief, with 37 states and 84% of the population believing the priority will rise in the next five years[41]. Sharing between local agencies and their own state agencies shows even higher numbers; 41 states with 91% of the population seeing a priority now, with 43 states and 92% believing the priority will be even higher in five years[42].

The intra-locality (whether actually implemented at the local, regional, or state level), locality-to-state and state-to-federal sharing paths should receive implementation and funding priority.

Direction: Use Internet-like Technology

Forty-six states with 96% of the national population have at least some TCP/IP networks, and all these states believe that that the proportion of total traffic carried on TCP/IP networks will increase in the next five years[43]. Thirty-four states with 81% of the population already serve local agencies with such networks and 44 states with 94% of the population believe those local-serving networks will carry more of the total traffic load in five years[44]. *Imposing a TCP/IP requirement for sharing will not pose a hardship.*

Forty-six states with 96% of the population report that persons and corporations can submit some documents to governmental agencies electronically, and they all believe that the usage will increase in the next five years[45]. In 35 states with 78% of the population, local agencies can now submit some documents to state agencies electronically, and 44 states with 95% of the population believe there will be more document types submitted electronically in five years[46]. *Sharing through document exchange will not run counter to trends and expectations.*

Only 14 states with 49% of the population use digital signatures now, but 45 states with 94% of the population believe that digital signatures will be more broadly used in five years[47]. The U.S. Con-

gress is considering the definition of national standards for digital signatures. Such standards would be helpful. *The use of digital signatures is consistent with current directions and expectations.*

Thirty-six states with 81% of the national population report that that some agencies use the *Internet* to access their own databases, and 43 states with 88% of the population believe that such Internet use will increase in the next five years[48]. Thirty-six states with 68% of the population report that some agencies already use the Internet to link with other agencies, and 41 states with 81% of the population believe that such Internet use will increase in the next five years[49]. *Internet use for information sharing is consistent with present experience and expectations.*

Thirty-four states with 73% of the national population report current use of *encryption* and 43 states with 89% of the population expect the use of encryption to increase in the next five years. *Use of encryption for information sharing is consistent with present experience and expectations.*

The use of Internet standards should permeate every aspect of the information sharing infrastructure.

General Strategy

How are we to get from here to there? *Here* has a very large number of governmental information systems, mostly operating in splendid isolation, although there are notable examples of sharing using modern telecommunication technology (e.g., NLETS, NCIC 2000, IAFIS) and tagged-field message syntaxes (electronic fingerprint and arrest information submissions to the FBI, interstate criminal history record). *There* has the same very large number of governmental information systems, plus many more information systems serving private organizations and individuals, in an information-rich interconnected environment, exchanging structured data for many different purposes, all via a single basic sharing technology.

The dimensions of the general strategy are these:

- Rapid development and endorsement of a *Concept of Operations* for nationwide sharing of governmental information;
- Rapid development and endorsement of a full set of *technical standards* needed for implementation of the Concept of Operations;
- Rapid development and endorsement of a specific set of sharing *message specifications*, with early focus on criminal justice needs and realization that the specification set will grow as the scope of sharing grows;
- Quantitative description of the existing telecommunications infrastructure available nationally at the local, state, interstate and state-to-federal levels for the sharing of governmental information;
- Support for the enhancement of this infrastructure to support sharing of the kind described in this report;
- U.S. Department of Justice to take a leadership role in encouraging state and local participation through special conditions on grant programs for Information Technology (IT) projects;
- U.S. Department of Justice to take a leadership role by implementing standard-compliant alternatives[50] to every local/state-to-DOJ and local/state-from-DOJ information flow;
- State CIOs to take a leadership role by building compliant statewide networks serving all local and state governmental entities;
- State CIOs to take a leadership role by using all available methods to encourage intra-locality information sharing.
- Workshops, tutorials, press coverage and other methods to convince the vendor community that this is the path to the future.

Next Phase Activity Areas

Very substantial progress can and should be made in each of the following activity areas during the next 24 months:

- **Activity Area**: Widespread *evangelization* concerning information sharing, based on a solid business case for sharing and a high-level concept of operations. The primary evangelization targets should be (a) membership groups that represent the governmental people affected by information sharing, (b) specialized press covering governmental interests, (c) agencies that regulate and make grant funds available for information system development in criminal justice and all other government areas, and (d) legislators and their staffs at all levels. (NASIRE to participate but not lead in this area except in area (d).)
- **Activity Area**: *Endorsement* by all major membership groups representing criminal justice information systems of the exchange-point model for criminal justice information exchange. (NASIRE to participate but not lead in this area.)
- **Activity Area**: *Definition* of major information-exchange documents required to accomplish sharing at the exchange points. (NASIRE to review contractor products in this area.)
- **Activity Area**: *Design* of a detailed technical Concept of Operations (ConOps) for a national criminal justice information sharing system capable of growing into a national governmental information sharing system; the level of detail must be sufficient for full understanding by CIOs and by technology vendors. The ConOps should describe both the use and the governance of the nationwide sharing process. (NASIRE to take a leadership role in this area.)
- **Activity Area**: *Selection* of at least two counties for implementation of pathfinder sharing systems for criminal justice, the sharing to incorporate all local-to-local, local-to-state, local-to-federal criminal justice information flows. Counties should be in at least two states, should have small and medium sized criminal justice agencies, should have a mix of custom-built, vendor-product and manual criminal justice information systems, and should be chosen primarily for their willingness to undergo change. One of the counties might be in a state where local criminal justice data sharing is implemented at the state level. (NASIRE to participate actively in selection process.)

- **Activity Area**: *Implementation* of pathfinder systems with full federal grant funding for all changes at the local and state level, and federal commitment to implement changes needed at the federal level. (NASIRE to participate actively in selection of pathfinders and overview of projects but not in project leadership.)
- **Activity Area**: *Design and implementation* of an observer-agency process involving all governmental agencies within the pathfinder counties, as well as a national selection of criminal justice agencies. (NASIRE to take a leadership role in design and conduct of the process in this area.)
- **Activity Area**: An *evaluation* which includes at a minimum the costs of implementing the national criminal justice information sharing system, including but not limited to the telecommunication infrastructure. (NASIRE to review contractor work in this area.)
- **Activity Area**: Decisive *actions* by U.S. Department of Justice assuring that all information interconnections involving DOJ and local/state criminal justice agencies are made available in a form compatible with the standards and specifications produced by this initiative. In order to provide for a gradual national implementation while encouraging local/state agencies to get on board, DOJ could announce, for example, that the new formats will be optional for five years and, thereafter, the old format will be discontinued. (NASIRE does not have a role in this area.)

33 CIO Survey, Questions 1, 2.

34 CIO Survey, Questions 3, 4.

35 CIO Survey, Questions 11, 12.

36 CIO Survey, Questions 13, 14.

37 CIO Survey, Questions 15, 17.

38 CIO Survey, Questions 28, 29.

39 CIO Survey, Questions 18, 19.

40 CIO Survey, Questions 20, 21.

41 CIO Survey, Questions 24, 25.

42 CIO Survey, Questions 34, 35.

43 CIO Survey, Questions 51, 52.

44 CIO Survey, Questions 53, 54.

45 CIO Survey, Questions 84, 85.

46 CIO Survey, Questions 88, 89.

47 CIO Survey, Questions 92, 93.

48 CIO Survey, Questions 94, 95.

49 CIO Survey, Questions 96, 97.

50 The purpose of this initiative is to show federal-level resolve and to provide opportunities for state and local entities to apply the new technology.

Glossary

The definitions provided here are meant to be specific to the information sharing environment being discussed in this report. No attempt was made to define the terms in a way applicable to the entire information technology environment.

Agency: A governmental unit; in the narrowest sense, a governmental unit of the executive branch.

ANSI: American National Standards Institute. See Standards Organizations.

Architecture: Those characteristics of a network, operating system and/or application program which facilitate information interchange.

Authentication: Any of the methods used to assure that the alleged source of the received data is the actual source, and that the message received is the same as the one sent in every respect.

Branch: In this report, this generally refers to the judicial branch of government.

CIO: Chief Information Officer. In this report, CIO is the highest state-level person responsible for policy concerning information systems and telecommunications systems.

CJIS-WAN: Criminal Justice Information Services Wide Area Network. A nationwide state-to-federal network operated by the FBI to serve fingerprint-based information exchange.

Collective Data: Distinguished from transaction data. A collective data object contains data from several transactions (e.g., an incident description plus all the outcomes of the incident, or all incidents in a time period with their outcomes).

Common Native Language: Information sharing technique which relies on multiple databases but a single data dictionary.

Concept of Operations: A description at a relatively high level of the participants in information sharing, the information flows involved and the functional requirements at each step of sharing.

ConOps: Shorthand for Concept of Operations

Consolidation: Information sharing technique that relies on tightly coupled application programs interacting with a single database.

CSS: Cascading Style Sheet, used to format structured data for display or printing. See also XSL.

CTO: Chief Terminal Officer. In each state, the single person responsible for intrastate connections with the information systems and networks provided by the FBI.

Data: The raw material of information. Data may be structured or unstructured; dynamic or static; textual or graphic. Raw data plus its associated metadata equals information.

Data Esperanto: Information sharing technique which relies on the ability of each sharing system to transform from its own database format to a single transfer format, and from that transfer format to it own format.

Data Schema: Definition of the permissible data to be included in a specified data element, or by extension to all data elements of a file, table or document. See Semantic Data.

Data Warehouse: Information sharing technique which relies on a separate database created by transforming data from several sources into a single database, along with application programs to retrieve the transformed data.

Digital Signature: Any of the methods used to assure that the alleged source of a message is the actual source, and that the attached message is the intended one.

Document: As used in this report, an information-exchange message for structured information; the document structure, data content and edit requirements are pre-defined before the information exchange takes place. See also Message.

DTD: Data Type Declaration in SGML and XML. See Semantic Data.

Encryption: Any of the methods used to protect the contents of a message from all but the intended receiver.

Entity: An information sharing unit. All agencies (see definition above) are entities; so are courts and legislative bodies. Private organizations which share (receive) governmental information are also entities, as are private persons.

Exchange Point: An event within a process at which information sharing does or should occur; either information is collected which is useful to another entity, or information from anther entity is needed, or both.

FBI: Federal Bureau of Investigation

Formatting data: Metadata which expresses the appearance of data on a page or screen. XSL and CSS style sheets contain formatting data.

Function: A capability of an application program, for example case initiation, meeting notification, decision outcome recording, etc.

Governance Model: A model associated with a specific ConOps which describes the rules for making decisions concerning ongoing operation of a system, in this case the information sharing system. Items include requirements for participation, placement of and services to be provided at the central site, expenses and payments, adoption of and changes/additions to the sharing vocabulary or document list, sanctions for inappropriate behavior.

HTML: A message tagging method used to instruct concerning the appearance of the message, primarily on screen. See Markup Language.

HTTP: Hyper Text Transmission Protocol. See Transmission Protocols.

IAFIS: Integrated Automated Fingerprint Identification System

III: Interstate Identification Index

IETF: Internet Engineering Task Force. See Standards Organizations.

Information: See Data. Data and Information are sometimes used interchangeably if that can be done without affecting clarity.

Information Architecture: See Architecture.

Information System: Computer hardware, software and personnel directed toward the collection, organization and dissemination of information.

IP: Internet Protocol. See Transmission Protocols.

ISO: International Standards Organization. See Standards organizations.

ISO 8859-1: International standard for an 8-bit character set. First 128 characters (7 bits) are the same as ANSI/ASCII codes.

LDAP: Lightweight Directory Access Protocol. A standardized way to connect with a directory which might hold passwords, addresses, public encryption keys, and other exchange-facilitating data.

Markup Language: A method of providing context to a message. The context may provide a description of how each portion of the message should appear on paper or in print (SGML, HTML, XML) or the semantic data (q.v.) for each portion of the message (XML). The method for providing the context is to enclose each message portion in beginning/end markers called tags, hence the description tagged-field formats.

Message: Means the same as Document in this report, other times can refer to unstructured data requiring human inference for interpretation.

Metadata: Data about data. There are at least three types of metadata of

interest in the context of this report: semantic data, which gives the meaning of the "raw" data; formatting data which describes the appearance of the data on-screen or on-page; and intellectual property data which describes data ownership conditions.

MIME: Multipurpose Internet Mail Extensions. A set of Internet standards used to express, in email format, data which does not fit the limitations of the basic standard.

Multi-Functional: Pertaining to an information exchange which crosses between two entities which have different operational objectives. School-to-probation exchanges are multi-functional; school-to-school district exchanges are not.

NASIRE: National Association of State Information Resource Executives, representing the chief information officers of the states.

NCIC: National Crime Information Center. An information system and nationwide network serving local, state and federal law enforcement agencies.

NLETS: National Law Enforcement Telecommunication System. Not-for-profit organization which provides a nationwide network for interstate communication among local, state and federal criminal justice agencies.

OMG: Object Management Group. See Standards Organizations.

Open Standard: Standard arrived at under the aegis of a Standards Organization (q.v.). So-called proprietary standards are not open, nor are most so-called industry standards.

Publish/Subscribe: An information sharing modality in which the subscriber user indicates a desire to be informed if certain events occur affecting a certain person, and the publisher system agrees to provide the information.

Pull: An information sharing modality in which an application program, upon sensing the occurrence of a specified event, automatically requests specified information from another information system.

Push: An information sharing modality in which an application program, upon sensing the occurrence of a specified event, automatically sends specified information to another information system.

QOS: Quality of Service. A guarantee of service quality for an information or telecommunication service; it may include promises concerning time between failures and time to repair failures, minimum bandwidth availability, database accuracy or other measurable descriptors of the service to be provided.

Query/Respond: An information sharing modality in which the human user of an application program requests specified information from another information system.

Rosetta: Information sharing technique which relies on the ability to transform from any one of many database formats to any other of the same large number of database formats; a many-to-many transformation capability.

SEARCH: SEARCH, The National Consortium for Justice Information and Statistics. Not-for-profit membership organization.

Semantic Data: Data about the meaning of the data in a message. This can be expressed as a data schema, a data dictionary, an XML or SGML DTD (Data Type Declaration).

SGML: A legacy tagging standard. Its progeny include HTML and XML. See Markup Language.

Signature, Digital: See Digital Signature.

Standards Organizations: Organizations which have defined procedures for the determination that a standard is necessary, the creation of a standard, and the periodic review of a standard. ANSI, ISO, IETF, WWWC (W3C), OMG are examples of standards organizations.

Structured Data: Data (q.v.) which carries with it the associated semantic data (q.v.) or a pointer to it.

Style Sheet: A method for describing the appearance of a document in print or on screen, such as CSS and XSL.

Tag: A marker within an information exchange document which points to a full description of the semantic data associated with the tagged data. <DateBorn> is an XML tag which points to a precise description of how to read the numbers which follow it as a date, and which date it is (the date of birth).

Tagged-field: Tagging is a method of imposing structure on a document. Each information field has a tag; each tag has a name which points to data-dictionary-like information such as meaning and edit criteria.

TCP: Transmission Control Protocol. See Transmission Protocols

Transaction Data: The descriptors or attributes of a single activity (e.g., the court disposition transaction data includes the court name, data, case, charges, decisions, sentences).

Transmission Protocols: Transmission protocols provide the mechanism for the transfer of information. IP controls transmissions between networks and is the fundamental mechanism of the Internet and many large WANs. TCP provides a mechanism for information transfer on a single WAN, and is often used with IP. HTTP is the key transmission protocol of the world wide web and provides for hot links to a URL.

URL: Universal Resource Locator. A unique address for a page on the world wide web.

USDOJ: United States Department of Justice

USDOJ/BJA: United States Department of Justice, Bureau of Justice Assistance. Administers federal grant programs for state and local criminal justice purposes.

W3C: World Wide Web Consortium. See Standards Organizations.

WAN: Wide Area Network. See Transmission Protocols.

XML: A message tagging method used to mark up a document with semantic data and style data. See Markup Language, XSL.

XSL: XML Style Sheet, used to format structured data for display or printing, and also to convert from one XML form to another.

NASIRE

167 West Main Street
Suite 600
Lexington, Kentucky 40507-1324
www.nasire.org