Document Title:	InfoTech Methodology For Data Integration
Author(s):	SPAWAR System Center
Document No.:	210416
Date Received:	June 2005
Award Number:	2001-RD-R-061

InfoTech Methodology For Data Integration

Dr. John Hoyt
Bruce Baicar

Table of Contents

Introduction

The critical need for public safety information sharing has prompted several technology efforts over the last decade. Many approaches have been tested and some successful tools have been developed. As a result, regions planning to undertake information sharing technology development efforts have several approaches to choose from. The law enforcement domain is an optimum situation for information sharing technologies. Communications infrastructures are continually established, and as such, action is required to effectively use these networks to their full potential. Information technologies that are best suited for the law enforcement domain must be evaluated and implemented in a cost-effective manner. Unlike the Defense Department and other large federal agencies, individual jurisdictions at both the local and state level cannot afford to expend limited resources on research and development of prototype systems. Therefore, we must identify enabling technologies that have matured in related domains and transition them into law enforcement at a minimum cost. Crucial to this measure, is the selection of the appropriate levels of information sharing technologies to be inserted.

Data storage mechanisms in the public safety domain differ greatly in the type and level of sophistication. The range of data storage mechanisms covers the full spectrum. Some jurisdictions maintain data in low-level databases such as Microsoft Access or any number of versions of Dbase. At the other end of the spectrum are large agencies that require major storage mechanisms to handle the volume of data collected by law enforcement personnel. In many cases old mainframe computer systems are still in place to handle the volume of data and access to the stored information is quite limited. Falling in the middle are medium to large jurisdictions that have implemented data storage mechanisms such as Oracle, Sybase, or Progress just to name a few. Information sharing technologies traditionally fall within two approaches: the data warehousing approach and mediation across distributed heterogeneous data sources.

The InfoTech program was initiated to answer the need for an integrated and secure information network solution to facilitate sharing of public safety information. In order to accomplish this goal, the InfoTech team set out to establish a global search capability, allowing criminal justice (CJ) professionals to query all relevant CJ repositories, and quickly obtain critical information.

The InfoTech program chose to use a technical approach providing mediation across distributed heterogeneous data sources, so that tailored data integration would satisfy specific state and local jurisdictional needs. In addition, this was layered with distributed, rule-based commercial security integration. Critical to the success of any shared data integration project is the development of shared standard data objects. To this end, NIJ has developed a set of standard data objects to which criminal justice agencies can map their own data objects through which the software can then share the information across data sources, irrespective of their specific data schema.

Compelling Need

"Police work is all about getting information. And the easier it is to access, the more likely they are to use it."

Mark Calhoon
Newport News Police Department

The National Institute of Justice's (NIJ) critical needs assessment of state and local Criminal Justice Agencies identified information sharing as the number one priority for information technology (IT) solutions. In response to this need, NIJ created the InfoTech Program to enable affordable information sharing solutions for criminal justice agencies. Early program goals and priorities were set based on needs expressed clearly by state and local law enforcement agencies. Specific stated needs included: secure access to local, state, and national law enforcement information, immediate access to pertinent information across multiple jurisdictions, critical information disseminated securely and rapidly to officers on patrol, and reduction in the amount of voice traffic over already overcrowded channels. Operational users overwhelmingly stated that quick and complete information is critical to officer safety, victim protection, offender treatment, and the effective prosecution of criminal justice.

InfoTech Vision

Early project vision took into consideration both operational needs and emerging state-of-the-art technical capabilities. In an effort to provide the right information to law enforcement officers at the right place and at the right time, the InfoTech program set goals to increase law enforcement situational awareness by providing context-driven information access. Further goals were to increase officer safety by highlighting information determined by the region to be critical, to increase effectiveness by providing information where it is needed most, regardless of officer location, and to provide access to existing law enforcement sources at all levels through rich information search, retrieval, and display technologies. The InfoTech team did not plan to provide just another single-source repository, requiring maintenance and additional operational costs. NIJ was committed to providing a low cost, secure solution that would allow agencies to keep their existing systems, while still providing rich cross-agency information search and retrieval capabilities.

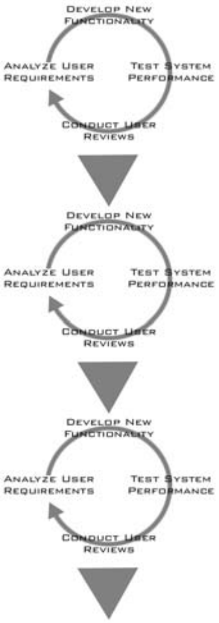
Project Approach Hypotheses

The InfoTech team based their technical approach on several hypotheses, to include these operational assumptions: new technologies should not require agencies to replace existing data sources, information sharing technology should provide role-based access-control to support the rigorous security requirements of criminal justice agencies, and information should be available to officers in the field. Another proposed hypothesis stated that it would be possible to define a set of grass-roots data objects that would be applicable to state and local criminal justice agencies across the U.S. The InfoTech team also believed that an iterative development process was the best way to deliver valuable technology.

Development Process

The InfoTech team develops and implements software utilizing an iterative cycle, incorporating testing and operational user feedback into every delivery. Key to a successful technology transition was creating involved teams of law enforcement operational experts, jurisdiction IT staffs, and experienced software developers. InfoTech software developers work directly with operational user groups, under the guidance of local jurisdiction leadership. This allows the developers to get direct user inputs and better understand the elements of information critical to regional officer safety. Developers also collaborate directly with the information technology staff so that the developers learn the subtleties of the individual data systems and IT staff members understand the technology being implemented. The key elements to making this process effective are committed leadership, rapid iterative development, and active participation from operational users and technical IT staffs.

The most operationally useful systems are created when law enforcement leaders and experienced technologists work together to set program goals and milestones. This collaboration ensures that the most important functionality is incorporated first, and that the system is delivered on time and on budget.


Develop New Functionality
Analyze User Requirements
Test System Performance
Conduct User Reviews
Develop New Functionality
Analyze User Requirements
Test System Performance
Conduct User Reviews
Develop New Functionality
Analyze User Requirements
Test System Performance
Conduct User Reviews

Using a rapid iterative development cycle, the team can analyze user requirements, develop new functionality, test system performance, and conduct user reviews. InfoTech's tight development cycle requires developers to test regularly in operational environments. These technology drops provide increased and reliable functionality to users.

As the system is developed, demonstrations are shown to operational users and agency IT staff for feedback. The comments from these reviews and tests provide critical quality control, and are invaluable towards implementing a system that provides significant operational benefits.

When leadership is committed, and operational users and IT staffs are actively involved, rapid iterative development results in key technical solutions. It allows users to see the impact of their inputs soon enough to change/refine the requirements, and provides a clear mechanism for developers to receive IT staff input.

Technology Impact

1998-1999

Rather than the InfoTech team seeking initial sets of users to act as testbeds, criminal justice community users were excited enough to come to them. Broward and Brevard counties in Florida volunteered and were chosen as the first operational testbeds. The primary tools of the initial InfoTech prototype provided operational users in Florida with single-point access to multiple heterogeneous data sources through a simple web-based interface. This new functionality significantly improved an officer's ability to quickly and effectively access the tremendous amount of available information. Although it provided exciting new capabilities, InfoTech was originally a custom built data sharing solution that was very static in nature. New sites, data sources and data objects had to be hard coded by developers into the system, and the user interface was not flexible. Remarkably, to this day operational personnel use that initial prototype system daily.

The InfoTech team also developed the Drivers License Image Retrieval System (DLIRS), which provided criminal justice users with secure on-line access to driver's license information and photos. This system is available via the Florida Criminal Justice Network (CJ-Net).

Public Safety Security Requirements
A fundamental requirement of all Public Safety systems is that they must enable users who may not know each other to establish a trust relationship and communicate securely. For InfoTech, several approaches were investigated. For the implementation in Florida, a mechanism was chosen that could issue electronic credentials that:

- Can be verified by multiple applications
- Can be used in an online transaction
- Are commercially supported in web browsers and other applications
- Satisfy State requirements for audit

An architecture that employed a public key infrastructure (PKI) was chosen to fulfill this purpose.

The InfoTech deployment in Florida was originally made up of a number of components distributed across the state. These components include the PKI, LEADS web server tie-ins to county networks, and the CJNET backbone.

The assurance that only authorized users have access to the DLIRS application is controlled by the use of digital certificates issued from FDLE's Public Key Infrastructure (PKI). Network administrators require tools and methods for monitoring and preventing unauthorized access into systems. Due to increasing threats from external connections, and the need to protect data, organizations have indicated a requirement for a secure communication solution to access various systems, applications and data. Organizations are looking for the capability to safely and reliably identify users and provide secure communications via the Internet. Users and organizations want the assurance that information assets are protected and that the security functions associated with the protection of information are provided:

- Authentication – proof that the sender is who they claim to be,
- Confidentiality – information is kept private,
- Authorization – protect against unauthorized use,
- Data Integrity – verification that no unauthorized modification of data has occurred, and
- Non-repudiation – assurance that the person sending data cannot deny participation in the transaction.

A Public Key Infrastructure (PKI) enables users of a basically non-secure public network to securely and privately transmit data across the network through the use of a public and private cryptographic key pair that is obtained and shared through a trusted authority. A PKI provides for a digital certificate that can identify an individual or an organization and directory services that can store and, when necessary, revoke the certificates. PKI protects your information assets in several essential ways:

- Authenticate identity – Digital certificates issued as part of your PKI allow individual users, organizations, and web site operators to confidently validate the identity of each party in an internet transaction
- Ensure privacy – Digital certificates protect information from interception during internet transmission
- Authorize access – Digital certificates replace easily guessed and frequently lost user IDs and passwords to streamline intranet/internet log-in security
- Authorize transactions – With PKI solutions, your enterprise can control access privileges for specified online transactions
- Verify integrity – A digital certificate ensures that the message or document the certificate "signs" has not been changed or corrupted in transit online and
- Support for non-repudiation – Digital certificates validate their users' identities, making it nearly impossible to later repudiate a digitally "signed" transaction.

The selection of the PKI was fundamental to the overall security architecture of Florida's InfoTech implementation. The chosen PKI needed to meet many requirements to satisfy the programmatic objectives and to provide adequate security. The following guidelines were used in choosing the PKI:

Must enforce strong authentication	Use digital signatures for controlling access to data and networks Support the RSA signature and encryption algorithm Support revocation of certificates
Must provide adequate privacy	Avoid overkill Allows migration from software to hardware tokens e.g., smart cards Support commercial grade cryptographic algorithms
Must be scalable	Must scale from small pilot to large deployment Impacts on Public Key Infrastructure must be

	accommodated
Must support key recovery	Law enforcement concerns must be addressed
Must be low cost	Low per-user cost Minimize hardware requirements Minimize overhead due to limited bandwidth Use of software based security when appropriate and acceptable
Must leverage commercial offerings	The PKI must support commercial-off-the-shelf (COTS) products such as Netscape and Internet Explorer.

InfoTech supported FDLE by investigating how the data-sharing applications being developed could use digital certificates to provide role-based access control.

Integration of a PKI into a process or workflow requires an understanding of the nature and number of changes that will need to occur. In 1998 at the onset of InfoTech, PKI deployments were growing, but not generally available. Thus, many applications supported the use of digital certificates differently. The de facto authentication scheme of the time was Username and Password. The initial applications had to be converted from this type of user authentication to digital certificate based.

To date, approximately 9000 digital certificates have been issued to the Florida law enforcement community. FDLE expects the system to grow to 20-30,000 certificates in the next few years.

2000-2001

In 2000, the InfoTech prototype underwent a major architecture revision that allows a more scalable and secure solution. InfoTech now provides a flexible solution that requires no changes to the existing system to incorporate new data sources. InfoTech provides dynamic discovery of new data sources (via JINI and the Query Server interface) that allows new data sources to instantly become available to users of the system. Also, the re-designed front end is flexible enough to provide a configurable interface to suit the target audience and target platform. Through a user-driven system design methodology, the initial data model from Florida was expanded to include additional criminal justice user contexts, to include courts and probations personnel. InfoTech was demonstrated to Florida court personnel, who validated that the concepts were beneficial to more operational users than just police officers, investigators and deputies.

The new InfoTech system architecture was implemented in San Diego enabling 38 agencies to seamlessly and securely share critical information. Authorized users in the region can now perform single point queries against multiple heterogeneous data sources, to include officer notifications, county wants and warrants, NCIC, the California Department of Motor Vehicles, and a regional booking photograph source. After less than a year of development, while the system was still in the first phase of operational testing, a San Diego detective was able to track down a suspected criminal using the InfoTech system. A crime analyst also reported finding an outstanding warrant on an individual that an unsuspecting officer was in-route to interview, and was able to get word to the officer in time for him to approach the individual more cautiously.

A demonstration system was developed in 2001 to provide secure single-point data sharing to Oregon's State Police, Department of Corrections, Youth Authority, and Judicial Department. This system marked the first time these state-level agencies were able to directly share information, providing their users with a significantly enhanced capability.

Also in 2001, the InfoTech team has initiated development of a certificate-based access-control capability, called Palladium. This technology was developed to provide administration, management and enforcement of access-control rules outside the context of a particular system, site or application. Palladium will be integrated with Florida criminal justice applications, utilizing the Public Key Infrastructure that the Florida Department of Law Enforcement has deployed. FDLE's PKI gives the users the capability to authenticate their identity through digital certificates. These digital certificates are electronic documents that are linked to a specific individual. Palladium will use information from the user's digital certificate to determine the access rights of the user. The first full integration of certificate-based access-control will be with the InfoTech technology currently being implemented to provide operational users access to all of Florida's jail management systems. Additionally, work is being done to provide a generic web server filter, to enable access control to a website to be brought under Palladium control. This will allow agencies to disseminate sensitive information quickly, securely, and only to the people who are authorized to view it.

From the very beginning, the InfoTech team has developed and evolved a set of well-defined data objects to facilitate data sharing. In 2001, these data objects were made available to any interested agency. Because these objects have been developed based on input from local, state and federal LE professionals, they capture the way front-line criminal justice professionals interact with data. This process of open and standardized data objects has successfully provided the foundation for state-state interoperability.

Data Sharing Standards and XML

Based on emerging criminal justice community interest in XML solutions, the InfoTech system was modified to allow XML inputs to the central query-mapping component. This capability is being evaluated in San Diego, CA, where an interface between the San Diego Police Department's Computer Aided Dispatch (CAD) system and the underlying InfoTech system is being considered. This capability would enable dispatchers to conduct a single query against many valuable legacy sources through their current CAD user interface (currently dispatchers must log-in to each data source separately to perform a query). These legacy data sources are a mixture of relational and network model DBMS's.

In order to provide third party application access via an XML interface, InfoTech technology accepts a socket-based XML interface. Exhibit 1 indicates the general structure of an installation using this interface:

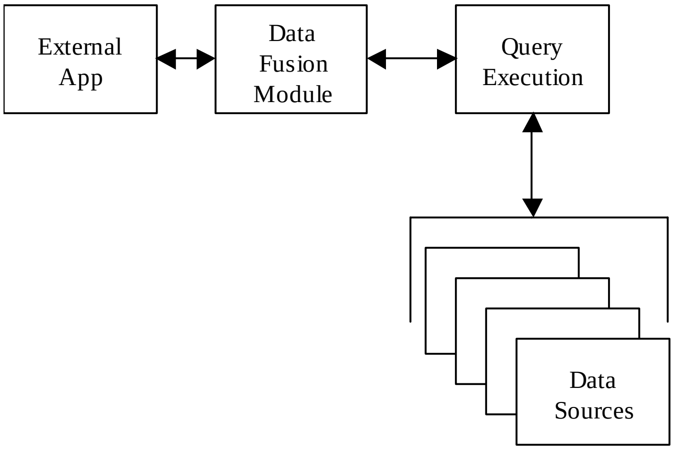


Exhibit 1. XML Interface Architecture

For this type of installation, the external application uses communication sockets to send XML-coded requests to InfoTech and to accept XML-coded responses. This protocol supports user interfaces such as an interactive user interface: This might be a computer-aided dispatch (CAD) system, an investigator's query system, or other system through which a user issues a query. The "external application" (from InfoTech's view) is initially responsible for accepting information from the user and formatting it into a legitimate InfoTech query, sent to InfoTech's data fusion module (DFM) via a new-query request. The external application formats the summary return and displays the result to the user. Subsequent drill-down or refresh actions on the part of the user cause the external application to send the appropriate request to DFM and to format the results appropriately.

In addition to providing an interface capable of supporting XML queries, the InfoTech team has developed a rich data dictionary of law enforcement standard data objects that can be used to generate an XML DTD valuable to many criminal justice agencies. The InfoTech team has found that it is possible to define a set of grass-roots data objects that are applicable to state and local criminal justice agencies across the U.S. The same data dictionary has been used to share regional data in California, Oregon, Virginia, Florida and South Carolina, requiring few changes to bring a new agency on-line.

Even in testing stages, InfoTech technology generated leads and information that improved officer safety and enhanced the abilities of law enforcement professionals to bring criminals to justice. In a significant vote of confidence, several regions requested follow-on efforts that they are internally funding.

Study Conclusions

Review of hypotheses

Sharing of information between agencies, without requiring changes to their existing data sources, has proven to be very valuable. Agencies that have data sources that currently support their internal needs should not be required to change their systems simply to share their information. Information sharing is critical to effective criminal justice, and every group involved with the InfoTech project has identified other jurisdictions/agencies in their region that are interested in InfoTech sponsored technology.

Criminal justice technology should provide role-based access-control to support the rigorous security requirements of LE agencies. Certificate-based access has been demonstrated as the best way to provide this capability. This approach is ideal because criminal justice information is sensitive in nature and often requires special safeguards (i.e. information on victim and juvenile records can only be viewed by certain authorities), and certificates are easily modified to support criminal justice professionals who often move between areas that have different levels of access authority. By providing the capabilities to specify roles and other ancillary information about a user, the system provides the capability to be flexible with the changing needs of the community it supports.

At the heart of the InfoTech effort is providing critical information to individuals at the right time and place, even if the officer happens to be in a squad car on the side of the road. Officers in the field have the need for a query capability that delivers cross-jurisdiction information, and limited bandwidth can be overcome if the technology is implemented properly. Information delivered to officers in the field provides increased officer safety by improving situational awareness, as well as clearing radio traffic for emergency situations.

The InfoTech team has found that it is possible to define a set of grass-roots data objects that are applicable to state and local criminal justice agencies across the U.S., requiring few changes to bring a new source on-line (the only changes required are if new types of information are being provided).

Even though it makes defining project scope more difficult, an iterative development process is the best way to deliver valuable technology. Users are able to impact the process and redirect development to ensure that minor misunderstandings do not result in an unusable system. By allowing users and IT staff to make mid-course corrections, the technology delivered is effective and usable, and the involved users are able to train new operators.

Summary of impact

Even in testing stages, InfoTech technology generated leads and information that improved officer safety and enhanced the abilities of law enforcement professionals to bring criminals to justice. A significant vote of confidence was provided as several regions requested follow-on efforts that they are internally funding. The InfoTech team was also nominated for a

Community Watchdog Award in California, touted as a project responsible for saving taxpayers money (based on personnel time savings and a cost-effective technical approach). Many regions had been attempting to solve agency-agency interoperability problems for decades prior to InfoTech providing a successful solution.

Lessons Learned

Strong Regional Leadership

Having an active regional leader involved throughout the effort is necessary to navigate the project to successful completion. Active leadership provides a consistent point of contact for operational users, IT staff, and InfoTech management. A strong regional leader also ensures that project goals are consistent with long-term organizational/regional needs. Finally, a dedicated leader helps overcome the inevitable political obstacles involved in agency-agency data sharing.

Committed Operational Users

Including a consistent set of operational users from the beginning of a project is critical to operational adoption of new technologies. This ensures technology will meet operational needs, and makes users part of the team, which helps promote use of the new technology. This approach also creates a group of trained users that truly understand how the system functions.

IT Staff Involvement

IT staffs need to be intimately involved in the development process, so they have the opportunity to share their insights and learn about system capabilities. IT staff members have significant, and valuable, insights into existing data systems, and how operational users use them.

Customer Education

Even today, FDLE estimates that approximately 60% of the law enforcement community in Florida understands what a digital certificate is and how it is used. Additional up-front marketing of the concept and benefits, tailored to the law enforcement community, might have generated a demand for these services earlier.

Make the Security Fit Existing Business Flow

In Florida, the standard configuration of the PKI installed for InfoTech required a manual review of each certificate request at a Registration Authority (RA). If the users were remote, the PKI vendor's concept was to deploy remote RAs each tied back to a common Certification Authority (CA) at FDLE. While possible, this did not meet the FDLE business model which dictated that there be one RA located in Tallahassee and to have regional Point-of-Contacts (POCs) approve and forward each certificate request to the RA. The RA operator position was then redundant. Significant work was expended to make the PKI conform to the FDLE requirement instead of changing the normal mode of operations for FDLE. Additional upfront system planning might have identified this issue earlier and allowed more time for changes.

Iterative Development

Keeping flexibility in the requirements process allowed valuable ideas to be incorporated late in the development cycle, but also made it difficult to wrap-up a project phase for roll-out.

Federal Government Support

And finally, federal government assistance in the development of software for use by local and state criminal justice agencies provides key funding and programmatic support to ensure the successful end-to-end implementation of new technology solutions.