AFRL-IF-WP-TR-1998-1528

INFRASTRUCTURE FOR I3 TECHNOLOGY TRANSITION FINAL REPORT
Intelligent Integration of Information (I3) Initiative

Nancy B. Lehrer

ISX Corporation
4353 Park Terrace Drive
Westlake Village, CA 91361

DECEMBER 1997

FINAL REPORT FOR PERIOD 07 OCT 94 – 07 OCT 97

Approved for public release; distribution unlimited

INFORMATION DIRECTORATE
AIR FORCE RESEARCH LABORATORY
AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7334

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government related procurement, the United States Government incurs no responsibility nor any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell patented invention that may in any way be related thereto.

This technical report has been reviewed and is approved for publication.

Tod J. Reinhart

TOD J. REINHART, Project Engineer
Embedded Information Systems Engineering Branch
AFRL/IFTA

James S. Williamson

JAMES S. WILLIAMSON, Chief
Embedded Information System Engineering Branch
AFRL/IFTA

James S. Williamson
For
Eugene C. Blackburn

EUGENE C. BLACKBURN, Chief
Information Technology Division
AFRL/IFT

IF YOUR ADDRESS HAS CHANGED, IF YOU WISH TO BE REMOVED FROM OUR MAILING LIST, OR IF THE ADDRESSEE IS NO LONGER EMPLOYED BY YOUR ORGANIZATION, PLEASE NOTIFY AFRL/IFTA, BLDG 620, 2241 AVIONICS CIRCLE, WRIGHT-PATTERSON AFB, OH 45433-7334 TO HELP US MAINTAIN A CURRENT MAILING LIST.

COPIES OF THIS REPORT SHOULD NOT BE RETURNED UNLESS RETURN IS REQUIRED BY SECURITY CONSIDERATIONS, CONTRACTUAL OBLIGATIONS, OR NOTICE ON A SPECIFIC DOCUMENT.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1997	Final Report: 07 Oct 94 - 07 Oct 97

4. TITLE AND SUBTITLE
INFRASTRUCTURE FOR I3 TECHNOLOGY TRANSITION FINAL REPORT
Intelligent Integration of Information (I3) Initiative

5. FUNDING NUMBERS
C: F33615-94-C-1556
PE: 62301E
PR: B765
TA: 01
WU: 01

6. AUTHOR(S)
Nancy B. Lehrer

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
ISX Corporation
4353 Park Terrace Drive
Westlake Village, CA 91361

8. PERFORMING ORGANIZATION REPORT NUMBER
I3BAA-9401

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Information Directorate
Air Force Research Laboratory
Air Force Materiel Command
Wright-Patterson AFB OH 45433-7623
POC: Tod Reinhart, AFRL/IFTA, 937-255-6548 x3582

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-IF-WP-TR-1998-1528

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
The major goal of the Infrastructure for I3 Technology Transfer (II3TT) project was to show the impact of I3 technology on operational systems development and to provide a framework for measures and metrics which verify this impact. To this end, the II3TT project had two main foci: 1) To provide an infrastructure for I3 program organization and communication based on the then budding young web technology, and 2) To use ISX's ties with other major Defense Advanced Research Projects Agency (DARPA) programs to find marriages between specific operational problems and I3 technology and to develop, where appropriate, feasibility and proof-of-concept systems, which if successful, should lead to major funded efforts of their own.

14. SUBJECT TERMS
Information Technology, Intelligent Agents, Information Support, Knowledge Acquisition, Services Based Architecture

15. NUMBER OF PAGES
26

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

Table of Contents

Table of Figures

1. Introduction

The major goal of the ***Infrastructure for I3 Technology Transfer (II3TT)*** project was to show the impact of I3 technology on operation systems development and to provide a framework for measures and metrics which verify this impact. To this end, the II3TT project had three main foci:

1. To provide an infrastructure for I3 program organization and communication based on the then budding young Web technology.
2. To use ISX's ties with other major DARPA programs to find marriages between specific operational problems and I3 technology and to develop, where appropriate, feasibility and proof-of-concept systems, which if successful, should lead to major funded efforts of their own.

2. I3 Program Coordination

One charter of the II3TT program was to use the World Wide Web to create an environment for the management and coordination of I3 technology programs. The II3TT effort spun off a program called TRADE to investigate these tools further. During the life of II3TT, ISX maintained a web site for the I3 program including tools to organize end of year technical reporting.

2.1 TRADE (Technology Roadmap, Analysis, and Development Environment)

Conducted early during this effort, TRADE focused on developing tools for DARPA program managers to coordinate and manage advanced technology programs. The specific tools adapted to TRADE are the Acronym, Calendar, Contact, Document and Roadmap Servers. The tools help coordinate and transfer data between DARPA program offices that are often duplicated within organizations and ensure the group's ability to have consistently current information and versions of documents and other information. The I3 program was used to capture an initial set of user requirements. The TRADE prototypes have now grown into a set of tools used within the DARPA Information Systems Office (ISO). The program is now known as ISO World.

3. I3 Feasibility Prototypes

A major goal of this program, was evaluation of the impact of I3 technology on full-scale operational programs. In order to identify good candidates for I3 technology and evaluate operational impact, the II3TT developed a framework for the assessment and evaluation of programs. In addition, as a result of examining the application of I3 technology to many domains (including many which did not result in feasibility prototypes), a consolidation of the I3 reference architecture was developed called the Key I3 Services (KIS). These services are now being developed more fully in ISX's ACME program (Architecture for Component-based Mediation Environments) project also managed at DARPA by Dave Gunning. This section describes the I3 Assessment and Evaluation Framework, Key I3 Services Framework and four Feasibility Prototype Demonstrations.

3.1 I3 Assessment and Evaluation Framework

An early goal of the II3TT project was to create a framework to assess and evaluate the potential impact of I3 technology for a particular domain problem. To this point in time, the I3 initiative had been postulating that I3 technology would be most effective in domains where there was a wide variety of types of information sources (relational, text, simulations), where there was a disconnect between the data represented in the source and how the end-user preferred to use the data, and where either an information source changes over time (i.e., a change in data schema) or the source of information changes over time (i.e., a new, alternate source). The purpose of the assessment and evaluation framework was to codify these concepts and describe domain problems in these terms.

Figure 1 shows the framework for evaluation of the impact of I3 technology with three main axis: operation effectiveness, breadth of coverage, and maintenance or evolution over time.

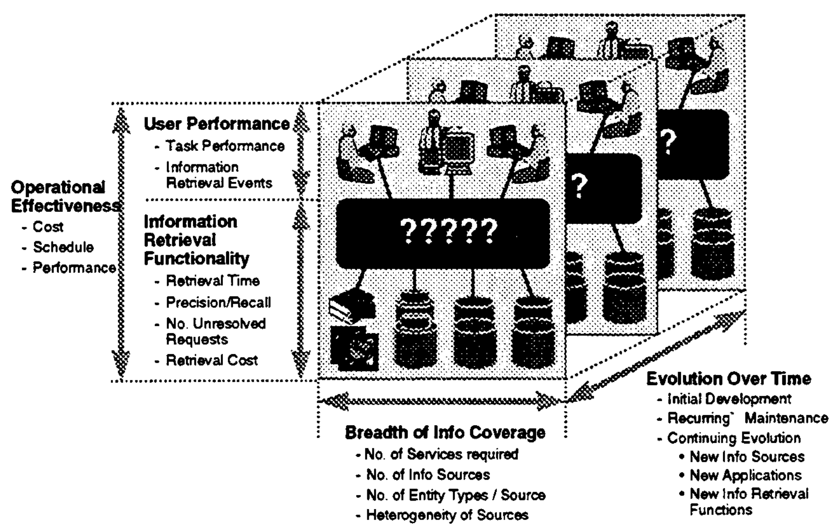


Figure 1. Axis of Evaluation

Figure 2 shows the economics of information access. The upper graph depicts the breadth of information coverage that can be achieved for a particular cost with and without mediation. The lower graph shows the frequency of queries. The area under the curve between the capability limit with and without mediation determines the number of unresolved queries, which occur without mediation (shown by the dotted vertical lines).

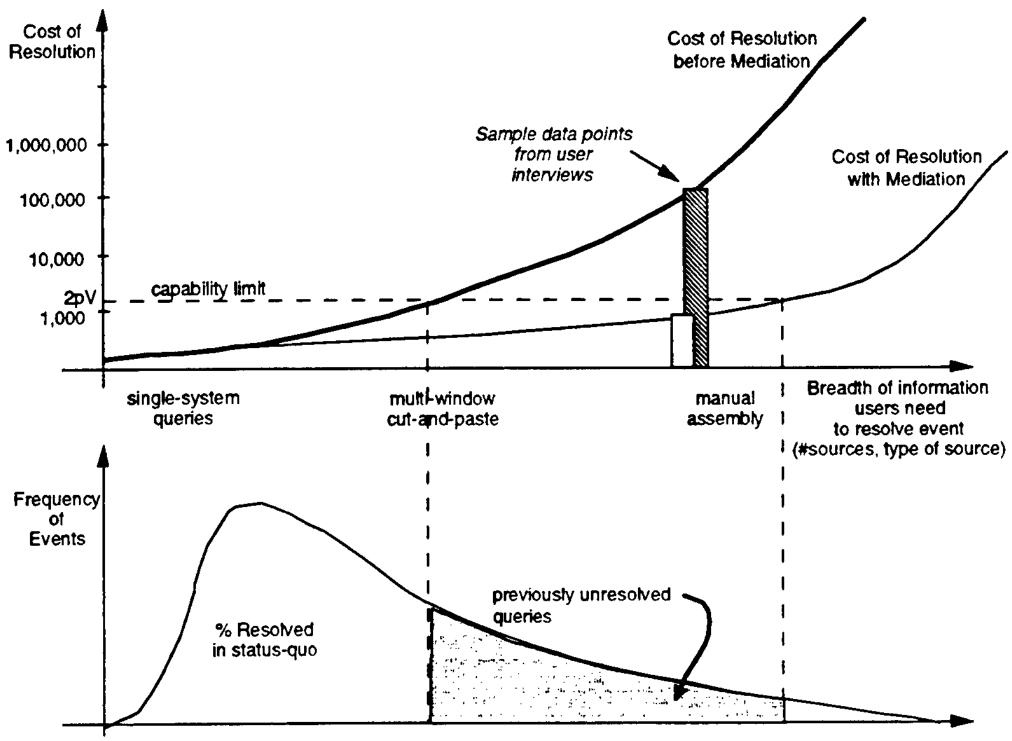


Figure 2. Query Economics

Finally Figure 3 illustrates the results of evaluating I3 technology insertion for the use in design for the F-22 airframe and for generating target lists in the Checkmate Air Campaign Planning Tool. We measured the benefits in four dimensions: quality, latency, cost, and maintenance and evolution over time.

Quality:

Quality measures the goodness of the information results. Accuracy, completeness, specificity, and match can measure this to user needs.

Latency:

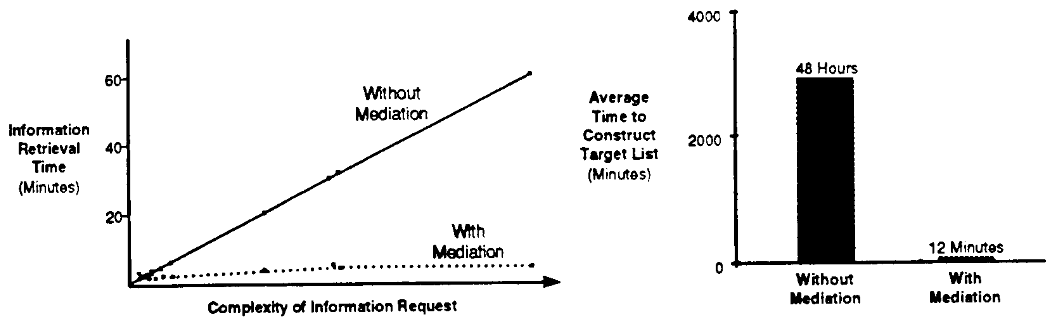
Latency measures the time elapsed between the identified need for information and the receipt of information. Factors affecting latency can include data source down time, bottleneck organizations for accessing information, data source complexity, and data source size.

Cost:

Cost measures the actual dollars or labor time required accessing information. Cost differs from latency in that it does not include idle time, i.e. time that is spent waiting for a resource to become available. Additionally, whereas latency includes only one dimension for time, cost multiplies time when more than multiple resources are working parallel. Addition factors effecting cost include connect time, expertise needed to access data and software or systems required to access data.

Maintenance and Evolution over time:

A major concern in data-centric organizations is the maintenance of data access and update systems as data sources change over time. The cost and time required to maintain these systems as user requirements change, new data collection opportunities arise, or database systems evolve must be factored into an overarching information services solution.



Engineers from the F-22 Airframe Integrated Product Team (IPT) used mediated non-mediated information access methods to retrieve design specs and guidelines while designing portions of the forward fuselage

Air campaign planners used mediated and non-mediated methods to access target databases while constructing target lists to satisfy a specific air campaign objective

Figure 3. Evaluation of I3 Technology Insertion

3.2 Key I3 Services (KIS)

The set of I3 services defined by the KIS is critical to the future success of the I3 initiative. They must be easy enough to be used by legacy and conventional information clients, yet complete enough to be able to take advantage of the advanced features provided by various advanced technology components.

The I3 Service-based architecture is based on a layered approach to defining service components. The I3 service definitions (displayed in Figure 4) are built upon a foundation of information services, which define information request, update and subscription, protocols. Built from these protocols are the Repository Services, and Mediation Services.

The Repository Services define interfaces and protocols for the storage and representation of objects including persistence, replication, caching, and data monitoring. As a specialization of an information service, repository services also define the interfaces for information requests, updates and subscriptions. Example Repository Servers could include CAD repositories, text corpuses (design manuals, etc...), image repositories, stream sources, traditional relational and object oriented databases. Each of these repository specializations can define additional layers to the interface protocols of their inherited parent. For example Domain and Ontology Services will define specialized services for the assembly, classification, and meta-object reasoning.

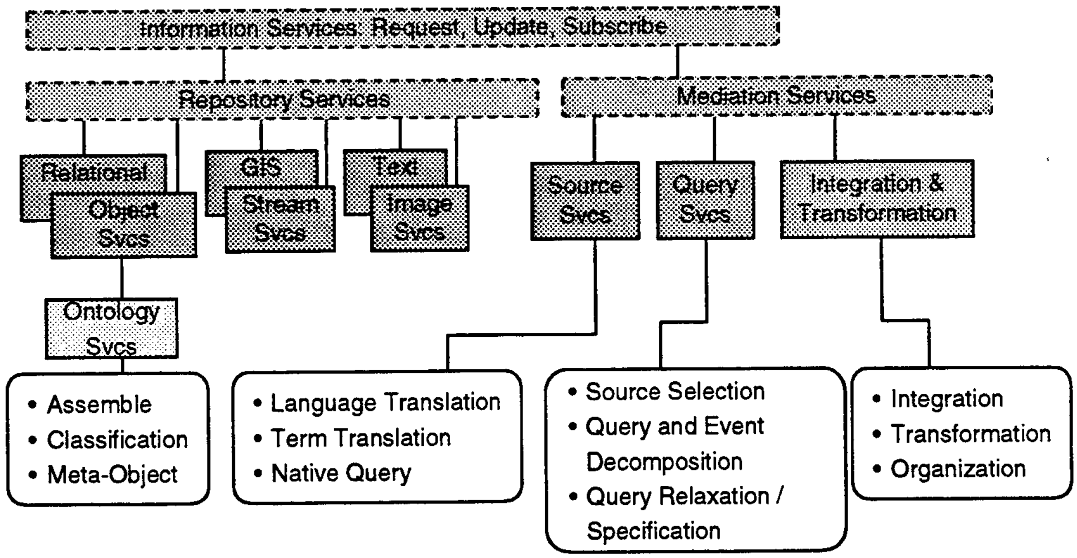


Figure 4. I3 Service Layers

Mediation services are comprised of Source Services, Query Services and Integration and Transformation Services. Source services provide those information services required for an information client to transparently access a specific repository. Typical functions include language translation (from an object-oriented language to say SQL), term translation (from a domain model representation to a source specific schema), and protocol translation (from an I3 IDL to a repository specific transport such as Oracle's SQL*Net). Source services are highly knowledgeable about the repository to which they communicate both in terms of its content (schema) as well as its communication protocol.

Query Services bridge information clients to source services by providing all means of query transformation, source selection, and query and event decomposition. Query services are highly knowledgeable about the terms in the users domain (via the Ontology Service) and the capabilities of specific source services (i.e., what kinds of questions can a particular source answer).

Finally, the Integration and Transformation Services provide (possibly) domain specific data correlation, translation, and fusion. Domain specific integration services might include unit conversions, name and place correlation (i.e., Ms. Lehrer is the same as Nancy B. Lehrer), and geographic correlation of information from multiple sources.

Figure 5 illustrates one scenario interaction between I3 services.

1. A query is sent from the Information Client to a Query Service
2. Modification is performed on the query to relax, specify, or translate query terms to/from one ontology to another. Ontology services may be used for this task.
3. The query is transformed from one multi-source query to several single source queries each directed to a specific source services. This may happen sequentially if information from one source is needed to direct a query to another source. Once again, ontology services may be used to accomplish this and other Query Service tasks.

4. Source services perform language, term and protocol specific transformations to the query and pass repository specific queries in repository specific languages using repository and operating specific protocols to retrieve query results. Ontology services may be used to accomplish this and other Source Service tasks.
5. Repository specific representations passed back to the source service for translation back into the organization and representation of the domain ontology.
6. Objects are passed back by each source service to the query service for object composition.
7. Query Services make use of Integration and Transformation services to perform object composition and possibly domain dependent correlation. Ontology services may be used to accomplish this and other Integration and Transformation Service tasks.
8. Composed objects are passed back to the query service for final management.
9. Domain objects are presented to the Information Client completing the cycle.

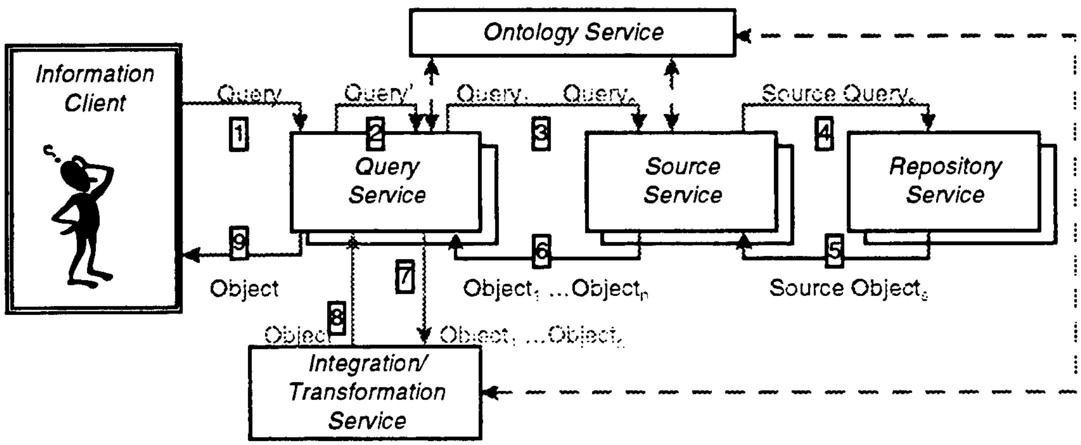


Figure 5. I3 Services Interaction

3.3 *Feasibility Prototypes*

The proof-of-concept systems, which were developed under II3TT, were called ***Feasibility Prototypes***. Their purpose was to identify specific operational challenges in multiple domains and to show the potential impact of I3 technology on these challenges. Impact was illustrated through the use a variety of techniques ranging from paper studies based on the metric framework developed under this program, no-operational system mockups, and highly focused operational prototype systems. The feasibility prototype efforts had the following goals:

- Glean I3 technology from the I3 community members and solicit their cooperation.
- Identify an I3 technology solution to be embedded in an on going DARPA system.
- Focus on specific full-scale operational problems using operational data and user requirements, but restricted and with integration into the larger system.
- Show a clear integration path into the full operational system.

The development vision is illustrated in Figure 6. I3 programs provide a technology inventory which is used to feed 10-month feasibility prototype efforts. Each effort, if successful, then feeds into a major operational program.

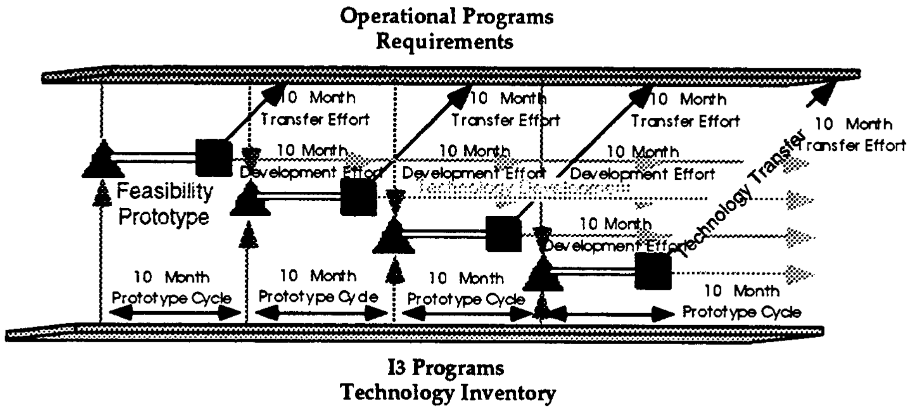


Figure 6. Technology Transition Roadmap

Four such feasibility prototypes were performed during the period of this contract: TRADE, ImInEx, QuickTurn, ISLE, and IWSDB. Each are described below in a short summary.

3.3.1 ISLE: Information Support for Law Enforcement

Under the II3TT, program an initial assessment was performed to evaluate the applicability of I3 technology for use by the FBI and local law enforcement agencies to develop and maintain case information. The result was a study using the framework for assessment and evaluation described previously in this document. This evaluation lead to the creation of a large program, funding by the National Institute of Justice and DARPA and performed under the II3TT contract vehicle. A full description of the ISLE development effort is attached.

3.3.2 IWSDB: Integrated Weapon Systems Database

The focus of the IWSDB project was to apply I3 technology to aid in the design and development of the F-22 airframe. In this prototype, we applied I3 technology from ISX and the Lockheed-Martin AI Center (LAIC) to be applied during the design of the F-22 airframe. Under the main funding of II3TT, ISX, LAIC, working with the Lockheed Aeronautical Space Company in Atlanta, Georgia, developed a small working system enabling design engineers to access several design, production, and bill of material data sources in a single integrated system. This system has since lead to a larger program for the F-22 and a program for the Defense Logistics Agency all aimed at providing integrated information access to aircraft design and procurement data.

3.3.3 ImInEx Summary by Ross Ashley

Image exploitation and image content understanding face great challenges in the next couple of years. As warfighters require more image information and as the volume of images increases exponentially, the Interactive Image Exploitation (InImEx) vision creates an environment in which image analysts can finally turn their attention to deriving target significant information rather than trying to discern images. InImEx provides a research and development vision for the image understanding community. The goal of this vision is to shift the analysts' focus from pixels to targets through the use of advanced image understanding and human computer interaction technologies.

To accomplish this task, three major milestones were achieved. First, research was conducted on current technology efforts within the image understanding community. Second, image analysts from the national and tactical community were briefed and interviewed on the evolving InImEx vision. Third, a WEB-based, interactive visual demonstration was developed for the image understanding community, complete with document server and a mechanism to provide further feedback.

Technology research was conducted to pole the image understanding community as to what research and development areas have the most potential pay off, both for the near and far term. Researchers from across the image understanding community were interviewed and briefed on the evolving vision to ensure that major areas of current research and potential new areas of research all provided value to the overall vision.

Image Analysts from the tactical community and managers from the national imagery community were briefed or interviewed to ensure that the InImEx vision provided a coherent goal for image understanding research and development. Results from these interviews and briefings can be found in the document server portion of the visual demonstration.

3.3.4 Quick Turn

The Quick Turn program was aimed at decreasing the analysis time for imagery exploitation by an order of magnitude by creating a collaborative, multi-media environment for image analysis. In this project, we applied mediation technology in order to allow an image fusion program (developed at Mitre Corp.) to transparently access from multiple image and relational repositories.

4. Summary

In summary, the Infrastructure for I3 Technology Transfer focused on creating a framework for assessing and evaluating the impact of I3 technology on real world problems. A major product of the II3TT program is the definition of the KIS framework which describes a methodology for defining I3 technology components in a manner in which they can be easily integrated into larger systems. This work will continue with a new DARPA contract called the Architecture for Component-based Mediation Environment (ACME).

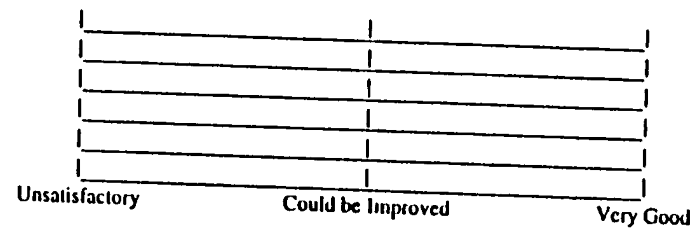
Domain Assessment Framework for I3 Technology Transfer

Nancy Lehrer, Allen Smith, Bill Swartout, Bill Mark

Operational Effectiveness

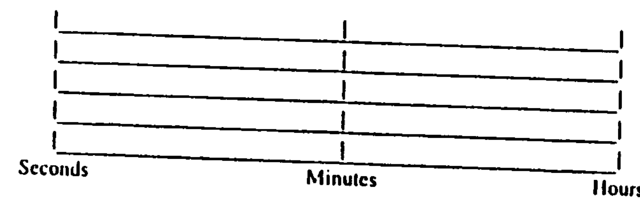
Quality. Under the current system, and within the time permitted, what is the users' assessment of the quality of the information integration service (either manual or automated) that is provided today?

- # of "hits" compared with the total universe of relevant data
- breadth of coverage; number of relevant sources tapped
- Completeness of information
- Correctness of response
- Timeliness of response
- Cost of response



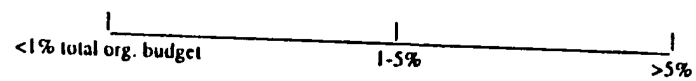
Elapsed Time. For the typical query, how long does it take to locate, access, compile, analyze, and report information? (o = average time, x = maximum acceptable time)

- Locate - where is the answer, what DB, text... who do I ask?
- Query generation and iteration
- Access - Query processing time
- Compile
- Analyze
- Report



Cost: What are the costs associated with information access and maintenance

Current cost to provide information integration services.



Complexity/Breadth of Information Coverage

Information Integration Functions. For typical queries, what does the user or the users' application have to do with information that is collected?

Simple query and reporting
Complex query
Filtering
Fusion
Inconsistency resolution
Abstraction
Analysis
Update data bases
Etc

Never — Sometimes — Always

Number and type of sources (Heterogeneity): For a typical query, how many different sources of information have to be accessed? (o = Typical x = Maximum)

Humans (count each as 1)
Paper files/books/manuals (count each location as 1)
Unstructured electronic data bases (i.e. electronic text)
Structured data bases (Count each as 1)

0 — 1-4 — 5+

Location of Information sources:

Human sources — Local / Geographically dispersed / Same Organization / Geographically dispersed/ Different Organization

Paper files — Local / Geographically dispersed / Same Organization / Geographically dispersed/ Different Organization

Unstructured data bases — Same Machine as User / LAN / WAN / Sneaker Net

Structured data bases — Same Machine as User / LAN / WAN / Sneaker Net

	Low	Middle	High
Size (Amount of Data):			
Humans (# of individuals providing information)	0	1-4	5+
Paper files (total # pages for all sources)	0	1,000	5,000+
Structured databases:			
total # of records (instances)	0	1000	500,000+
number of attributes desired per instance	0	1-10	5+
Complexity (structured databases):			
Number of tables	1	10-20	100+
Average number of columns per table	1	10-20	100+
# of joins	0	1-3	5+
Degree of heterogeneity of sources used for a typical query			
# of different types accessed	1	10-20	25+
# of different schema	1	2-3	5+
Security/Authorization Requirements	None	Single Level	Multi-level

Evolution/Maintenance

How often do schema/data definitions change?	>Yearly	Yearly	<Monthly
How often is new information added to information sources?			
Paper sources	>Yearly	Yearly	<Monthly
Unstructured data bases	>Monthly	Weekly	<Daily
Structured data bases	>Monthly	Weekly	<Daily
How often are applications changed that use the information?	>Yearly	Yearly	<Monthly
How often must new information retrieval/integration functions be added?	>Yearly	Yearly	<Monthly
How often are new information sources added?	>Yearly	Yearly	<Monthly

Cost

Cost of:			
Changes to schema/data definitions	0	??	??
Adding information to data sources			
Paper sources	0	??	??
Unstructured data sources	0	??	??
Structured databases	0	??	??
Adding Applications	0	??	??
Adding/Changing Information Integration functions	0	??	??
Adding new data sources	0	??	??

Advanced information management tools for investigation and case management support in a networked heterogeneous computing environment

T E Clifton III (tclifton@ dc.isx.com)
Nancy Lehrer (nlehrer@isx.com)
Mark Klopfenstein (mklopfen@ dc.isx.com)
Belinda Hoshstrasser (belinda@east.isx.com)
Rachel Campbell (rcampbe@dc.isx.com)

ISX Corporation
2000 North 15[th] Street, Suite 1000
Arlington, Virginia 22201

ABSTRACT

The right information, at the right time and place, is key to successful law enforcement. The information *exists*; the challenge is in getting the information to the law enforcement professionals in a useable form, when they need it. Over the last year, the authors have applied advanced information management technologies towards addressing this challenge, in concert with a complementary research effort in secure wireless network technology by SRI International. The goal of the combined efforts is to provide law enforcement professionals the ability to access a wide range of heterogeneous and legacy data sources (structured, as well as free text); process information into digital multimedia *case folders*; and create World Wide Web-based multimedia products, accessible by selected field investigators via Fortezza-enhanced secure web browsers over encrypted wireless communications. We discuss the results of our knowledge acquisition activities at federal, regional, and local law enforcement organizations; our technical solution; results of the one year development and demonstration effort; and plans for future research.[1]

Keywords: Knowledge Acquisition, Services Based Architecture, Database, Text Processing, Case Management, Wireless Networks, World Wide Web, Fortezza

1. INTRODUCTION

The right information, at the right time and place, in the most useable form is key to successful law enforcement. Whether it's an officer responding to a call, an investigator at a crime scene, or a public affairs officer at a community meeting, accurate and up-to-date information can make a significant difference. Unfortunately, the information these police officials need is not always readily available. The information *exists*. The challenge is pulling it out of a host of different law enforcement systems (national crime databases such as NCIC, bookings, correctional databases, investigatory reports, Uniform Crime Reporting files, etc.); refining the data into valuable information packages; and getting these packages into the officer's hands.

The US Department of Defense's Advanced Research Projects Agency (DARPA) has invested considerable resources over the years in basic information technologies applicable to this challenge. DARPA's Operations Other Than War/Law Enforcement (OOTW/LE) INFOTECH program is funded by the Department of Defense/Department of Justice Joint Program Steering Group to demonstrate the application of these technologies to the OOTW and law enforcement domains. Dr. John Hoyt (FBI, assigned to DARPA) created a team in 1995 to use DARPA technologies in a demonstration of capability, to provide

[1] This effort was funded under DARPA's Operations Other Than War/Law Enforcement INFOTECH Program (Dr. John Hoyt, DARPA/TTO & FBI), with support from Dr. Allen Sears, DARPA/ITO; Mr. David Gunning, DARPA/ISO; Dr. Barry Leiner, DARPA/ITO (now of MCC); and Mr. Robert Ruth, DARPA/ITO (formerly US Army CECOM). A description of the INFOTECH program together with a demonstration of the functionality can be found at http://isx.com under the "Programs" section, "ISLE Project"; e-mail the authors for information regarding permission to access this site.

law enforcement professionals the ability to access a wide range of heterogeneous and legacy data sources (structured, as well as free text); process information into digital multimedia *case folders*; and create World Wide Web-based multimedia products, accessible by selected field investigators via Fortezza-enhanced secure web browsers over encrypted wireless communications.

SRI International took charge of the communications and security aspects of this effort. In order to answer the rigorous security needs of the law enforcement community, SRI International provided a secure wired and wireless communication capability. SRI used the Fortezza encryption technology developed by the National Security Agency (NSA) to ensure the security of the information transferred over wired and wireless communication paths. To meet specific law enforcement requirements for the ability to get timely information to officers in the field, wireless laptops were designed to communicate over a variety of channels, to meet the bandwidth requirements for certain types of information and the physical restrictions of the individual receiving the information. For example, an analyst working in a command center will have access to wired channels, while an officer in the field may be able to get images through a high-bandwidth satellite network or may be limited to text only over low bandwidth networks (e.g., secure pagers). Also part of SRI's contribution to the Infotech program, were Fortezza enabled firewalls and in-line encryption devices that allowed the use of public networks to exchange information securely.

The rest of this paper will focus on ISX Corporation's portion of the INFOTECH program, Information Support to Law Enforcement, or ISLE. We focused on the information management portions of the INFOTECH challenge - database access, case folder management, World Wide Web package creation and dissemination, and electronic submission. We will present the results of our knowledge acquisition work with the FBI, Washington/Baltimore High Intensity Drug Trafficking Area (W/B HIDTA), Baltimore (City) Police Department (BPD), and various other law enforcement organizations; describe our technical solution; discuss the results of our demonstrations; and present avenues for future research.

2. KNOWLEDGE ACQUISITION

2.1 KA, ACE, and USE: Successful transfer of advanced technology into operational domains, such as law enforcement, demands intimate domain expert involvement in the development of a future systems vision, a process we call Advanced Concept Engineering, or ACE, supported by intensive user-centered systems engineering (USE). Since the purpose of technology is to enable the human user to perform tasks better and more effectively, the user's information requirements, tasks, environment, and corporate culture must all be taken into account. Thus, knowledge acquisition (KA) is the cornerstone to successful technology transfer.

KA can be described as the process of:

- eliciting domain knowledge from domain experts,
- capturing and organizing knowledge in a useable form, and
- implementing the captured domain knowledge into an intelligent system.

The *useable form* we are referring to is one that is helpful to software engineers developing intelligent systems. A variety of different forms is appropriate, including: models, a knowledge base, or a system requirements specification. The documents, models, and analyses resulting from the KA process provide a bridge between the domain experts and the technologists. This captured domain knowledge is used to identify user and domain requirements and constraints, and design and develop an appropriate information system. The goal is to make these systems as helpful and easy to use as possible.

2.2 Techniques: The ACE process begins with an initial version of the future systems vision, essentially a hypothesis of what technologies might support which key domain requirements based upon a high level requirements screening. ISLE's initial concept involved supporting crisis management centers with a combination of technologies from three DARPA programs - Intelligent Integration of Information (I^3), TIPSTER, and Global Mobile Information Systems (GloMo) programs. I^3 involves technologies to access and exploit information from legacy and heterogeneous data sources; TIPSTER provides technologies for the processing of free text into databases: and GloMo is a research initiative focusing on radical improvements to mobile secure communications. I^3 and TIPSTER are described more fully in section 3. KA techniques are used to elicit further requirements which build upon the initial future systems vision.

Central to effective KA is choice of appropriate domain experts. Since the goal of the ISLE project was to build a system that would be applicable over a broad range of law enforcement agencies, the experts needed to be from a broad range of agencies. The experts we interviewed came from three major groups:

- Local Law Enforcement - the primary local law enforcement experts were from the Baltimore Police Department, although we did interview some experts from the Alexandria Police Department as well.
- Federal Agencies - Several groups from the FBI provided perspective from the point of view of a large federal agency.
- Hybrid Regional Groups - The Washington / Baltimore HIDTA was an excellent source of a variety domain experts. This unique group provided experts from a wide range of law enforcement agencies including: local law enforcement (Baltimore Police Department, DC Park Police); federal agencies (FBI, DEA); and the military.

We performed an initial round of structured interviews to obtain a foundation of domain knowledge. This broad understanding of law enforcement allowed us to define the initial concept of the case folder, identify an initial set of information requirements, and recommend a set of functions ISLE could provide that would be useful from the domain perspective. For each session we created narrative session reports, consisting of summaries of what was discussed or learned during the KA sessions. These reports were sent to the domain expert for review prior to dissemination to the ISLE team. The narrative session reports were used primarily by the Knowledge Engineers to perform the required analyses. The session reports were also a useful source for clarification and background information for the system designers and developers.

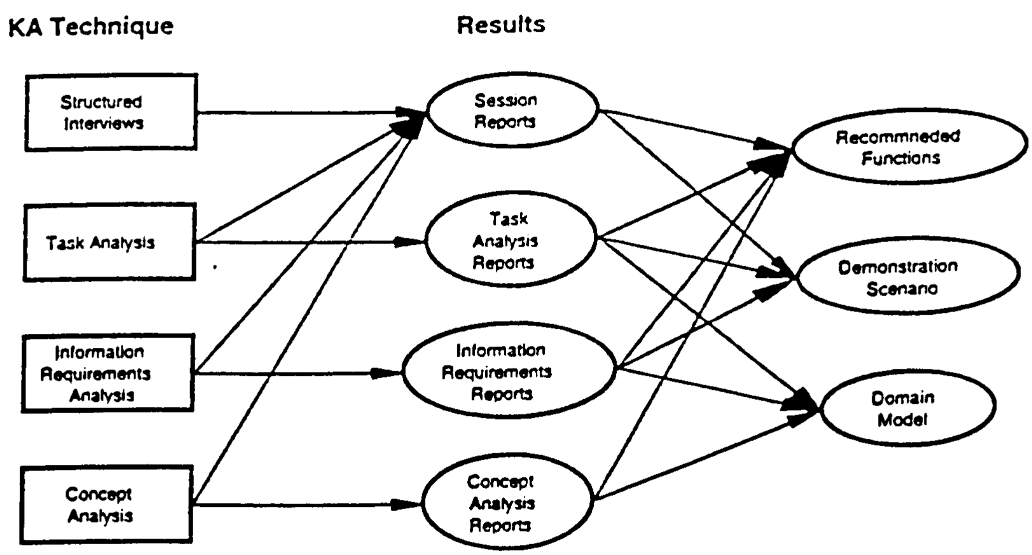


Figure 1. KA Techniques and Results

In follow on sessions, we focused on task and information requirements analyses. These sessions provided detailed information about how a law enforcement professional performs his or her job and allowed the knowledge engineers (KEs) to identify the information needed by the officer. Information requirements reports from these sessions include two main sections: a task hierarchy and the information requirements grid. The high level task hierarchy is simply a textual hierarchy showing the main functions performed by the expert. The task hierarchy forms a basis for specific information requirements. The information requirements grid identifies the information elements used in performing a task, the range of values of the information. and the possible sources. Identifying information requirements involved identifying the high level information needs of the users. then drilling down to discover the elemental information requirements. A suspect's criminal history is an example of a high level information need. The specific details of a crime, such as date. time, place. victim. offense. etc. are examples of elemental information requirements.

Task analysis was performed at a high level. just detailed enough to allow us to organize information requirements around tasks. Supporting task analysis was the more in-depth concept analysis technique, which typically required another session following the completion and expert approval of the reports and analyses from the task and information requirements sessions. The concept analysis sessions enabled the KEs to map the concepts used by the experts and the relationships between those concepts. forming the domain model. This model was captured in concept session reports. and evolved through successive KA sessions.

Understanding problems or difficulties in a domain. particularly those that relate to information requirements. enables us to identify functional requirements of an information system for the law enforcement officers and determine appropriate areas for the application of DARPA technologies.

2.3 KA and ISLE Systems Development: While the ISLE knowledge acquisition effort generated a number of valuable products and findings. its ultimate purpose was to provide input and support for the development of the ISLE system. and the greater INFOTECH project. The KA team worked closely with the technical development team to transition domain knowledge and provide guidance from the end user's perspective. The primary methods of communicating the distilled domain knowledge to the development team were:

- KE participation in the monthly design meetings,
- completed KA session reports.
- revised versions of the Recommended Functions for ISLE report.
- releases of the domain model following significant expansion and revision. and
- day-to-day interaction between the KE's and the technical team members.

Throughout the course of the project. the KA team actively solicited from the development team comments, requests, and suggestions on the KA plan. reports. and analyses in an effort to ensure that the KA team was providing maximum benefit. The result was a technical solution which satisfied the user's information. task. and environmental requirements..

2.4 Findings: Currently, most law enforcement organizations do not have the latest and best information systems. As a result, there are many problems in trying to use the wealth of information in existence. Information management challenges identified in the course of our KA efforts fell into four categories: database access, case management. information dissemination, and digital capture:

- Database Access: Many of the databases in existence in police departments run on large mainframe computers with limited connectivity. Also, most of the databases on the mainframe computers were created to support management or reporting needs, not to support criminal investigations. Most law enforcement agencies have access to federal databases maintained by the FBI. DEA. INS. etc. However, the officers who could benefit from the information may not make use of the information because of the difficulties encountered in accessing and using the information. Each database has a unique login procedure and query method; results are returned in different formats, often as summaries or a collection of hard to read attributes. The officers do not have the flexibility to ask for just the information they need. Instead, they must wade through voluminous, cryptic reports to locate the information of interest. Another difficulty with these databases is that each one must be accessed individually; there is no method available for issuing a single query to several databases at once. This limitation requires the officers to keep in mind what specific data elements are stored in each of the databases, so that queries may be issued to the databases most likely to contain the desired information. Many departments inside police departments have their own "home grown" databases to store information relevant to their investigations.

- Case Management: Most case folders are paper-based. The ease with which information may be located is dependent on the filing skills and discipline of the members of the investigation team Since the folder is a single physical entity, it is difficult to share effectively and efficiently across a team of investigators. In large cases, the case folder may actually grow to be the size of several filing cabinet drawers. Efficient searches over a body of reports to locate specific information are not possible. The officers cannot ask questions like, "Has the suspect in the case I'm investigating been investigated before?" Maintaining only paper-based reports and case folders renders information effectively inaccessible and therefore virtually unusable. Also, there are often key relationships between individuals, artifacts and places that cannot be shown in a traditional case folder.
- Information Dissemination: A significant amount of information could be made available to other departments, *if* they knew of its existence. Much could be gained if departments could easily share information and collaborate on cases. When conducting field investigations, officers have little or no ability to receive sensitive, case-related data. Investigations are sometimes delayed while sensitive information is hand-carried to the officer who needs it to continue.
- Digital Capture: Much of the reporting done by officers is still kept on paper only. Even in those situations where reports and case folders *were* stored electronically, often there were large barriers to effective use of the information. At some of the agencies we observed, the data entry process lagged the report generation process by as much as a full week.

3. ISLE TECHNICAL SOLUTION

In addition to providing knowledge acquisition services for the entire INFOTECH program, ISX had two main tasks:

- develop an extensible architecture within which a wide range of relevant DARPA information technologies could be tested against, tailored to, and delivered for the law enforcement domain; and
- test, tailor, and deliver an initial solution based on technologies from DARPA's I^3 and TIPSTER programs.

Our architecture and solution had to work in conjunction with secure wireless communication technologies from DARPA's Global Mobile (GloMo) Information Systems research initiative. For the first year's effort, we worked with the simplifying assumption that SRI would provide us transparent Internet-like connectivity to wireless mobile PCs; future research would involve more in-depth integration with the information management aspects of GloMo (e.g., modifying the task-level information management behaviors based upon *Quality of Service* features provided by the underlying wireless network.). The first year solution needed to address the four main law enforcement information management requirements identified by our knowledge acquisition effort: database access, case management, information dissemination, and digital capture.

3.1 Architecture: We designed the ISLE architecture to allow for the transition of advanced technologies, such as those being developed by DARPA, into operational systems. In the ISLE system, we have concentrated on information, collaboration, and mobile communications technologies. The architecture however is open to allow for the integration of new technology components as they mature and become available.

The ISLE architecture is comprised of many asynchronous and independent software agents requiring the support of specific tools to support agent interaction, communication, and cooperation. The Common

Object Request Broker Architecture (CORBA) provides these mechanisms. An architecture using CORBA protocols allows for a true peer-to-peer object-oriented interaction between software agents and interfaces. Software agent interfaces are described as object definitions which can be called from any platform containing a CORBA implementation. This allows for true interoperability across both development platforms (Unix, NT, Macintosh) and development languages (C++, Java, Ada, Smalltalk).

For the past several years, ISX has been developing multi-agent systems using CORBA and other tools. Although CORBA tools are still fairly young, our experience has proven that CORBA enabled peer-to-peer communication has greatly reduced development while improving and enabling system functionality. Our experience extends across domains including environmental restoration, hospital emergency room information management, and air campaign planning, and across system functionality from information integration systems to objective-based planners.

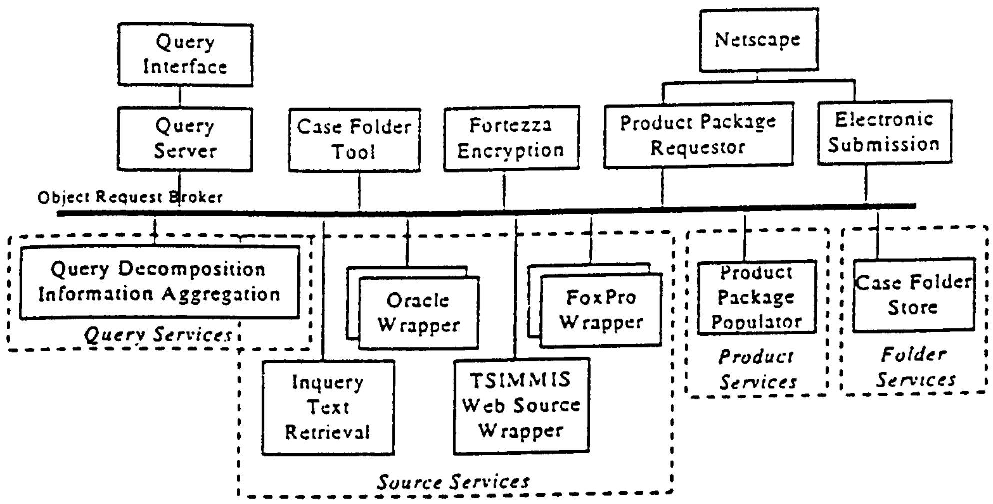


Figure 2. ISLE Services Based Architecture

An architecture is a framework for developing solutions to a class of domain applications. There are two distinct types of software components in the architecture: domain dependent components and domain independent, or general purpose, components. The domain dependent components or *applications* implement specific capabilities needed for one specific type of user such as the case folder interface for the Baltimore City Police Department. The domain independent components or *services*, implement a capability which is reusable across domains. Furthermore, services describe a general class of problem which can be solve in parts by multiple software components or with alternative implementation strategies or algorithms. Examples of services in the ISLE architecture include Query Services, Source Services, Product Services and Folder Storage Services. Figure 3 shows the application types and services implemented in the ISLE architecture.

3.2 Information Access: The single most compelling requirement identified during the course of ISLE was access to the widely varyingarray of information sources. To address this need, we integrated data mediation technology from DARPA's I^3 program and text retrieval technology from DARPA's TIPSTER program. The data mediation technology allows the semantic translation of questions in user terminology to a generic query language (*Query Services* in Figure 2). These queries are then decomposed into specific queries for individual data sources (*Source Services* in Figure 2). The results are then semantically integrated and presented back to the user. For text retrieval we integrated the Inquery search engine which was developed at the University of Massachusetts under the TIPSTER program. Inquery allows the indexing, searching, and rapid retrieval of text documents.

Domain users accessed the Query and Source Services through a domain-specific Query Interface, such as that shown in Figure 3. Users were able to enter simple constraints and return values to form a query, which was then applied to a number of different databases, generating specifically the information required. Users were able to access a wealth of information from a wide range of databases, *without* needing any special knowledge or training for the individual databases.

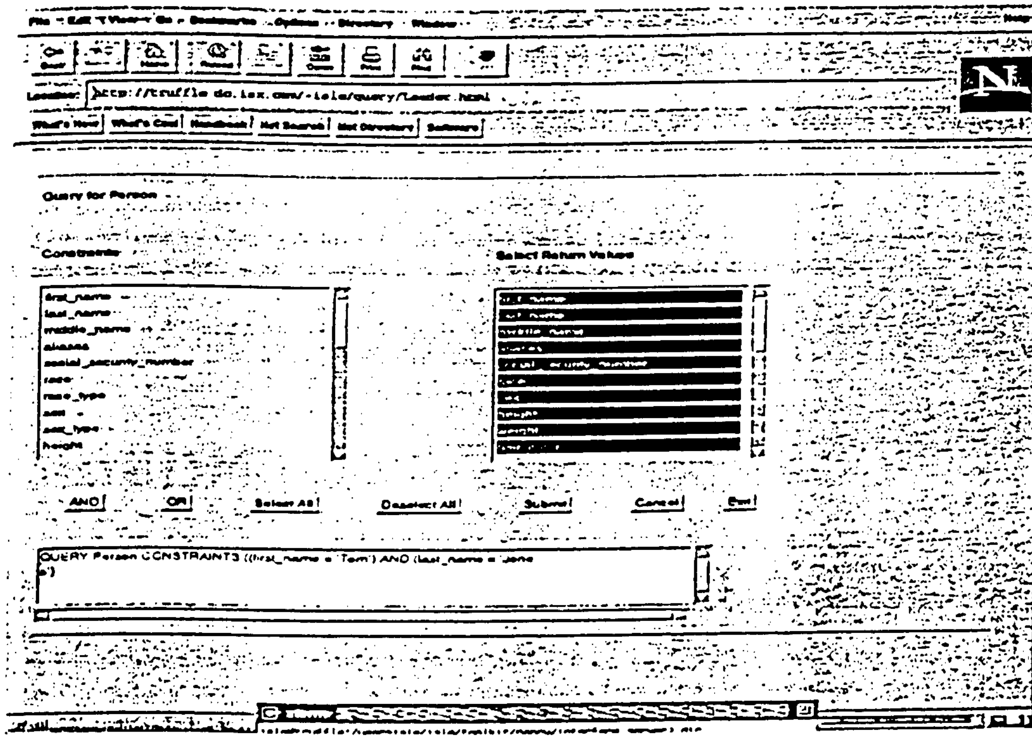
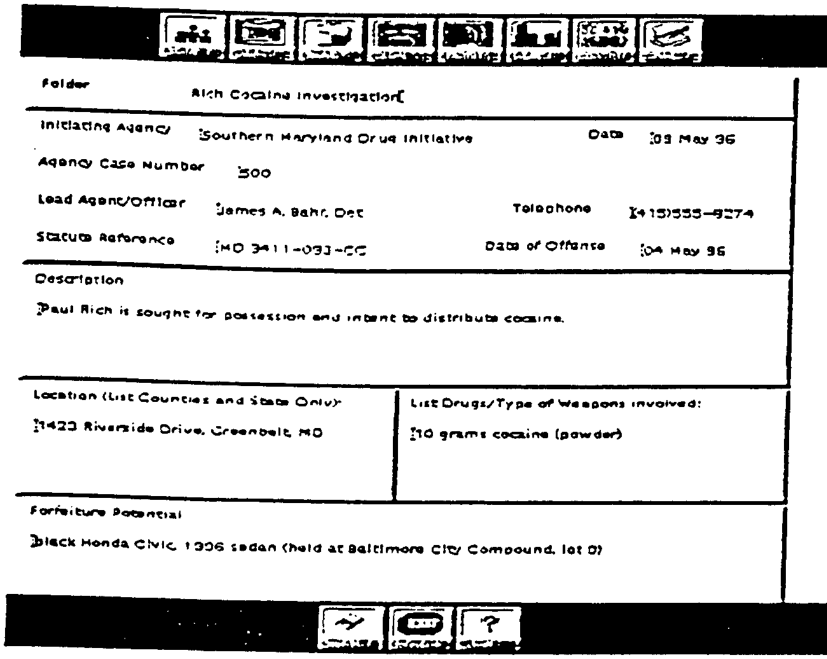


Figure 3. Query Interface Screen Example

3.3 Case Management: Second only to basic information access was the need for arranging relevant information into cases. *Folder Services* provides a domain independent method for managing, storing, and organizing such complex hierarchical containers of different media types (including text, tables, and images). Folder Services resides on top of a commercial object oriented database, in this case ObjectStore. Our domain assessment suggested that law enforcement professionals tend to organize cases around people, places, things, and events. The Case Folder Tool provides the domain-specific electronic case folder overlay onto these services.



Folder: Rich Cocaine Investigation

Initiating Agency: Southern Maryland Drug Initiative
Date: 03 May 96

Agency Case Number: 500

Lead Agent/Officer: James A. Bahr, Det
Telephone: (415)555-9274

Statute Reference: MD 3411-093-CC
Date of Offense: 04 May 96

Description

Paul Rich is sought for possession and intent to distribute cocaine.

Location (List Counties and State Only): 1423 Riverside Drive, Greenbelt, MD

List Drugs/Type of Weapons Involved: 10 grams cocaine (powder)

Forfeiture Potential

black Honda Civic, 1996 sedan (held at Baltimore City Compound, lot D)

Figure 4. Case Folder Screen Example

The case folder provides a shared repository for case related information retrieved with the query tools or entered directly by those officers having access. Officers are able to selectively insert information returned from the query results by cutting and pasteing. This means that only the information relevant to the case is included, not an entire lengthy formal report from a legacy database, as is the current practice with paper-based case folders.

3.4 Dissemination: After the relevant information is accessed, organized, and stored, it needs to be disseminated to those who need it. The *Product Services* together with the Product Package Requester application (see Figure 3) supports the generation of tailored, World Wide Web (WWW) based information products drawn from the electronic case folders. These products are based on a series of templates that query the case folders for specific information and dynamically generate specific information in the format most useful for a particular user, including plain text and HTML pages, with and without images.

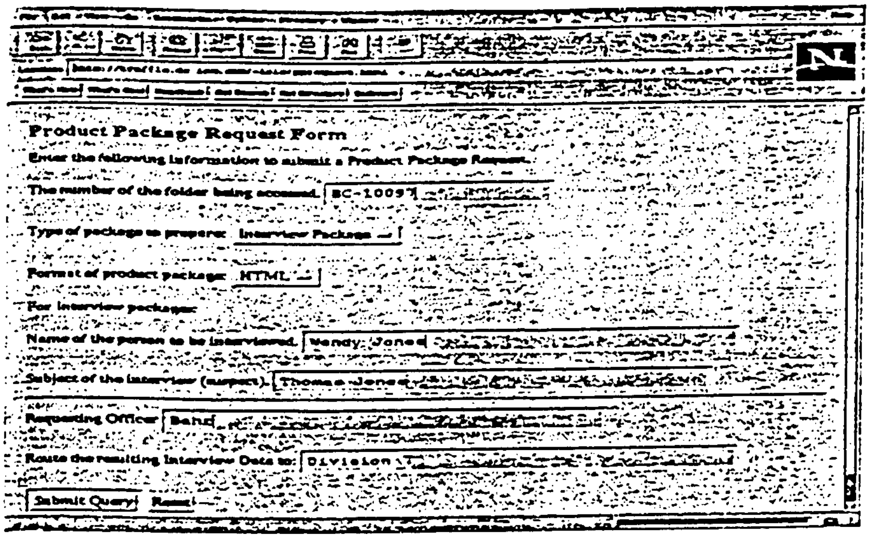


Figure 5. Product Package Request Form Example

3.5 Digital Capture: We supported the last major information management requirement through the integration of WWW forms technology with the Inquery indexer. WWW forms technology allows law enforcement officials to use standard WWW browsers from any computer (securely and wirelessly connected to the ISLE network), to enter information onto WWW versions of standard law enforcement forms (e.g., a WWW forms version of an incident response form). The Inquery indexer processed these into a format suitable for efficient search and retrieval.

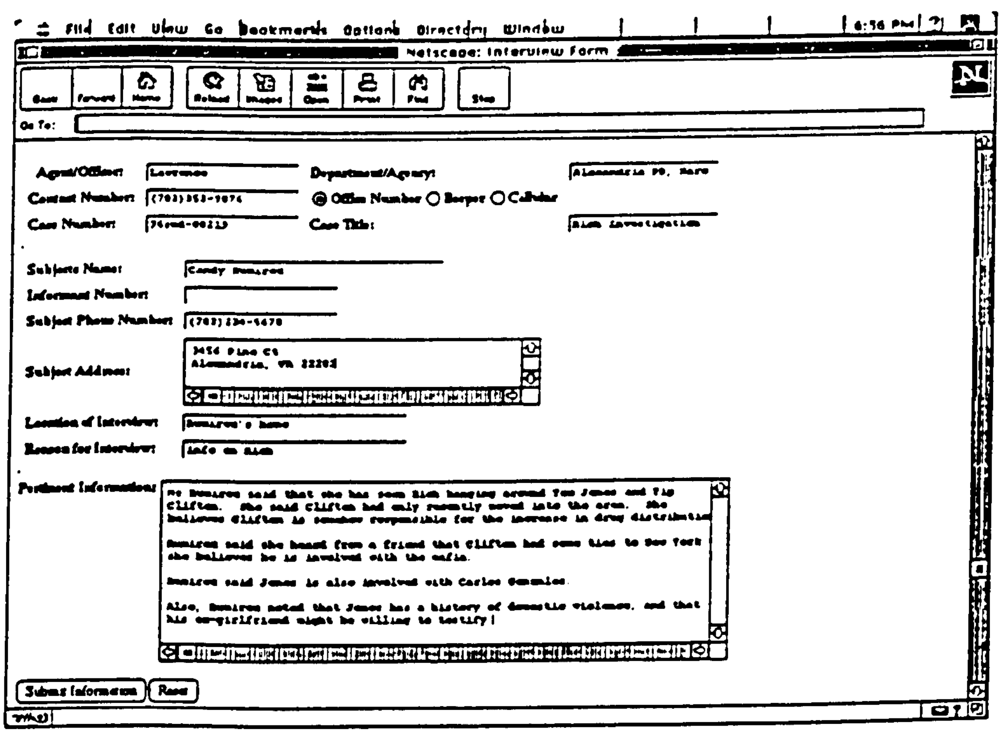


Figure 6. Electronic Submission Example Screen

Information security for both dissemination and capture was provided through the use of WWW server and browser software that incorporated the hardware-based Fortezza encryption algorithm developed by the

National Security Agency (NSA). The Fortezza encryption algorithm is expected to be approved by the Department of Justice for protecting sensitive law enforcement data. The use of the server and browser software in conjunction with the Fortezza PCMCIA cards provided us with a mostly transparent set of security services.

3.6 The ISLE System: We integrated these four primary modules into one system with a *button-bar* like interface that allowed users to activate any of these functions. All interfaces were programmed using HTML, CGI scripts, and JAVA, which made the system extremely portable across different platforms. The result was demonstration of a powerful set of tools that allowed law enforcement professionals to easily access a wide range of databases; create, manage, maintain, and exploit electronic case folders; disseminate key sets of information out of these case folders widely to other law enforcement professionals through ubiquitous WWW browser technology; and digitally capture officers' reports.

4. RESULTS AND FINDINGS

The first phase of the ISLE project culminated in late September with a demonstration of the various technologies in a law enforcement scenario at the Baltimore Police Department, and also in a military investigation scenario at the US Army's Communications and Electronics Command (CECOM), Ft Monmouth NJ. In conjunction with SRI's secure wireless communications, we demonstrated the following:

- integrated access to multiple information sources, including several relational databases, a website, a commercial PC application, and a corpus of text documents. These information sources are hosted on several different platforms including Sun Sparc workstations and PC platforms.
- information management via electronic case folders, which included the ability to cut and paste query results into a folder, execute pre-built queries to populate parts of the folder, and incorporate images with some capability to annotate these images.
- generation of tailored information packages extracted from the case folders. Available products included Folder Summary Packages, Interview Preparation Packages, Person Data Packages, and Community Crime Reports, which were available in HTML with images, HTML without images, and plain text formats.
- viewing/downloading of these information packages by remote users via secure (Fortezza-encrypted) web browsers over wired and wireless communications networks.
- electronic submission of form-based information by remote users via secure (Fortezza-encrypted) web browsers over wired and wireless communications networks.

On the whole, the law enforcement personnel at the demonstrations thought the capabilities were valuable and needed, and saw great potential should the ISLE technology be deployed in their domains. There were a number of lessons learned regarding our prototype:

- For the average law enforcement user, the simpler the interface the better.
- Foremost in the minds of all law enforcement officials is security. Security means prohibiting access from unauthorized individuals outside the agency, as well as keeping significant internal control over who can access product folders within the agency.
- Due to sensitivity and security concerns, obtaining access to real law enforcement information sources proved too difficult, so we created several dummy databases for demonstration purposes and to prove feasibility.
- Different law enforcement agencies have different requirements and policies for information security. The federal agencies tended to be more averse to (demonstrations of) technologies that seek to facilitate easier access to information.

- Different information sources have different levels of credibility in the law enforcement domain. Therefore, being able to document the source of certain information is important.
- While we successfully demonstrated the ability to transmit information securely to and from the field using Fortezza encryption, there are many issues such as key management that are yet to be resolved.
- Significant work is still required to tailor some of these technologies to operationally deployable systems; however, some of the capabilities such as electronic submission could be hardened for deployment without a major amount of effort.

5. FUTURE RESEARCH

The ISLE research effort resulted in an extensible architecture within which a wide range of relevant DARPA information technologies could be tested against, tailored to, and delivered for the law enforcement domain. Within this architecture we experimented with the application of I^3 and TIPSTER technologies to the law enforcement domain, and amassed quite a bit of operational knowledge regarding the use of these technologies in the law enforcement domain. We intend to build upon this success by improving our ability to use the architecture as a vehicle for technology transfer. Future research falls in to three main categories: hardening and deployment; extensions to current research; and new research areas.

A number of capabilities developed and demonstrated under ISLE are mature enough for deployment into real law enforcement applications. These capabilities include (among others) WWW-based electronic submission, a simplified version of WWW-based access to multiple databases, product package dissemination via the WWW, and a simplified version of the case folder. We plan over the next year to select key capabilities out of the ISLE toolkit, harden them into supportable tools, and deliver them into key operational domains at FBI and Baltimore Police Department (among others). In order to be successful, we will need to improve the ease and intuitiveness of the user interfaces as well as the ease with which these tools can be maintained, managed, and supported.

ISLE proved the applicability of I^3 and TIPSTER technologies to the law enforcement domain, but also identified areas where the two initiatives could be extended. Security implications in the law enforcement domain, especially the need to track sourcing information, provide an impressive set of challenges for the I^3 and TIPSTER research communities. In addition, there's a wide range of technologies from these two programs that could be applied to improve the ease of use, sophistication, and functionality of future ISLE systems. In particular, next year's research needs to focus on improving the ease with which additional data sources can be added; adding tools to help system administration personnel manage the domain models; integrating TIPSTER's text processing technologies more fully with the case folder and database access functionality; and creating a more robust information mediation framework for the multiple database access.

Through the process of building ISLE, we identified a number of other research areas which are potentially highly relevant to the law enforcement domain. These include advanced user interface research, to include voice and gesture recognition; more robust digital capture research, to include document recognition, pen-based computing, etc.; information survivability and security; collaboration; and visualization. ISLE's focus on just two research initiatives in the first year worked out well from the perspective of a reasonable stretch; we expect our customer to choose two new research areas for the next year to complement I^3 and TIPSTER.

From a domain perspective, ISLE started out with a focus on crisis center operations. As the project progressed, this focus shifted towards normal operations of analytical cells within organizations such as FBI, HIDTA, and the Baltimore Police Department. However, the functional model didn't shift - we were still working with a central location which pulled information in, packaged it, and then pushed the information back out. We expect to broaden the domain perspective towards multiple heterogeneous distributed law enforcement entities, a virtual community of internetworked agencies sharing and collaborating over shareable case folders.

Recommended Reading:

McGraw, K. L. and Harbison-Briggs, K. (1989). *Knowledge Acquisition Principles and Guidelines.* Englewood Cliffs, NJ: Prentice-Hall.

Rouse, W.B. (1991). *Design for Success: A Human-Centered Approach to Designing Successful Products and Systems.* New York: Wiley.