Concept Paper

For

Operational

Justice

XML

Registry/Repository

Capability

Acknowledgment Page

Many of the Concepts and content in this planning document were abstracted from (and in some cases copied verbatim) from the following two sources that proved invaluable to the Office of Justice Programs during the drafting of this document:

1. XML.gov Registry / Repository Business Case drafted by Booz Allen Hamilton in support of the Federal CIO Council's XML.gov Committee and the General Services Administration acting as agent for the CIO Council.

2. Draft Strategic Plan for the Development of a Justice Information Sharing Registry drafted by D.J. Atkinson, Telecommunications and Information Technology Division, Institute for Telecommunications Sciences, National Telecommunications Information Administration, U.S. Department of Commerce

This document was prepared by staff of the Bureau of Justice Assistance (BJA) Office of Information Technology Policy and was developed for the purpose of educating key executives concerning the need for a Justice XML Registry / Repository Capability. It should be treated as a Draft Concept Paper that does not carry the weight of an official policy document concerning this issue.

Table of Contents

List of Figures

List of Tables

Operational Justice XML Registry/Repository Capability
Concept Paper

1.0 Information Sharing: The Strategic Goal. In order to improve Public Safety for all Americans and increase our Homeland Security it is necessary to increase the opportunities for appropriate sharing of information among jurisdictions of the Justice and Public Safety community. The best way to accomplish that objective is to improve the effectiveness of Justice and Public Safety functions by applying information technologies designed for the express purpose of facilitating the exchange of information. The Extensible Markup Language (XML) is a key technology that was created for this purpose.

2.0 XML: A Key Enabler for Information Sharing. XML is generally recognized as one of the primary enablers for increasing the sharing of information and has emerged as a key technology to assist commercial and government organizations in exchanging information and conducting business over the Internet and intranets. XML is the 'glue' that promotes interoperability. It will allow the systems being developed to communicate with each other and paves the way for future or expanded collaboration between agencies and jurisdictions in the Justice and Public Safety community. XML lays the technological foundation that supports interoperability. An XML Registry / Repository, depending on its use, could support a common language and act as a catalyst in the definition of common processes across government. Some examples of common processes that are cross-agency are workflow, messaging and the definition of common functions.

The primary benefit of XML is that it is extensible and can be used to describe both low-level elements and business processes. This is also one of the primary detriments. Because it is so easy to use and can be used to 'glue' so many things together – for example, it can glue legacy system information to web-based application information – an overabundance of XML elements, schemas, or other components are being developed. As a result industry and government are already flooded with 'legacy' XML elements even though XML is relatively new. This presents a burden to the developer, IT manager, or system designer that is trying to reuse components that may be stored somewhere in multiple locations.

The first building block in defining an XML vision is to document the data that is in use in any particular community. This includes identifying the data elements, aggregations of data and semantics for those elements (sometimes referred to as Metadata). After that, common actions, activities or business processes can be developed. Upon common actions, activities or business processes, common services can be developed. The accumulation of data elements, definitions, processes, etc. into a Registry / Repository of XML Artifacts will foster reuse among developers and integrators and create a vehicle for standardized information exchange within and between agencies and jurisdictions.

Groups of public and private organizations are using XML today to define standard data components and business processes that can be built into web applications, enabling them to use the Internet in innovative ways. As communities of interest define XML elements to be used in their business transactions, other business partners seeking to implement XML solutions – either within these particular communities or for their own internal use – can access the results of these

standardization efforts. These XML definitions can be published via XML registries, which are data stores created according to industry standards to facilitate the sharing and reuse of XML Artifacts.

When information is shared electronically, it must not only be exchanged, it must also be manipulated and interpreted. XML has been a tremendous enabler for the sharing of electronic information because it provides a flexible, non-proprietary way to annotate or 'tag' information so that it can be exchanged and interpreted by disparate computer systems. However, the need within the government to use XML emerged in advance of the development of data standards. This created a situation in which programmers were free to create their own XML data structures in isolation. While awaiting the evolution of data standards and without a central storehouse and clearinghouse of XML Artifacts, it is likely that many programmers are engaged in redundant efforts to develop XML structures that may already exist in part or in total.

The Office of Justice Programs (OJP) and the Global Justice Information Network Advisory Committee (Global), with support from the U.S. Department of Commerce's Telecommunications and Information Administration (NTIA) and the Georgia Tech Research Institute (GTRI), have identified and documented a set of approximately two thousand unique, standardized data elements needed by Justice Information Systems at the state and local government level. This set of data elements is commonly referred to as the Justice XML Data Dictionary (JXDD).

Version 3.0 of the JXDD, an object-oriented version of the dictionary, was developed at GTRI in response to guidance provided by an XML Standards Task Force acting on behalf of Global and OJP. They accumulated a collection of approximate 16,000 data elements from 35 different XML sources from the Justice and Public Safety community. They removed the redundancies and duplications and resolving the semantic differences required to develop the set of 2,000 unique data elements which were then used to form a set of approximately 300 core data object types.

To facilitate the dictionary development, GTRI created a database in Microsoft Access. The Access database has a variety of search and data export capabilities that facilitate the process that GTRI is employing. In addition, use of Access provides a commonly available format for distribution of the database to the JXDD development team. When Version 3.0 of the JXDD becomes fully operational, the native format provided will be an Access database. However, if and when the JXDD becomes embedded within an Information Sharing Registry / Repository capability, a more production-oriented database may be a better choice than Microsoft Access.

The Data Dictionary is the foundation for an XML Data Model that is capable of generating XML schemas. Referenceable schema versions of the Data Dictionary have been published, and current efforts are focused on creating an object-oriented version of the dictionary. The object-oriented version has been eagerly awaited by the state and local government Justice and Public Safety community of developers. It was released for comment and feedback on April 11, 2003 and is scheduled for an initial operational capability in mid-2003.

However, there is a significant barrier to achieving widespread use of the product. It needs to be made available in a form that makes it easily accessible, understandable and available for downloading and use. In order to make effective use of the product it will be necessary to develop a

Justice XML Registry / Repository (hereafter referred to as JXRR) to house the Data Dictionary, various XML Artifacts (Schemas, Document Type Definitions (DTDs), XML Instance Documents, style sheets, etc.), data element definitions, Metadata (information about the data elements), related business processes, and various other supporting components.

A JXRR will not only provide access to the Data Element Dictionary, but also the technical tools needed to assist practitioners in developing procurement specifications and creating Justice and Public Safety Information Systems. It will assist practitioners in discovering information that could help them meet their mission objectives and aid them in establishing agreements with jurisdictions that could provide needed information.

3.0 What are the Primary Strategies for Leveraging XML?

There are four primary strategies for leveraging XML:

1. There is a need for a strategy to discover information about the availability of information (What's out there).
2. There is a need to have access to a compendium of XML Artifacts, business processes, models, and other technical tools that will enable system developers to build the structures necessary to facilitate the sharing of information.
3. There is a need for a wide range of Reusable XML Components that can be located using Registry / Repositories, downloaded in their entirety, modified for local use, subsetted for local use, and that will be accessible at run-time via operational Registries / Repositories and assigned Namespaces on the Internet.
4. There is a need for a feedback mechanism for those parties using the standards to introduce new elements / components based on local adaptation.

4.0 The Need for a Justice XML Information Sharing Registry/Repository Capability

Three main factors drive the importance of an investment in a Justice Registry / Repository capability: the business requirement to share information, the widespread use of XML as an information sharing mechanism, and the need for efficiency and effectiveness in the application of information to mission needs.

The government must share information with its stakeholders, citizens, and interested parties. Sharing information electronically is critical to government operations. The first step in enabling information sharing is to provide a place where developers and practitioners (users of information) can go to find out what already exists in the way of information that can assist them in carrying out their mission. That place is an XML Registry / Repository that contains XML Artifacts, business processes and Metadata.

An XML Registry / Repository is the foundation upon which the components to support data transmission between disparate systems is built. It also provides a common taxonomy for systems, which allows them not only to share information but also to interpret and / or translate information that is received. It is for this reason that a Justice XML Registry / Repository would be a key enabler of current and future coordination within the Justice and Public Safety community.

4.1 What is an XML Registry/Repository? An XML Registry / Repository is a clearinghouse for XML Artifacts and other related artifacts. Artifacts include Schemas, Data Elements, Metadata, XML transformations (e.g., style sheets), Data Type Definitions, XML instance documents, business processes, models, and files describing the use and modeling of the XML Schemas. The Registry / Repository naturally amplifies artifacts associated with XML Schemas through logical associations. It must be searchable and referenceable and be capable of incorporating the requirements of new users. The overarching goal of the Registry / Repository solution is to foster standardization of data definitions, data exchange business processes, and data sharing. This is accomplished by creating a directory of XML related artifacts. The ultimate purpose is to support more effective and efficient electronic business across government.

Implementation of an XML Registry / Repository consists of a distributed environment with some or all of the supporting technologies likely being housed in separate facilities from the users. It can be developed utilizing Commercial Off-the-Shelf (COTS) products, and designed and implemented following industry and government standards and best practices.

The function of the Registry is to manage external interfaces with users, applications and other registries. To this end it is responsible for managing the lifecycle functions of the Registry / Repository and for handling all queries originating externally. In essence it is responsible for shaking hands with all entities external to the Registry / Repository. Since it handles all external interfaces, it is essential that the Registry employ industry standards for the exchange of communication with users of the Registry / Repository and with other registries / repositories. The currently accepted standard for such exchange of communications is the OASIS Standard for ebXML.

The function of the Repository is to store XML Artifacts, business processes, models and other related artifacts that need to be accessible to external users via the Registry function. Communications between the Registry and Repository need not be subject to the same degree of standardization that is required for external interfaces to the Registry.

Initial implementation of a Justice XML Registry / Repository is likely to be limited to a Registry and a single Repository, which could either be integrated into a single logical entity or could be created as totally separate but linked entities. However, future evolution is likely to entail multiple Repositories associated with a Primary Registry, use of Internet Namespaces, and interfaces with other Registries / Repositories. Ultimately, the Justice XML Registry / Repository will need to be a node on a Federated Network of Registries / Repositories that is envisioned the Federal CIO Council's XML.GOV Committee, the General Services Administration and the Federal Office of Management and Budget. To this end, early implementation must take the longer view into consideration when selecting strategies for early adoption.

4.2 Value of an XML Registry/Repository. A Justice XML Registry / Repository capability would provide a consolidated approach to XML definition and storage and reuse among Justice and Public Safety jurisdictions and their partners. Moreover, by providing a means of electronic information sharing that will enable other e-Gov functions and transactions, the initiative would directly benefit customers and stakeholders, including federal, state and local entities, private

industry, and the public. A Justice XML Registry / Repository capability that supports interoperability objectives, if made available to system developers across state and local governments, would facilitate the design and implementation of current and future e-Gov initiatives.

Through use of a Registry / Repository, XML Artifacts can be leveraged within and across communities of interest that transverse organizational boundaries. The value of the Registry / Repository is fully realized when XML Artifacts are leveraged within communities of interest as well as across them. The Registry / Repository supports the process of XML Schema creation at various points of its lifecycle by providing useful pre-existing schemas and elements, as well as non-XML artifacts surrounding those schemas. Additionally, refined metadata capture increases the value and propagation of XML Schema standards government-wide, improving semantic meaning. Overall, the collective data architecture activities surrounding XML and supporting artifacts in the Registry / Repository enable improvement in efficiencies and effectiveness with respect to data exchange and the application improved coordination of data semantics between organizations.

Registries / Repositories describing and pointing to Web Services (i.e., software processes that enable fetching, adding, editing, or deleting data) offer higher value than registries only describing or defining XML entities such as XML Schemas. This is because Web Services are inherently tied to an operational business process interconnected with a functioning software system, whereas XML Schemas by themselves may exist independent of any operational system and may only provide semantic understanding and compliance aspects of data. The Registry / Repository should support the goal of government organizations to programmatically and systematically describe their Web Services and specify how they prefer to conduct business so that partner organizations can quickly, easily, and efficiently integrate with those organizations. A future implementation of a federation of registries possessing capabilities similar to UDDI Registries (Universal Description, Discovery and Integration) at a later point could simplify the effort of integrating disparate government business processes.

4.3 Who are the Stakeholders? There are five categories of stakeholders:

(1) System Architects – those who are designing systems and work processes for government clients. They will need to adjust their design approaches based on appropriate XML Artifacts, and to accommodate technical interoperability with the appropriate XML Registry / Repository.

(2) XML data standards developers (contributors of elements and data object schema components to the Data Dictionary and developers of document schemas that specify protocols for information exchanges);

(3) XML application developers (those persons who incorporate standards in computer programs that either receive or respond to requests for information, issue requests for information and then present that information to the user when it is available);

(4) Information providers (persons who are responsible for storing and maintaining bodies

of information); and

(5) Information users (those that rely on information in order to safely and effectively do their job).

4.4 Concept of Operations. XML standards developers will have a resource to assist them in finding and reusing available data components (or fragments of those components – e.g., data elements), determine when a new data component is necessary, provide guidance in the development of that component, and assemble components into document schemas. Projects and programs will come to the XML Registry / Repository during their early planning activities, in order to determine if an artifact (e.g., an XML Schema) has been developed that could be leveraged to satisfy their needs. If no artifact exists, the project would develop new XML artifacts. These artifacts will undergo a series of validity checks to ensure that they comply with minimum technical and other standards, and will then be placed in the Registry / Repository for management and future reuse by other projects and programs as appropriate.

The preliminary phase of the Registry / Repository lifecycle needs to avoid being overly ambitious in demanding common standards For XML Schemas across the Justice and Public Safety domain, while accepting that multiple XML definitions will exist. As agencies recognize parallel efforts within their own organization as well as external organizations, the XML data harmonization of a small number of XML Schemas will lead to a proliferation of high traction and highly reusable XML Schema core components. These core components will benefit future initiatives between trading partners as well as efforts with yet-to-be-identified trading partners. Community of interest leaders will then be responsible for evangelizing XML Schemas, thereby identifying and advocating the benefits and merits of competing XML definitions in order to build overall awareness.

When the standards are in place, or even while they are still being developed, XML application developers can utilize those standards to guide their product development. Information sources can register their available services (i.e., what data they can provide) and the procedures and policies related to accessing those services (sometimes referred to as a Collaboration Protocol Profile). A data user could then approach the registry, search to find who might have the data that he / she needs, determine whether he / she will be granted access to that information, and then initiate an information sharing agreement between the user jurisdiction and the data source jurisdiction (this agreement is sometimes referred to as a Collaboration Protocol Agreement).

Appendix A provides examples illustrating how the categories of users might specify and establish a vehicle history information exchange.

4.5 Architecture for an XML Registry/Repository. An ebXML oriented registry architecture has two primary functions – the registry client and the registry service. A registry client can be a web interface to the registry services or a separate application that resides and runs on a registry user's computer. The registry service has two operating sub-functions: a Query Manager (QM) that enables a registry client to find the information that he / she needs by browsing services with drill down queries or through ad hoc queries on the whole registry; and a Lifecycle Manager (LM) that provides management functions for a registry component from submission through removal. The

ebXML specification permits messaging between the registry client and the registry service to be implemented using either Simple Object Access Protocol (SOAP) or the ebXML Messaging Standard (MS) or both. Given the diversity of the Justice and Public Safety communities, it may be necessary for the Information Sharing Registry / Repository (ISR) to provide both types of messaging.

The purpose of the Query Manager is to enable registry users to discover and make use of the wealth of information contained in the registry and / or repositories.

The Justice ISR architecture will require additional components – Repositories. Figure 1 provides an illustrated model of this expanded architecture. Furthermore, these functions could be networked in a Federation of Registries / Repositories. Figure 2 provides an illustrated model of this approach which is the alternative proposed by the Federal CIO Council's XML.GOV

Sub-committee of the Architecture and Infrastructure Working Group.

Figure 1

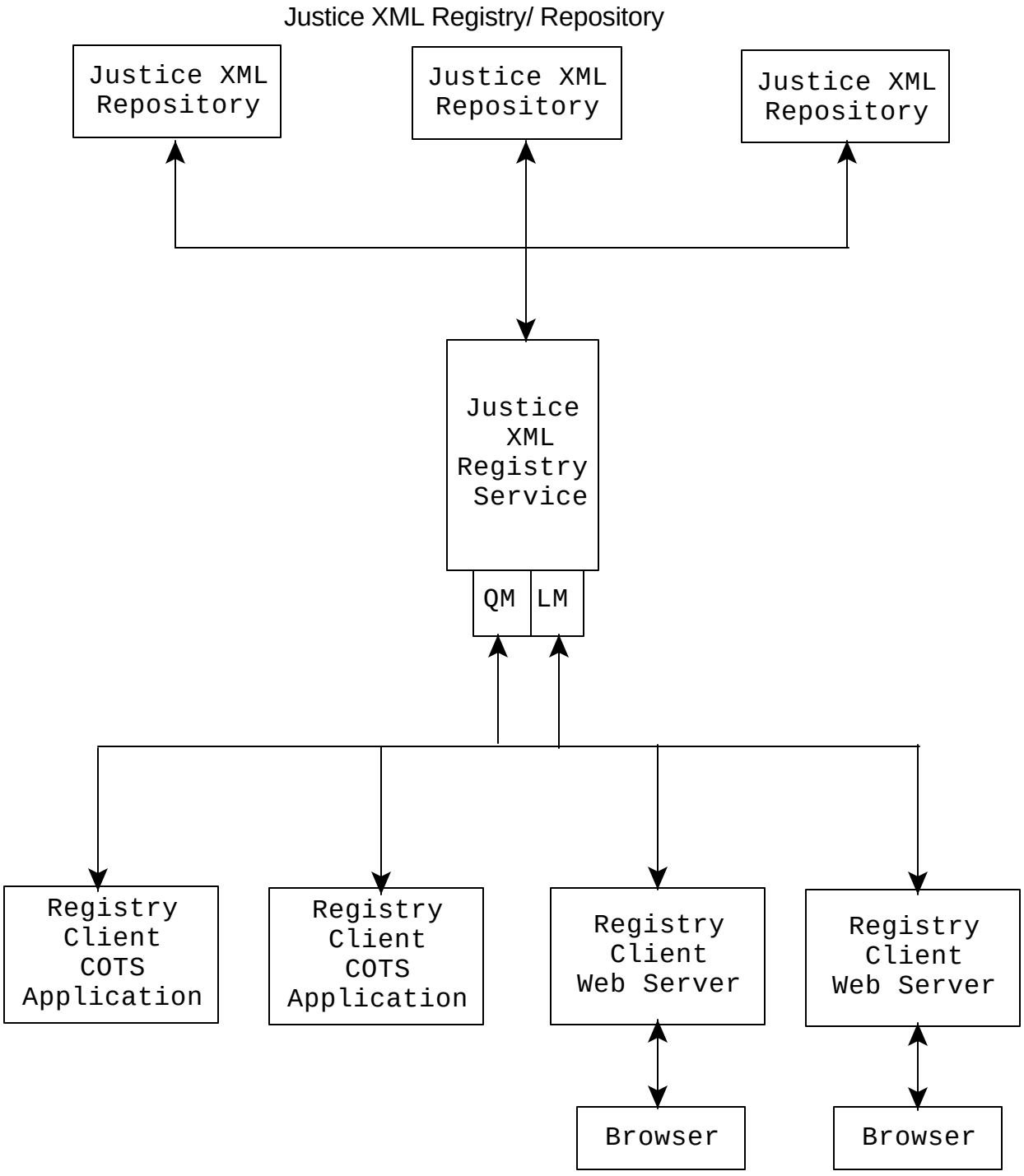
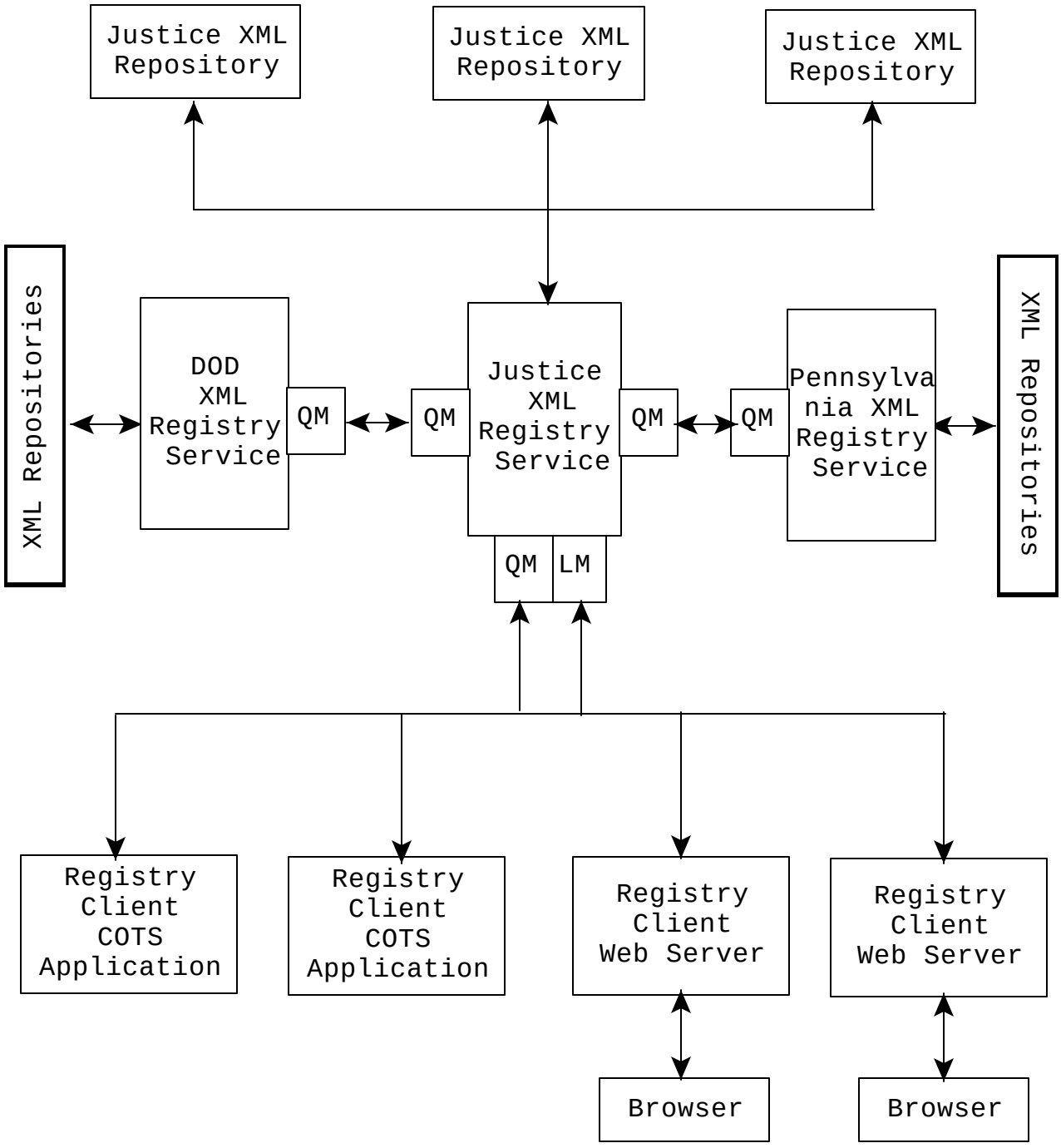


Figure 2

Federated Network of XML Registries/Repositories



As stated earlier, the Registry service has two operating sub-functions: QM and LM. The Query Manager enables registry users to discover and make us of the wealth of information contained within the registry and / or repositories. The Query services required in support of the Justice XML Registry / Repository are listed in Table 1.

Table 1. Query Services Required by the ISR

Browse by classification	Using one or more classification schemes to find the object that the user requires
Search by Tag Name	The search can be matched against a tag name or a partial tag name
Search by Keywords	Search all fields of the registry for the relevant key words
Restricted Search by Object Type	Search all fields of the requested object type for the relevant keywords
SQL Query Support	
Generate XML Dictionary Schema	Using the data object schema components and data elements in the registry, generate a referenceable XML Schema
Generate RDF Dictionary Schema	Using the data object schema components, data elements, and data object relationships in the registry, generate a referenceable RDF Schema
Assemble Document Schema from Component List	Given a list of data object components, generate a RDF or XML document schema or document schema fragment

The Lifecycle Manager provides management functions for a registry component from submission through removal. The four steps contained in the ebXML Registry Services Specification include Submission, Approval, Deprecation, and Removal. The Justice XML Registry / Repository will need to add the capability for versioning of components to the Lifecycle Manager functions.

The Justice XML Registry / Repository will need to store a variety of components. The ebXML Registry Services Specification provides a list of minimum requirements. Table 2 lists the ebXML components required. Table 3 provides a list of additional components that will be needed by the Justice XML Registry / Repository.

Table 2. Minimum Object Storage Capability as Defined by ebXML

CPA	An XML document describing a Collaboration Protocol Agreement (CPA) that indicates how two parties will communicate with each other using a specific protocol.
CPP	An XML document describing a Collaboration Protocol Profile that provides information about the rules for communicating with a particular party.
Process	An XML document describing a process
Software Component	A portion of a software package (e.g., an object library)
UML Model	A UML-based model
XML Schema	An XML Schema for a document, dictionary, or data element
Registry Package	Provides a grouping of logically related registry objects
External Link	Uses a Universal Resource Identifier (URI) to associate content in the registry with content that may reside outside the registry.
External Identifier	Provides identification information that is assigned by outside authorities (e.g., DUNS, SSN)
Association	Provides a mechanism to describe many-to-many associations among instances of objects in the registry

Classification Scheme	Provides a classification structure (i.e., taxonomy) for assigning classifications to registry objects
Classification	The assignment of a classification to a registry object by referencing a particular node within a classification scheme
Classification Node	Any node in a classification tree
Auditable Event	Provides a long-term record of events that effect a change in a registry object
User	Each auditable event is associated with the user that generated that event.
Organization	Provides a means for providing information on organizations (e.g., the organization that a user works for or that submitted an object to the registry)
Service	Provides a description of the services that are offered or required.
Service Binding	Provides the technical description of the requirements (i.e., protocols) for accessing a particular service
Specification Link	Provides a linkage between a service binding and one of its technical specifications
Unknown	An object whose content type is unspecified or unknown

In addition to these items, in order to fully function with Version 3.0 and future versions of the Justice and Public Safety XML Data Element Dictionary, some additional objects will be required. These are shown below in Table 3.

Table 3. Registry Objects that are Beyond the ebXML Minimum Specification

XML Tags	Tag Names (and their definitions) that are used in building Data Objects
XML Schema Fragments	Portions of schemas that can be inserted into other schemas
XML Data Types	Data types that can be used for creating data elements
XML Namespaces	Namespace definitions for communities of interest
XML Data Object Schema Component	Groups of one or more data elements or components that are specifically packaged for reusability and referenceability
XML Data Object Relationships	A description of how XML data objects can be related to one another
RDF Schema	As appropriate. Also includes RDF schema fragments, RDF data objects schema components, and RDF data types

4.5.1 Required Functions. The following is a list of functions required for the Justice XML Registry / Repository:

1. Ability to add, edit, and delete a hierarchical classification scheme and its related classification branches that support systematically organizing registered XML Artifacts.

2. Ability to navigate and view all content of the hierarchical classification scheme. Viewable content includes an XML Artifact, related artifacts, classification schemes to which an XML artifact belongs, and metadata about the XML artifact.

3. Ability to search the classification scheme by branches, XML Artifacts within the scheme, contents of XML Artifacts, or metadata about an XML Artifact. The search should support using human-to-system search interfaces and system-to-system interfaces.

4. Ability to add, edit, and delete an XML Artifact within one or more classification branches.

5. Ability to add, edit, and delete one or more of the metadata attributes of an existing XML Artifact within the classification scheme.

6. Ability to add, edit, and delete artifacts that supplement an XML Artifact with useful information, such as XML instance documents and files containing information about the XML Artifact.

7. Ability to review and validate XML Artifacts added to the system.

8. Ability to view, add, edit, and delete groupings of related XML Artifacts.

9. Ability to provide version control for XML Artifacts and metadata attributes.

10. Ability to establish selective security and access control capabilities for the Registry / Repository itself, classification scheme branches, XML Artifacts and their related artifacts.

11. Ability to organize Registry / Repository users within roles containing specified privileges so as to manage content and capabilities within the Registry / Repository.

12. Ability to view, add, edit, and delete actual data files in XML (e.g., agencies may publish quarterly reports in XML and centrally distribute those reports as XML files on the Registry / Repository).

13. Ability to support run-time XML validations via the HTTP protocol by comparing XML content against XML Artifacts that exist in the Registry / Repository (e.g., state governments have state water quality XML data files that must be validated against an EPA water quality XML Schema stored in the Registry / Repository)

14. Ability to provide basic XML Artifact collaboration to capture feedback and ideas about specific XML Artifacts from those leveraging the XML Artifact itself.

15. Ability for multiple Registries / Repositories to synchronize Registry metadata information from one registry to another in order to facilitate interoperability.

4.5.2 Use Cases/Scenarios. Based on the functional requirements, the Justice Registry / Repository has a variety of use cases and user scenarios that convey its potential uses. The following high-level use cases and associated descriptions specify anticipated possibilities for how users and systems would interact with the Registry / Repository:

1. A user uploads a new XML Schema. The user logs in, identifies a relevant classification scheme, uploads the XML schema file, has the XML schema automatically validated for correctness by the registry, receives upload confirmation, and waits for em-mail review and acceptance by the community of interest.

2. A user performs a text search for an XML Schema. The user navigates to the registry, types in search criteria, receives a list of XML Schemas with schema text matching search criteria, selects a desired XML Schema from the list, then views the detailed XML Schema contents and metadata to

determine applicability to their technology initiative.

3. A user performs a metadata search for an XML Schema. The user navigates to the registry, selects an option for a metadata search, views a list of standard metadata fields for searching, enters metadata search criteria, then views resulting XML Schema in the Registry / Repository.

4. A system different from the Primary Registry / Repository references an XML Schema in the Primary Registry / Repository. System A uses a SOAP-based message over HTTP to consume a Web service on System B. System B responds that the Web service requires XML compliance to an XML Schema located in System C's Registry / Repository. When receiving the SOAP-based message from System A, System B consumes the XML Schema from System C using HTTP, validates System A's message, and responds with XML data using the Web service.

5. A developer uses an XML software development tool to validate an XML instance file against an XML Schema residing in the Registry / Repository. The developer loads an XML instance file into the development tool, adds the registry's URL for the referenced XML Schema within the XML instance file, selects an option in the development tool to validate the XML file, the development tool references the XML Schema in the Registry / Repository, then the tool identifies the XML instance file as either valid or invalid.

6. A community of interest administrator approves an XML Schema submitted by a user. A community of interest administrator logs into the registry, selects an op[tion to view submitted, unapproved XMLO Schemas, views a list of submitted XML Schemas, selects a schema from the list, views the schema details and related metadata, and approves the XML Schema Assuming IT correlates into the community of interest classification scheme. The registry then automatically sends an e-mail message to the submitter with the notice of approval.

4.6 Governance. Architecture configuration management is an imperative aspect of governing the Justice XML Registry / Repository and the entire XML landscape. It is assumed that organizations will inevitably need to innovate and extend beyond the baseline functionality based on ebXML specifications. Evaluation and analysis of capability necessity must be employed on each enhancement to core ebXML capabilities. A Change Control Board (CCB) is recommended to manage future plans for processes, prioritization of objectives, and high-level scope activities. Essentially, the CCB works with core leadership in the Registry / Repository community to ensure that any non-ebXML, custom Registry / Repository requirements can be synchronized across all organizations affected by it to minimize stovepiped efforts.

There are at least three primary roles that have been identified associated with the maintenance of an XML Registry / Repository. In the case of a Federated Network of Registries / Repositories, roles at the agency and community of interest level could overlap with those needed for the support of a local agency XML registry, if they exist.

The first role is the need for a Registration Authority or Executive Agent whose primary responsibilities include coordinating and implementing registry functionality, insuring interoperability with other registry / repositories, monitoring national XML registry standards, maintaining relationships and ensuring coordination and convergence with other government

bodies, determining policy and procedures regarding submissions to the registry, coordinating XML education and outreach, determining levels of conformance, defining legal liabilities, reviewing and approving new entries to the registry, reviewing and approving new agencies' connection to and use of the registry, and functioning as a Program Manager for those agencies who cannot support a local registry but wish to use the Primary Registry.

The second role is the Program Manager also known as the Responsible Organization. The Program Manager is the person or group at each agency who maintains the local agency XML Registry, if one exists, and provides the link between the local and Primary Registries. Maintenance of submissions to the Primary Registry would be an extension of their current functions of providing oversight for their local registry. Responsibilities of the Program Manager include maintaining the local registry, making submissions to the Primary Registry if there is one, coordinating efforts between the local registry and the Primary Registry (if applicable), searching the Primary Registry for objects for reuse, registering new XML Artifacts in the Primary Registry if necessary artifacts are not found in the Primary Registry and linking new artifacts to the local registry, managing artifact lifecycle, maintaining artifact stability, managing version control, managing audit trails and maintaining quality control.

The third role is Community Manager, also know as the Submitting Organization. The community managers are the primary end-users of the registry. This group might include developers, data modelers, integrators, database administrators, business process managers, etc. This group could also include people from related communities of interest, which may have representation from private industry, standards bodies, working groups, subcommittees, etc. Responsibilities of the Community Managers include representing communities of interest, representing local working groups, reviewing submissions for conflict with entries in the Primary Registry, and verifying appropriateness of proposed submissions.

5.0 Concept for a Federated Network of XML Registries/Repositories. The Vision of the Federal CIO Council's XML.GOV Committee is for a Federation of XML Registries / Repositories linked together in a network that interfaces the Registries. Such a Federated Network would help promote a variety of federal, state, local and tribal government initiatives and missions, including, but not limited to, e-Gov and Homeland Security initiatives. It is an important concept to understand that the word "Federation" does not imply Federal. It is very likely that many of the Registries / Repositories that eventually are linked together in the proposed Federated Network will be operated by state and local governments (possibly even Tribal Governments at some point). The preferred alternative of a "registry of registries operated in a federated model, linked by a primary (or flagship) registry" would provide a non-intrusive way for existing and planned registries to co-exist and be used by all agencies without significant additional effort and resource allocations on their part.

Such a Federated Network of Registries / Repositories would be operated in a federated fashion in order to balance specific agency needs and the requirements of standardization, interoperability, and reuse. The federated model consists of a Primary Registry, presumably operated by XML.GOV, and a government-wide coordination function that links individual agency registries and ensures their interoperability with one another and the Primary Registry. The Primary Registry would also serve as the Registry / Repository for agencies without individual registries.

Any individual registry / repository within the federated registry / repository model architecture would be able to negotiate and exchange information with other registries within the federation. Additionally, a federated registry model could facilitate cross-searches of other registry indexes for relevant information and provide pointers to relevant XML Artifacts for either human or system consumption. As the propagation of new XML registries will inevitably continue within the government at all levels, registry synchronization and interoperability goals become increasingly imperative. Enterprise reuse and establishing a distributable technology architecture that different agencies can leverage is critical within the federated registry / repository model.

During the initial lifecycle deployment of the federated registry / repository model – preliminary ramp-up and standardization coalescence – existing registry efforts would plan to align their registry architectures with the Primary XML Registry / Repository architecture. After aligning technology plans and moving forward with architectural changes in the federation, registry / repository synchronization can then be implemented as specifications are released that provide guidance on enabling synchronization between registries. The federated registry / repository architecture model reasonably assumes that within two years from the initiation of the registry / repository effort, specifications will exist from either ebXML or another standards body that provide guidance in synchronizing registries. Various stakeholders with the ebXML community have noted that such synchronization is already identified as a key future specification intended for availability in upcoming ebXML specification documents. In the event that ebXML fails to put forth appropriate synchronization specifications, the federated registry / repository model will support a workable concept of synchronizing registries through the use of peer Web services existing on each registry / repository in the federation.

An early approach to a Federated Model would enable the government to immediately proceed with semantic standardization efforts before approaching the complexities of registry / repository standardization and interoperability issues. In summary, the complete federated model establishes a distributed XML knowledge management system for creating government-wide efficiencies in leveraging XML Artifacts and more effectively incorporating XML into electronic business processes.

6.0 Applicable Standards. The need for standards can be grouped into the need for Metadata standards and the need for operational standards. Metadata standards include such items as ISO 11179 and Dublin Core. Currently the focus for operational standards seems to be on ebXML as the interfacing standard between registries and users and / or user application software.

The concept of industry-embraced Web services should not be thought of as a W3C owned and controlled set of specifications. OASIS is actively working on many critical specifications surrounding non-ebXML Web services including WS-Security, UDDI, and other specifications beyond the work of the W3C. Thus, it is not necessarily valid to say that the W3C standards are the only foundation “XML Family of Standards”. Standards from various organizations such as OASIS and OMG must be examined in tandem with the W3C standards when making forward thinking decisions for government architectural initiatives. This concept further lends credibility to OASIS sponsored ebXML specifications.

As open market Web services are much less rigid, ebXML provides what many consider to be higher-end and semantically rich Messaging / Web Services specifications. Additionally, ebXML provides more robust ties between ebXML compliant artifact registries and ebXML-based Web services. Nevertheless, an ebXML Web services implementation requires more significant interactions between trading partners.

Both ebXML and Web services initiative are characterized as enabling systems in different organizations with distinct sizes and locations to discover one another, conduct automated data integration processes with one another, and subsequently do more business with one another in the future. The ideal integration scenario involves companies being able to negotiate business transactions easily and efficiently with minimal non-automated interaction. ebXML Web services primarily differentiate themselves from non-ebXML Web services by channeling integraged Web service into pre-defined and highly structured business process specifications by which trading partners abide. This level of semantic integration helps minimize invalid uses of Web services messaging, whereas non-ebXML Web services allow an organization to interact with trading partners using limited semantic understanding and highly flexible downstream integration options.

No formal relationship currently exists between UDDI and ebXML specifications. However, UDDI and ebXML exist as complementary technologies such that organizations using ebXML Web services can register their business and Web services in a UDDI registry. Subsequently, UDDI Web services, service types, and specification pointers have the ability to point to an ebXML registry / repository for business and technical descriptions of the services. Moreover, organizations that register XML Artifacts such as XML Schemas or XML Style Sheets in an ebXML compliant registry / repository can also register the existence of these artifacts via a URL as “service types” in a UDDI registry. This enables UDDI Model specification entries to point to these remote XML Artifacts that actually reside in the ebXML Registry / Repository.

Standardizing XML data definitions for data exchange efforts is easily feasible. However, standardizing XML business processes is highly complex and not as practicable given the implicit disparities that exist between government organizations using Commercial Off-the-Shelf (COTS) and custom software applications. The standardization of XML business processes across the government will require a significantly greater level of effort than the standardization of XML data definitions.

7.0 Other Related Issues. The following are a few key issues that need to be understood by system developers and users.

7.1 Namespace Policy Issues. One of the goals of the Justice XML Registry / Repository is to disambiguate elements using useful, logical, and relevant XML Namespaces that are closely mapped to functional communities of interest. By creating and reusing components that are stored on various Registry / Repositories throughout Government, developers can enjoy economies of scale that will reduce cycle times, improve the quality of software and provide solutions faster, cheaper and more reliably. This in turn will provide interoperability benefits that will further enhance performance.

7.2 Interoperability Issues. Overall, well-defined interoperability facilitates system collaboration between a heterogeneous set of platforms, applications, and programming languages by transparently supporting XML data negotiations between systems. ebXML specifications state that the primary reason for conforming to ebXML is to increase the probability of the interoperability between XML registry implementations being successful. Ideally, registry / repository specifications from standards bodies must be integrated to ensure traction and the ability for software product vendors to develop consistent software models that map to clearly understood specifications with minimal to nonexistent overlap. Registry interoperability must be dealt with at a level beyond specification-by-specification software patches. The government needs to expect that ebXML interoperability standards will show longer term, clearly defined vision. Without such well-defined interoperability even the most sophisticated ebXML registry / repository specification will inevitably lack market traction.

7.3 Need for a Security Plan. As part of its security plan for the proposed business transformation, the Justice XML Registry / Repository will:

1.Identify threats that could adversely affect critical operations and assets, such as intruders, criminals, disgruntled employees, terrorists, and natural disasters.
2.Estimate the likelihood of such threat based on historical information and the judgment of knowledgeable individuals.
3.Identify and rank the value, sensitivity, and criticality of the operations, data and assets associated with the federation of registries.
4.Estimate the potential costs associated with the impact on the most critical and sensitive assets and operations if a threat should materialize, including recovery costs.
5.Identify cost-effective actions to mitigate or reduce risk, which might include new organizational policies and procedures, as well as technical or physical controls.
6. Document the results of the risk assessment and draft an action and risk management plan.

7.4 Need for Long-Term Funding for the Operational Capability. It is essential to begin planning and budgeting for the funding necessary to support the long-term operational use of a Justice XML Registry / Repository capability. Although significant progress can be made through Office of Justice Program grants to supporting organizations such as the Georgia Tech Research Institute in order to pilot test the use of a Registry / Repository capability, it will take significantly greater resources to operationalize the capability over the long term. Action is needed to begin the process of identifying and justifying the needed funds.

Appendix A

Case Example: Establishing a Vehicle History Information Exchange

Schema Documents are Developed. An XML standards developer, in this case AAMVA and NLETS, recognize the need to exchange information on vehicle histories using XML. The developers search the Information Sharing Registry / Repository (ISR) for an existing vehicle history schema and they find that no existing schemas meet their needs.

When the developers decide they need to develop a new document schema, they search the ISR for components applicable to a vehicle history record. The developers use the tools of the registry to assemble available data object schema components and elements into a document schema. If the existing components and elements are insufficient, the registry assists them in creating and submitting additional components and incorporating those into the document schema. Once the developers are satisfied that the document schema meets the information sharing needs of their constituency, they can finalize it and have it automatically submitted to the registry.

Business Process Documents are Submitted. The standards developer, again AAMVA and NLETS, develops and submits the business process descriptions regarding how driver history information is to be shared. These documents contain the rules about who can access the data and under what conditions the information is provided to the user.

Applications are Developed. Possibly together, but usually separately, the information user's agency and the information provider's agency will develop software applications. In the case of the information provider, this may be as simple as providing a new interface to an existing database or as complex as a full system replacement, depending on their current application lifecycle and their customers' needs.

In the case of the information user's agency, the application may be as simple as a web-based form accessed from a laptop or desktop computer, or as complicated as a custom application, tailored for the agency's needs, that runs on a wireless palmtop computer.

In both cases, however, the registry provides a vital resource. It allows the agencies to investigate available XML standards and document schemas as well as helping to provide a referenceable specification that all vendors have equal access to. This includes the business process documents described previously.

As XML application development becomes more mature, the applications can be developed to be registry-aware. That is, the applications will be able to perform many of the functions described in the remainder of this section in an automated manner that still maintains the integrity and security of the information provider's data while allowing the information user almost instant access to the data that he / she needs to perform their job efficiently.

Information Provider's Collaboration Protocol Profile is Submitted. The information provider (e.g., a state Department of Motor Vehicles, DMV) publishes its Collaboration Protocol Profile (CPP)

to the Registry. The CPP describes the information provider, its role, its offered services, its information access policies, and the technical details on how to access those services. The Information Provider uses the Registry's classification facilities to categorize its CPP so that information users can discover the information provider using the searching functions of the registry.

Information User Discovers the Provider. The information user (e.g., a police investigator) browses the registry using classification schemes and other search functions defined within the registry to discover an appropriate information provider. For example, the information user may want to search for a particular vehicle's history in all states in New England. The user discovers the provider's CPP and decides whether he/she wants to request information from that provider.

Provider and User Establish a Collaboration Protocol Agreement. The user can, based on his own and the provider's CPP, create a Collaboration Protocol Agreement (CPA) which is then submitted to the provider. Once the provider accepts the CPA, the parties can begin to exchange information.

The Information User's Perspective. The previous sections provide a lot of information that an information user really doesn't care about. This sections utilizes our vehicle history example to describe the information user's perspective of using registry-aware applications where the information providers allow automatic CPA negotiation.

1.As a Boston Police officer pulls over a car for speeding, he enters the New Hampshire license plate number into his palmtop computer.
2.As he prepares for his encounter with the driver, the software in the palmtop computer contacts the registry to determine which data provider has information about New Hampshire license plates. Using the Officer's identification information, the software automatically establishes a 30-minute CPA with the New Hampshire DMV and requests the information on the license plate.
3.As the officer approaches the automobile, his palmtop receives the information from the New Hampshire DMV. While the NH report contains no information that there are any problems with the plates, the officer notes that the make and model of the car do not match those listed by the DMV.
4.The officer retrieves license, registration, and insurance information from the driver. In addition, he records the VIN listed on the dash plate of the vehicle.
5.The officer returns to his patrol car where he enters the vehicle VIN into his palmtop computer.
6.As the officer begins entering information required to issue a citation into his palmtop, the search application contacts the NH DMV again to search for information by VIN. When NH returns a "No Information" response the application contacts the registry again to determine what other data providers might have information on the VIN in question. The application negotiates new CPAs with the DMVs in the states that border NH and initiates a VIN query with each of them.
7.As the officer enters information for the citation, the search returns indicating that the vehicle with that VIN was stolen in Vermont the day before. The officer then follows department procedures to arrest the driver and have the car impounded.

This scenario, while perhaps a bit simplistic, demonstrates the potential for automated establishment of information sharing agreements between agencies that can be facilitated by a true

information sharing registry.

Appendix B

Glossary

DTD Document Type Definition

Interoperability Interoperability is the ability of two or more systems or components to exchange information and to use the information that has been exchanged (GAO definition). This definition implies both technical interoperability and semantic interoperability. Technical interoperability means that the architecture must work with lower level technical details such as protocols (e.g., HTTP), languages (e.g., XML), and potentially even traverse systems such as firewalls that may explicitly prohibit interoperability by limiting data exchange. IN contrast, semantic interoperability implies that the architecture supports systems' use of the exchange information based on shared comprehension and agreement on the meaning of that data. Two systems may be able to technically interoperate with one another successful, but one system may have a significantly divergent interpretation of the data being exchanged if the semantic meaning of that data was not understood or delineated within the data exchange process.

Metadata Descriptive information about other data. When applied to an XML context, metadata typically conveys information relating to XML structure, context, and use so as to help further describe an XML Artifact.

ebXML Electronic business XML is ajoint consortium between OASIS and UN / CEFACT. OASIS typically has a commercial focus as it sets objectives per the direction of its predominantly industry-based members, whereas UN / CEFACT acquires direction from bodies represented in international governments as found in the United Nations. Relevant ebXML specifications for the Justice XML Registry / Repository initiative provide business process collaboration, exchange of XML business semantic knowledge (i.e. XML Registries), messaging protocols, and trading profile management and discovery.

Overall, ebXML specifies methodologies and specifications for describing business processes and organizing XML Artifacts in a registry to facilitate their use with business processes. ebXML Registry / Repository products specify a semantic registry. A semantic registry primarily stores meaningful XML-based semantic content (e.g., metadata information about elements), as well as classifications and associations of that semantic content. ebXML provides specifications for ebXML messaging service capability, establishing a more sophisticated specification for XML-based Web Services that aims to solve a similar problem as competing Web Service specifications. Although no formal ebXML product certification is

currently available, product companies are able to build products based on the interpretation of specifications. Moreover, NIST is in the process of establishing an ebXML compliance and conformance standard to verify validation points of ebXML compliant registry software.

Registry/Repository Refers to electronic listings of specifications (schemas, DTDs, etc.) as well as the locations where the schemas, DTDs, metadata, etc. reside. A mechanism used to discover and retrieve documents, templates, and software (i.e. components and resources) over the Internet or intranet. A registry is the mechanism to discover the component(s). The registry provides information about the component (metadata), including its location. A repository is where the component resides. A user retrieves a component from a repository.

Related Artifacts Refers to any supplementary documents, XML Files, design documents, or any other files that support XML Artifacts themselves and convey more precise information about those particular XML Artifacts.

SOAP Simple Object Access Protocol. A means for a program running on one kind of software platform to communicate with a program on the same or different kind of platform using standards based on HTTP and XML as the mechanisms for information exchange. SOAP specifies how to encode an HTTP header and an XML file it so that a program on one system can call a program on another system and pass it information. SOAP also specifies how the called program can return a response, thus making it an overall data exchange governance standard.

UDDI Universal Description, Discovery, and Integration. An OASIS specification that defines an XML-based registry for organizations to list and describe their Web services on a network. UDDI enables organizations to find beneficial software capabilities on the Web and interoperate with those systems for data exchange and commerce. UDDI is often compared to a telephone book's white, yellow, and blue pages; it allows businesses to be listed by name, product, location, or the Web services they offer. UDDI is intended to operate as a general registry for Web services, as well as XML Artifacts and schemas specially designed with those services.

Web Services Self-describing, self-contained, modular units of software application logic that provide deined business functionality. Web services are consumable software services typically including some combination of programming and data. They may also include human resources made available from an organization's UDDI server for Web users or other web connected software. Web services can be aggregated with one another to establish a larger workflow or business transaction. Inherently, the architectural components of Web services support messaging, service descriptions, registries, and loosely coupled interoperability.

WSDL An XML-based language used to describe the services offered by an organization and provide a way for individuals and organizations to access and search those services electronically. WSDL language describes services listed on an UDDI directory and is, therefore, a cornerstone of the UDDI initiative.

XML Extensible Markup Language. XML is a tagging language used to describe and annotate data so it can be consumed by human and system interactions. XML is typically arranged hierarchically using XML elements and attributes. It also uses semantically rich labels to describe elements and attributes to enable meaningful comprehension.

XML Artifacts Refers to and includes XML Schemas, Document Type Definitions, XML Instance Documents, XML Transformations, etc.

XML Schema An XML Schema Definion (XSD) is a W3C specification that defines how to formally describe XML elements within an XML document. XSD descriptions enable a system to verify that all XML data adheres to the element description rules within the XML document. An XML Schema represents an abstract characterization of the XML elements that indicates an object's relationship to other objects in the XML document. A system considers an XML document valid when it conforms to the specifications in an XML Schema.