

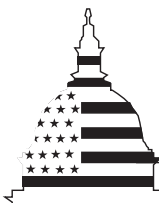
GAO

Report to the Chairman, Committee on
Governmental Affairs, U.S. Senate

April 2002

ELECTRONIC GOVERNMENT

Challenges to Effective Adoption of the Extensible Markup Language



G A O

Accountability * Integrity * Reliability

Report Documentation Page

Report Date 00APR2002	Report Type N/A	Dates Covered (from... to) -
Title and Subtitle ELECTRONIC GOVERNMENT: Challenges to Effective Adoption of the Extensible Markup Language	Contract Number	
	Grant Number	
	Program Element Number	
Author(s)	Project Number	
	Task Number	
	Work Unit Number	
Performing Organization Name(s) and Address(es) General Accounting Office, PO Box 37050, Washington, DC 20013	Performing Organization Report Number GAO-02-327	
Sponsoring/Monitoring Agency Name(s) and Address(es)	Sponsor/Monitor's Acronym(s)	
	Sponsor/Monitor's Report Number(s)	
Distribution/Availability Statement Approved for public release, distribution unlimited		
Supplementary Notes		
Abstract This report responds to your request that we review the status of Extensible Markup Language (XML) technology and the challenges the federal government faces in implementing it. XML is a flexible, nonproprietary set of standards designed to facilitate the exchange of information among disparate computer systems, using the Internets protocols. Specifically, we agreed to assess (1) the overall development status of XML standards to determine whether they are ready for governmentwide use and (2) challenges faced by the federal government in optimizing its adoption		
Subject Terms		
Report Classification unclassified	Classification of this page unclassified	
Classification of Abstract unclassified	Limitation of Abstract SAR	
Number of Pages 73		

Contents

Letter		1
---------------	--	----------

Executive Summary		2
	Purpose	2
	Background	2
	Results in Brief	4
	Principal Findings	6
	Recommendations for Executive Action	9
	Agency Comments and Our Evaluation	10

Chapter 1	Background: Features and Current Federal Use of XML	11
	Standardized Data Tagging Facilitates Information Exchange among Disparate Systems	13
	XML Supports Internet-Based Data Exchange	16
	XML's Technical Standards Provide the Tools to Describe and Exchange Data over the Internet	18
	XML Was Designed to Accommodate Numerous Extensions	19
	XML Can Enhance Information Search, Retrieval, and Analysis	20
	XML Usage Complements Traditional Electronic Data Interchange Applications	22
	Federal XML Projects Vary in Size and Scope	26
	Objectives, Scope, and Methodology	31

Chapter 2	A Comprehensive Set of Standards for Implementing XML Is Only Partially in Place	33
	XML Technical Standards Have Largely Been Defined	33
	Additional Standards Have Been Proposed for Using XML to Conduct Electronic Business	35
	Business Process Standards Are Less Well-Developed than Technical Standards	36
	Potentially Useful XML Vocabularies Are Not Ready for Governmentwide Adoption	41

Chapter 3	The Federal Government Faces Challenges in Realizing XML's Full Potential	44
	Implementing XML Presents Pitfalls	44
	Governmentwide Actions to Promote XML Adoption Have Focused on Education and Outreach	47

	Federal Government Needs Have Not Been Consolidated for Input to Standards-Setting Bodies	50
	XML Interoperability across the Government Depends on an Effective Cross-Agency Registry	52
	XML Implementations Can Be More Effective within the Context of an Enterprise Architecture	56
Chapter 4	Conclusions and Recommendations	58
	Conclusions	58
	Recommendations for Executive Action	60
	Agency Comments and Our Evaluation	61
Appendix I	Comments from the National Aeronautics and Space Administration	62
Appendix II	Comments from the National Archives and Records Administration	64
Glossary		65
Tables		
	Table 1: Comparison of HTML and XML	17
	Table 2: Basic XML Components	18
	Table 3: Comparison of EDI and XML	26
	Table 4: XML Technical Standards as of February 2002	34
	Table 5: Representative Industry-Specific XML Vocabularies	36
	Table 6: Strengths and Pitfalls of XML	46
Figures		
	Figure 1: A Hypothetical XML-Based State Driver's License System	15
	Figure 2: XML Code Example	16
	Figure 3: XML Can Facilitate the Use of Different User Interfaces and Display Devices	22

Figure 4: A “Request for Quotation” Formatted as an EDI Message	23
Figure 5: Typical Flow of Business Transactions Based on EDI Standards	25
Figure 6: Representative ebXML Transaction	38
Figure 7: Using a Registry of XML Data Elements and Structures	54

Abbreviations

ANSI	American National Standards Institute
CIO	chief information officer
DISA	Defense Information Systems Agency
DOD	Department of Defense
DTD	document type definition
ebXML	electronic business XML
EDGAR	Electronic Data Gathering, Analysis, and Retrieval
EDI	Electronic Data Interchange
EPA	Environmental Protection Agency
GAO	General Accounting Office
GSA	General Services Administration
HTML	Hypertext Markup Language
IT	information technology
NIST	National Institute of Standards and Technology
OASIS	Organization for the Advancement of Structured Information Standards
OMB	Office of Management and Budget
SEC	Securities and Exchange Commission
UDDI	Universal Description, Discovery, and Integration



United States General Accounting Office
Washington, DC 20548

April 5, 2002

The Honorable Joseph I. Lieberman
Chairman, Committee on Governmental Affairs
United States Senate

Dear Mr. Chairman:

This report responds to your request that we review the status of Extensible Markup Language (XML) technology and the challenges the federal government faces in implementing it. XML is a flexible, nonproprietary set of standards designed to facilitate the exchange of information among disparate computer systems, using the Internet's protocols. Specifically, we agreed to assess (1) the overall development status of XML standards to determine whether they are ready for governmentwide use and (2) challenges faced by the federal government in optimizing its adoption of XML technology to promote broad information sharing and systems interoperability. The report recommends that the director of the Office of Management and Budget (OMB) take steps to improve the federal government's planning for adoption of XML.

As agreed with your office, unless you publicly announce the contents of this report earlier, we plan no further distribution until 30 days from the report date. At that time, we will send copies of this report to the ranking minority member, Committee on Governmental Affairs, and interested congressional committees. We will also send copies to the director of OMB. Copies will be made available to others upon request. The report will also be available on our home page <http://www.gao.gov>.

If you have any questions concerning this report, please call me at (202) 512-6257 or send e-mail to mcclured@gao.gov. Other major contributors included Barbara S. Collier, John de Ferrari, Chetna Lal, Steven Law, Anh Le, John C. Martin, and Mark D. Shaw.

Sincerely yours,

David L. McClure
Director, Information Technology Management Issues

Executive Summary

Purpose

The Extensible Markup Language (XML) is a flexible, nonproprietary set of standards for annotating or “tagging” information so that it can be transmitted over a network such as the Internet and readily interpreted by disparate computer systems.¹ It is increasingly being promoted by information technology (IT) developers as the basis for making computerized data much more broadly accessible and usable than has previously been possible. As a result, many organizations, including both private businesses and federal government agencies, are building applications that try to take advantage of XML’s unique features. Given the widespread interest in adopting this new technology, the chairman of the Senate Committee on Governmental Affairs asked GAO to assess (1) the overall development status of XML standards to determine whether they are ready for governmentwide use and (2) challenges faced by the federal government in optimizing its adoption of XML technology to promote broad information sharing and systems interoperability.²

Background

Advances in the use of IT—especially the rise of the Internet—are changing the way private sector businesses, government agencies, and other organizations communicate, exchange information, and conduct business among themselves and with the public. The Internet offers the opportunity for a much broader and more immediate exchange of information than was previously possible, because it provides a virtually universal communications link to a multitude of disparate systems. However, although the Internet can facilitate the exchange of information, much of the information displayed to users is delivered only as a stream of computer code to be visually displayed by Web browsers, such as Internet Explorer or Netscape Communicator. For example, an economist might visit a Web page that displayed statistical information about the production of various agricultural commodities over a number of years. Typically, such a Web page would only display this information to the economist to examine visually on his or her computer screen. Without special translation software, it would likely be difficult for the economist

¹ Tagging is accomplished by labeling each element of a data set to clarify what kind of information is being provided. For example, “1600 Pennsylvania Avenue” could be tagged to show that it refers to an address. In XML, the result would be <Address>1600 Pennsylvania Avenue</Address>.

² Interoperability is the ability of two or more systems or components to exchange information and to use the information that has been exchanged.

to transfer the information to a separate computer program for further statistical analyses.

An agreed-upon standard for labeling or “tagging” each element of the computerized data set could facilitate the automatic identification and processing of such information. For example, the economist’s Web page would likely display many numbers representing specific pieces of information. The number “2,400,000.00” might appear, representing the value of soybeans produced in a given place at a given time. Even if the economist’s computer had been programmed to analyze agricultural cost data, it would not be able to recognize that “2,400,000.00” referred to a specific value for soybeans at a given place and time, unless the number were tagged with that descriptive information in a format the computer system understood. Tagging data according to standard formats and definitions would allow systems that recognize those standards to readily understand and process the data.

Currently, the XML set of standards is generally considered to be a primary candidate for filling the role of an Internet family of standards for tagging data. If implemented broadly and consistently, XML offers the promise of making it significantly easier for organizations and individuals to identify, integrate, and process complex information that may initially be widely dispersed among systems and organizations. For example, law enforcement agencies could potentially better identify and retrieve information about criminal suspects from any number of federal, state, and local databases. Further, XML could also make it easier to conduct business transactions over the Internet, because it offers a standard way to label and package the information that needs to be exchanged to conduct electronic business.

Rather than a single specification, XML is a collection of related standards. Two types of standards are essential for effective use of XML across organizations in either the public or private sector: (1) technical standards, which define the basic rules for tagging, structuring, and displaying information; and (2) business standards, which provide the vocabulary and protocols for conducting business electronically. The core XML standard was designed to accommodate a wide variety of supplemental standards, or extensions, to address additional functions and meet specialized needs.

XML is not the first attempt by IT developers—or the federal government—to standardize the process of data exchange. Much effort, for example, was spent over many years to develop the Electronic Data Interchange (EDI) standards, which remain in use today and are expected

to continue in use alongside XML. However, EDI use has been largely limited to data exchanges among large organizations, because implementing EDI generally entails buying customized proprietary software and setting up expensive, private communications networks. XML has the potential for broader implementation because it requires less customization and uses the Internet's data communications infrastructure, which is already in place.

Federal XML projects undertaken to date have varied significantly in size and scope. In many cases, agencies have used XML to enhance data exchange within well-defined communities of interest with well-defined data exchange requirements. In addition, several larger agencies have been making efforts to define XML-related data standards for larger communities of interest. For example, the Environmental Protection Agency has been working with state environmental agencies to develop XML data standards for a national network of environmental information.

Results in Brief

While XML's technical standards—such as specifications for tagging, exchanging, and displaying information—have largely been worked out by commercial standards-setting organizations and are already in use, equally important business standards are not as mature and may complicate near-term implementation. For example, standards are not yet complete for (1) identifying potential business partners for transactions, (2) exchanging precise technical information about the nature of proposed transactions so that the partners can agree to them, and (3) executing agreed-upon transactions in a formal, legally binding manner. Many standards-setting organizations in the private sector are creating various XML business standards, and it will be important for the federal government to adopt those that achieve widespread acceptance. However, it is not yet clear which business standards meet this criterion. In addition, key XML vocabularies tailored to address specific industries and business activities are still in development and not yet ready for governmentwide adoption.

Given that a complete set of XML-related standards is not yet available, system developers must be wary of several pitfalls associated with implementing XML that could limit its potential to facilitate broad information exchange or adversely affect interoperability, including (1) the risk that redundant data definitions, vocabularies, and structures will proliferate, (2) the potential for proprietary extensions to be built that would defeat XML's goal of broad interoperability, and (3) the need to maintain adequate security. In addition to these pitfalls, which all systems developers must address, the federal government faces additional

challenges as it attempts to gain the most from XML's potential. Specifically:

- No explicit governmentwide strategy for XML adoption has been defined to guide agency implementation efforts and ensure that agency enterprise architectures address incorporation of XML. Although agencies need flexibility to tailor XML-based systems to meet their unique needs, they risk building and buying systems that will not work with each other in the future if their efforts do not take place within the context of a well-defined strategy.
- The needs of federal agencies have not been uniformly identified and consolidated so that they can be represented effectively before key standards-setting bodies. It will be important for the federal government to leverage and build upon commercially developed standards and XML vocabularies as they become mature and widely accepted. If federal requirements are not better understood and consolidated, the government may be unable to effectively provide input to these standards while they are still under development.
- The government has not yet established a registry of government-unique XML data structures (such as data element tags and associated data definitions) that system developers can consult when building or modifying XML-based systems. Without such a registry, developers are less likely to build systems using compatible data definitions, which would likely defeat the goal of broad data access and exchange. In order to establish such a registry, policies and procedures for adding tag definitions and maintaining the system would also be needed and have not yet been developed.
- Much also needs to be done to ensure that agencies address XML implementation through enterprise architectures so that they can maximize XML's benefits and forestall costly future reworking of their systems.

To address these challenges, GAO is making recommendations to the director, Office of Management and Budget (OMB), to enhance federal planning for adoption of XML.

Principal Findings

A Complete Set of Standards for Implementing XML Is Only Partially in Place

Key technical standards for XML have been largely worked out under the auspices of the World Wide Web Consortium (W3C).³ These technical standards are focused on providing the generic structure and tools to tag data, transmit it over the Internet, and allow it to be processed by the computer systems that receive it.

Business standards, though equally important, are generally less well-developed, and reaching agreement on them is proving to be difficult when multiple communities of interest are involved. Business standards are needed to provide a more complete framework for conducting business over the Internet, including advertising products and services so that potential buyers and sellers can find each other, proposing and agreeing upon electronic transactions, and executing the agreed-upon transactions. Business standards are also needed to define vocabularies for the specific data elements that are to be exchanged when these transactions are conducted.

Unlike XML technical standards, which are all established and maintained by the W3C, business standards are developed by a variety of public and private sector organizations, including industry consortia, and are not always universally supported. For example, a number of different approaches to addressing the process of conducting business transactions have been proposed, including electronic business XML (ebXML), RosettaNet, and XML-based Web services. These different approaches continue to vie for support and offer functionality that is in part overlapping and incompatible. Because uncertainty remains about which business standards will ultimately prevail, applications based on any of the current proposals may be at risk of being incompatible with future standards. In addition, without universally accepted standards, commercial IT vendors may be using XML extensions that are nonstandard and divergent and that may limit interoperability.

In industries and professions where needs are well-defined and cohesive communities of interest exist, standard data vocabularies have been

³ The W3C was founded in 1994 by Tim Berners-Lee, the inventor of the Web, to lead development of common protocols that promote the evolution of the Web and ensure interoperability.

successfully developed. For example, mathematicians have created an XML vocabulary called the Mathematical Markup Language that allows them to insert equations into Web pages that can then be copied into specialized software applications and immediately used for calculations. Some of these vocabularies, once fully developed, may be useful to the government as well. However, many of these potentially useful standard vocabularies are still in the initial stages of development and do not provide all the data structures needed to support current needs. Using them at this time would mean taking the risk that future developments could diverge from these early standards and limit interoperability with them. As a result, they are not yet ready for governmentwide adoption.

The Federal Government Faces Challenges in Realizing XML's Full Potential

Although XML offers the potential to greatly facilitate the identification, integration, and processing of complex information—both within the federal government and externally—system developers face a number of pitfalls in implementing the technology. One risk is that markup languages, data definitions, and data structures will proliferate. If organizations develop their systems using unique, nonstandard data definitions and structures, they will be unable to share their data externally without providing additional instructions to translate data structures from one organization and system to another, thus defeating one of XML's major benefits. Likewise, software vendors and system developers may be tempted to add proprietary extensions to the XML standards when they build specific systems. Such systems might then be less able to freely exchange information with other XML-enabled systems. In addition, implementing XML in an organization could create new security vulnerabilities if steps are not taken in designing the system to mitigate this risk.

In addition to these pitfalls, which all systems developers must address, the federal government faces additional challenges as it attempts to gain the most from XML's potential. Specifically:

To date, neither OMB, which is responsible for developing and overseeing governmentwide policies and guidelines for agency IT management, nor the National Institute of Standards and Technology (NIST), which is responsible for developing federal information processing standards and guidelines, have formulated an explicit governmentwide strategy for XML adoption to guide agency implementation efforts and ensure that agency enterprise architectures address incorporation of XML. Activities within the federal government to promote broad governmentwide adoption of XML technology have been limited. Most governmentwide coordination

has been limited to the activities of the XML Working Group, chartered by the federal Chief Information Officers (CIO) Council. The working group's activities have focused on education and outreach rather than developing a strategy for adopting XML. Without agreement on a governmentwide implementation strategy, agencies risk building and buying systems that will not work with each other in the future.

The federal government as a whole has neither identified cross-agency and governmentwide requirements for XML nor developed a dictionary of inherently governmental data tags and definitions. Further, no process has been defined for consolidated collaboration with commercial standards bodies to ensure that government requirements are identified and incorporated. Past experience coordinating federal requirements for EDI suggests that an effective approach is to task a central committee with collecting requirements from federal agencies and representing the government on key standards groups.

Given that it is challenging to agree upon predefined XML vocabularies, other approaches can be adopted to encourage broad, consistent use of data definitions and structures. Specifically, a "bottom up" approach is to establish a centralized registry of key XML data elements and structures and coordinate its use by XML systems developers. With this arrangement, developers have the incentive to reuse data structures found in the registry because doing so reduces costs and brings about interoperability with other existing systems. The federal XML Working Group, chartered by the CIO Council, is working to create a pilot version of a governmentwide registry, based on a registry previously developed by the Defense Logistics Agency. However, further work will be needed to set policies and guidelines to ensure the effectiveness of the registry in promoting governmentwide systems interoperability.

Another avenue for promoting interoperability is to ensure that sound XML implementation strategies are adopted and documented on an agency-by-agency basis through development of enterprise architectures. Effective XML implementation depends on complete and well-established data definitions and structures, which can be best obtained through the process of defining and adopting an enterprise architecture. Such an architecture provides the foundation for maximizing XML's benefits and forestalling costly future reworking of agency systems.

If these challenges are not addressed, the use of XML in the federal government may have only limited benefits and may not achieve the

technology's promise of facilitating broad interoperability among disparate systems.

Recommendations for Executive Action

Given the statutory responsibility of OMB to develop and oversee governmentwide policies and guidelines for agency IT management, we recommend that the director of OMB, working in concert with the federal CIO Council and NIST, develop a strategy for governmentwide adoption of XML to guide agency implementation efforts and ensure that the technology is addressed in agency enterprise architectures. This strategy should, at a minimum, address how the federal government will address the following tasks:

- Developing a process with defined roles, responsibilities, and accountability for identifying and coordinating government-unique requirements and presenting consolidated, focused input to private sector standards-setting bodies during the development of XML standards. This process could be patterned after the current process that is in place for EDI coordination among federal agencies, or OMB might consider adapting the EDI process to cover XML as well. Guiding the overall process should be the presumption that mature, agreed-upon commercial standards will be adopted by the government whenever possible.
- Developing a project plan for transitioning the CIO Council's pilot XML registry effort into an operational governmentwide resource. This plan should include identifying time frames and resources needed to implement and maintain an operational registry linked to agency repositories of standard data structures.
- Setting policies and guidelines for managing and participating in the governmentwide XML registry, once it is operational, to ensure its effectiveness in promoting data sharing capabilities among federal agencies. These policies should clarify the roles and responsibilities of specific agencies and should consider including definitions of classes of compliance, which could be used to categorize how rigorously organizations adhere to the policies. Further, these policies should promote the consistent use of XML namespaces to resolve potential ambiguity in data references across XML documents.

In addition, as part of its ongoing process for reviewing agency IT architectures and annual budget requests, we recommend that OMB ensure that agencies' business needs for XML technology are defined in their enterprise architectures. Specifically, OMB should specify requirements for documenting the usage of XML standards and products

in the standards profile section of the architecture—the section that defines the set of rules governing systems implementation and operation.

Agency Comments and Our Evaluation

In oral comments on a draft of this report, officials from OMB's Office of Information and Regulatory Affairs, including the Information Policy and Technology Branch chief, generally agreed with our findings and conclusions and stated that they would consider our recommendations. The officials also provided information on recent OMB actions aimed at promoting the adoption of XML by federal agencies. We have incorporated this updated information in the report. We view these recent OMB actions as positive steps. Nevertheless, we also believe that OMB can improve on these actions by implementing the recommendations in this report.

We received oral comments from the co-chairmen of the XML Working Group; officials of NIST's Information Technology Laboratory; and the deputy associate administrator, Office of Electronic Commerce, General Services Administration. We also received written comments from the chief information officer, National Aeronautics and Space Administration; and the director for policy and communications staff, National Archives and Records Administration. Letters from these latter two agencies are reprinted in appendixes I and II. All of the agency officials who reviewed the draft agreed with the overall content of the report. Officials from the XML Working Group and the National Archives and Records Administration expressed concern that the draft overemphasized the value of a "top down" XML implementation strategy that emphasizes executive direction and guidance as opposed to a "bottom up" approach relying on individual initiative at lower management levels. We believe that it is important to strike a balance between the two approaches. In response to this concern, we are including language in the final report to emphasize that a balance between the bottom up and top down approaches is needed. In addition, each agency provided technical comments, which have been addressed where appropriate in the final report.

Chapter 1: Background: Features and Current Federal Use of XML

Advances in the use of information technology (IT)—especially the rise of the Internet—are changing the way organizations communicate, exchange information, and conduct business among themselves and with the public. The Internet offers the opportunity for a much broader exchange of information than was previously possible, because it provides a virtually universal communications link to the multitude of disparate systems operated by private sector businesses, government agencies, and other organizations.

However, although the Internet can facilitate the exchange of information, much of the information displayed to users is delivered only as a stream of computer code to be visually displayed by Web browsers, such as Internet Explorer or Netscape Communicator. Without human intervention, such information cannot be extracted and reused for other purposes. For example, an economist might visit a Web page that displayed statistical information about the production of various agricultural commodities over a number of years. Typically, such a Web page would only display this information to the economist to examine visually on his or her computer screen. Without special translation software, it would likely be difficult for the economist to transfer the information to a separate computer program for further statistical analyses.

An agreed-upon standard for annotating or “tagging” each element of the computerized data set could facilitate the automatic identification and processing of such information. For example, the economist’s Web page would likely display many numbers representing specific pieces of information. The number “2,400,000.00” might appear, representing the value of soybeans produced in a given place at a given time. Even if the computer system had been programmed to analyze agricultural cost data, it would not be able to recognize that “2,400,000.00” referred to a specific value for soybeans at a given place and time, unless the number were tagged with that descriptive information in a format that the computer system understood.

Tagging data in a standard way allows any system that recognizes the standard to readily understand and process data that conforms to that standard. In tagging, a standard format is used to label each element of a data set with metadata¹ that clarifies what kind of information is being

¹ Metadata are data containing descriptive information about other data. For example, a block of numerical data might be identified in metadata as representing unit cost in dollars.

provided. Common tagging systems for electronic information—also known as markup languages—use labels set off by angled brackets to show where data elements begin and end: for example, in <label> data </label>, the second tag includes a slash to indicate that it is a closing tag.

The Extensible Markup Language (XML) is a flexible, nonproprietary set of standards for tagging information so that it can be transmitted over a network such as the Internet and readily interpreted by disparate computer systems. If implemented broadly with consistent data definitions and structures, XML offers the promise of making it significantly easier for organizations and individuals to (1) identify, integrate, and process information that may initially be widely dispersed among systems and organizations, and (2) conduct transactions based on exchanging and processing such information—a key element for federal agencies positioning themselves to provide electronic government services to citizens and businesses.

In a previous attempt to standardize the process of data exchange, much effort was spent over many years to develop Electronic Data Interchange (EDI) standards, which are in use today and will probably continue to be used alongside XML. However, their use has been largely limited to data exchanges among large businesses and government agencies, because implementing EDI generally entails buying customized proprietary software and setting up expensive, private communications networks. XML has the potential for broader implementation because it was designed to take advantage of the Internet's capabilities and protocols, which are already in place.

Federal XML projects undertaken to date have varied significantly in size and scope. In many cases, agencies have used XML to enhance data exchange within well-defined communities of interest with well-defined data exchange requirements. In addition, several larger agencies have been making efforts to define XML-related data standards for larger communities of interest. For example, the Environmental Protection Agency (EPA) has been working with state environmental agencies to develop XML data standards for a national network of environmental information.

Standardized Data Tagging Facilitates Information Exchange among Disparate Systems

Identifying, exchanging, and integrating information from different and perhaps unfamiliar sources are functions that are essential to the effective use of networked information for a wide range of goals, including the provision of electronic government services. Federal agencies exchange data with many external entities, including other federal and state agencies, private organizations, and foreign governments. For example, federal agencies routinely use data exchanges to transfer funds to contractors and grantees; collect data necessary to make eligibility determinations for veterans, social security, and Medicare benefits; gather data on program activities to determine if funds are being expended as intended and the expected outcomes achieved; and share weather information that is essential for air flight safety.

If a data exchange does not function properly, the data being received by a computer system could cause it to malfunction or produce inaccurate results, or the data may not be received at all. However, because systems providing information to an organization are frequently external or were developed for other purposes, they may structure and format the needed information in incompatible and unpredictable ways, making data exchange problematic. Effective data sharing among computer systems faces many problems, including

- incompatible operating systems and hardware platforms,
- incompatible computer applications written in different programming languages,
- inconsistent or poorly developed data definitions, and
- incompatible data transmission protocols.

Without predefined standards in place, systems developers may need to define in detail the precise steps to be taken to carry out the exchange of a set of data, and these definitions must be encoded in the software and hardware of both transmitting and receiving systems—a potentially complex, time-consuming, and expensive process.

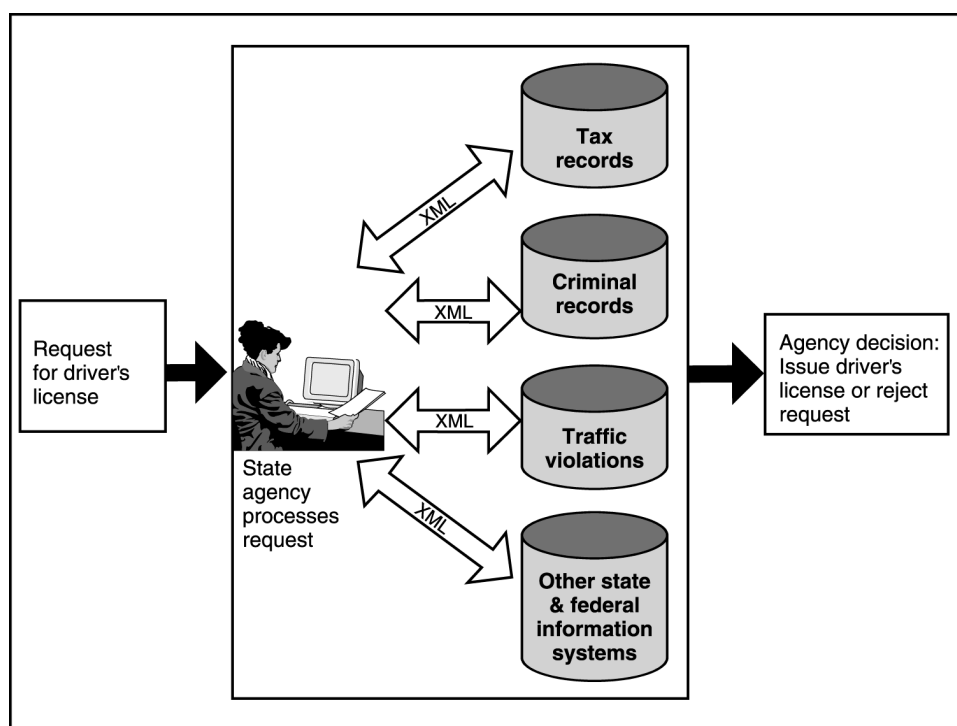
In contrast, if standards are in place for how data are structured and tagged, it can be more efficient and less expensive to develop interfaces, and as a result data exchange can be facilitated. A hypothetical state driver's license system offers a good conceptual example of the potential benefits of a data tagging standard for (1) interfacing disparate systems and (2) locating and sharing data among these systems. In processing an application for a driver's license, a state government agency might want to consult a number of local, state, or federal databases before issuing or renewing the license, including records of residency, traffic violations,

criminal convictions, tax payments, and others. In today's environment, each of these systems could be operated by a different entity and could use incompatible systems software and computer applications, which could cause data-sharing problems. One solution would be to tag data in a standard way so that it could be easily shared among all these systems and databases.

Standardized tagging helps solve the problem by formatting both the data and relevant information about the data according to a standard that can be readily interpreted by any other system that recognizes that format and understands the data definitions and structures that are used. In our example, each state agency may have relevant information about a drivers' license applicant stored in a different format. The applicant's name might be called "Name" in one system but divided into "Lastname," "Firstname," and "MiddleInitial" in another system. Further, the database system software running at each agency might use different commands and programming syntax to access and query its databases, requiring that any system wanting to connect and access its data conform to that agency's unique structures. However, if the data were made available to other organizations using a standardized tagged format, these agency-unique discrepancies could be overcome. All name information, for example, might be consistently tagged as <Name>. Even if it did not use this standard tag internally, each state agency would be responsible for matching up its internal data structures to the appropriate standard data tags, which would have agreed-upon definitions. The standard tags would make it easy to connect to each agency and exchange relevant information, because each exchange would use the same format to transfer the data and annotate (tag) what it means. Of course, policies and procedures would still be needed to ensure that the data were exchanged only for authorized purposes, and each system would have to conform to the standards in use and agree on standard data definitions and structures.

Figure 1 shows the role that a set of tagging standards such as XML could play in facilitating data sharing among disparate agencies.

Figure 1: A Hypothetical XML-Based State Driver's License System



Source: GAO.

Tagging data in a consistent, standard way can also make it much easier to locate information that is dispersed among incompatible computer databases and difficult to access. In the example of the driver's license application, the fact that an applicant had a criminal record might remain unknown to the licensing agency if the information was stored in an incompatible—and thus inaccessible—database. On the other hand, consistent, standardized tagging would help make the information much easier to find, because the licensing agency could perform a search based on a standard tag definition, knowing that all relevant information should be tagged in the same way and thus should be identified by that search. The standardized tagging of data has the potential to bring a similar benefit to individuals searching for information over the Internet. Instead of simply finding instances of text that match a given string of characters, Web-based search engines could locate and report on data by examining tags reflecting the content of the data. In all likelihood, such searches would produce more focused and useful results.

XML Supports Internet-Based Data Exchange

XML is a nonproprietary set of standards for tagging information so that it can be transmitted over a network such as the Internet and readily interpreted by many different computer systems. It is platform-independent, meaning that it can operate on any combination of computer hardware and XML-enabled software. The core XML standard, known as XML 1.0, was adopted in 1998 by the World Wide Web Consortium (W3C), which has jurisdiction over the Internet's technical standards. It is a subset of the well-established Standard Generalized Markup Language, which was approved and published by the International Organization for Standardization in the 1980s² and is used primarily in large organizations for tagging technical documents.

XML code is designed to be clearly intelligible to a human reader and involves embedding descriptive tags around data in a computerized text file. Figure 2 shows a simple example where “President George Washington” has been tagged in XML to indicate what kind of data each of the three words represents. The “NAME” tag uses a hierarchical structuring capability to distinguish two subcategories of tags, “FIRST” and “LAST.” All XML documents have the ability to structure data in a similar hierarchical manner. The example also includes the use of a data attribute—a rank of “1” has been assigned to the office of the president.

Figure 2: XML Code Example

```
<?xml version="1.0"?>
<OFFICE RANK="1">President</OFFICE>
<NAME>
  <FIRST>George</FIRST>
  <LAST>Washington</LAST>
</NAME>
```

Hypertext Markup Language (HTML), the current standard for displaying information on the World Wide Web, also uses tags embedded in text files and is also a subset of the Standard Generalized Markup Language.

² *Standard Generalized Markup Language*, ISO 8879:1986.

However, unlike XML, HTML's tags are predefined and are used solely to transmit instructions for displaying information on Web pages. HTML tags describe document structures (that is, whether text should be treated as a heading, a list, a quotation, and so on) and document appearance (such as whether text should be emphasized, larger or smaller than surrounding text, or in a particular type font or color). A Web browser that receives an HTML file simply displays the stream of data that it receives according to the HTML instructions, without "understanding" what information it is displaying. Table 1 summarizes the differences and similarities between HTML and XML.

Table 1: Comparison of HTML and XML

	HTML	XML
Differences	Tags are predefined and are intended to provide formatting and display instructions.	Data tags are not predefined and can be used to label data according to any hierarchical structure.
	Data in HTML documents generally cannot be interpreted and processed without human intervention.	Data in XML documents can be automatically interpreted and processed by XML-enabled systems.
	Strength is in displaying information on a Web browser.	Strength is in facilitating data exchange.
	HTML is designed to overlook syntactical errors and focus on displaying information.	XML is designed to check for syntactical errors and ensure conformance with data structures (or templates), when specified.
Similarities	Both are nonproprietary W3C standards that can potentially work on a variety of computer systems.	
	Both are designed to rely on Internet protocols as a means of providing connectivity to a broad range of systems.	
	Both are based on the Standard Generalized Markup Language and thus are structured as text files with tags that can be read and understood by humans.	

When a system using XML is developed, several basic components may be needed to provide ways to do such things as (1) define the tags that are used in an XML document, (2) validate the correct use of a document's tags, and (3) provide formatting instructions for displaying the data. Table 2 summarizes important basic components that are often part of XML implementations currently in use.

Table 2: Basic XML Components

Component	Description
XML document	A text document marked up with descriptive tags and attributes. An XML document can also begin with declarations that refer to other files providing further instructions for interpreting and displaying data elements.
Document type definition (DTD) or XML schema	A DTD is a file that describes the structure of XML documents and defines how markup tags should be interpreted. A DTD can be used to automatically interpret multiple documents in a uniform way. XML schemas serve the same function as DTDs but provide greater definitional power and are more flexible. For example, XML schemas can specify what type of data a tag refers to—such as whether it is an integer or a text string.
Parser	Software that reads an XML document and determines the structure and properties of the data in the document.
Style sheet	A text file that provides instructions for formatting and displaying the information in XML documents. Style sheets can include variations depending on the type of device used to access the document. For example, the same XML document could be displayed differently on a handheld wireless computer or a desktop computer, based on different style sheets.
XML namespace	A unique identifier, such as a Web address, referenced at the start of an XML document as a source for definitions of the tags and other data structures used in the document. An XML document can reference more than one namespace.

XML’s Technical
Standards Provide the
Tools to Describe and
Exchange Data over
the Internet

Because the core W3C XML 1.0 standard provides only limited features, an entire family of related technical standards has been developed to define and structure in greater detail the ways in which XML is to be used. XML’s technical standards define the basic rules for using XML components to tag, structure, and display information. Technical standards can be divided into two groups: core standards and supplemental extensions. Core technical standards developed by the W3C provide the fundamental rules for using XML and include the following:

- XML 1.0 specifies how to use markup symbols to define and describe the content of data elements and their associated attributes. By design, XML 1.0 does not focus on providing specifications for document processing, such as specific presentation formats and processing instructions. Rather, these issues are addressed by other standards.
- XML Stylesheet Language (XSL) describes how to use electronic files called style sheets to provide instructions for formatting XML documents for display in a variety of visual media. Different style sheets are created and used to display the same XML document on different media, such as a desktop computer or a palm-sized device. XSL includes two extensions of its own—XSL Transformations (XSLT) and XSL Formatting Objects (XSL-FO). XSLT makes it possible to convert (or transform) the original structure of an XML document to match the structure of another XML

document. XSL-FO provides the formatting vocabulary to carry out such a transformation.

- The XML Schema standard provides a superset of the capabilities found in XML 1.0 for document type definitions (DTDs). It offers comprehensive instructions for describing the structure and constraining the contents of XML documents. The XML Schema standard also specifies a robust system of data types, including a number of predefined data types that can be associated with XML data elements and attributes to help manage dates, numbers, and other special forms of information.
- The XML Namespace standard provides guidelines for uniquely identifying the data definitions that appear in an XML document, thus avoiding ambiguity among data elements with the same name that may come from different sources.

In addition to these core standards, a number of supplemental standards have been developed or are proposed to codify how additional functions should be performed. When developers identify a need for new functions to be incorporated into XML technology, new supplemental specifications can be developed as extensions to the core XML standards. These supplemental specifications have been designed as separate standards so that they can be used when needed as modular enhancements to individual implementations. Examples of supplemental technical standards include the following:

- The Document Object Model (DOM) is a platform-independent and language-neutral application-programming interface. DOM allows programmers to develop applications that can dynamically access and update the content and structure of XML documents.
- The XML Linking Language (XLink) standard allows XML documents to contain links similar to HTML hyperlinks. While XLink is similar to HTML linking, it adds new features to make links more flexible and precise. For example, XLink allows a link to point to a specific reference within an external file rather than simply pointing to the file as a whole, as in HTML.
- XML Path Language (XPath) provides a common syntax and semantics for addressing specific parts of an XML document. XPath gets its name through its use of a path notation for navigating through the hierarchical structure of an XML document.

XML Was Designed to Accommodate Numerous Extensions

An important advantage of XML is that it is flexible enough to accommodate an unlimited number of uses. Each new use is accommodated by the development and standardization of extensions to the core set of XML standards. This is what makes XML “extensible”; its structure can be adapted (or extended) to meet many different needs.

In addition to the supplemental technical standards already discussed, XML can accommodate extensions to suit the needs of specific communities of users, such as chemists, travel agents, and numerous others. As a result, many efforts are under way to define specialized tags and other XML data structures and processing protocols to suit a variety of specific business purposes. For example:

- Electronic business XML (ebXML) is being developed as a complete, modular suite of specifications to enable the conduct of business over the Internet.
- Mathematicians have created an extension of XML, called the Mathematical Markup Language, that allows them to insert equations into Web pages that can then be copied into specialized software applications and immediately used for calculations. The W3C has approved the Mathematical Markup Language as a standard.
- The HR-XML Consortium, an industry coalition, is developing XML vocabulary and data structures to meet the needs of the human capital field, including such functions as exchange of staffing data and payroll transactions.
- The Extensible Business Reporting Language (XBRL) was developed by a consortium of industry and public sector organizations as a standard for reporting and analysis of financial information.

XML Can Enhance Information Search, Retrieval, and Analysis

If widely implemented using consistent data definitions, XML can be a very effective tool to facilitate searching for, identifying, and integrating information from different and perhaps unfamiliar sources. For example, because XML uses data tags (as discussed earlier), it can be used for more precise data queries and collections, both locally (for a specific organization) and across the Internet. XML's data tags can be used to precisely identify individual data elements, allowing XML-based systems to collect and integrate specific types of data relatively easily from a variety of sources and create reports or support other kinds of analysis that otherwise might require a much more labor-intensive effort. For example, the federal government annually produces many reports with large amounts of tabular data, such as cost figures and other numerical statistics. If tagged in XML using agreed-upon data definitions, specific data elements could be located within these tables, retrieved, and recombined to form a new kind of analysis. In fact, the data could be dynamically retrieved each time the analysis was examined, if up-to-the minute information were desired. Officials from the EPA and other federal agencies are currently working on a centralized Web site for federal

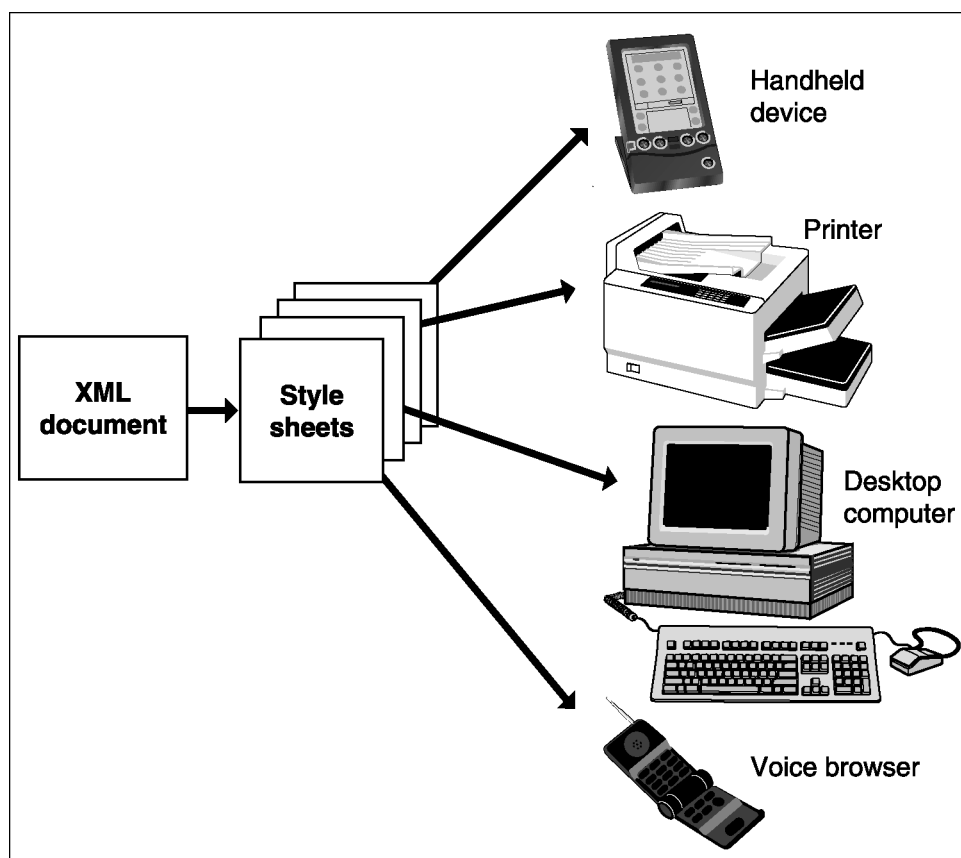
government statistical information—called FedStats—with the objective of using XML to provide this kind of capability.

Similarly, XML could be used to enhance general Web search engines. As mentioned earlier, the use of data tagging would provide for more precise searching than current approaches, which are based on relatively crude quantitative measures, such as the frequency of occurrence of a given string of text or the proximity of one text string to another. Some databases have already been developed to take advantage of this feature of XML. The news agency Reuters, for example, which has archived over 800,000 news stories, used XML tags to classify these into 775 searchable categories.

Once XML code is written, not only its creators but also external parties can potentially reuse it. For example, after Amtrak created an XML system to access its application and database system, the associated data tags and structures were reused for a voice recognition reservation system. According to XML experts, additional cost savings may be realized in the future as well, because it will likely be easy for new systems and applications to recognize and make use of XML data.

XML's extensibility also facilitates interaction among a variety of devices. The same XML document can be interpreted through different style sheets to suit any number of different display devices. Figure 3 illustrates this benefit.

Figure 3: XML Can Facilitate the Use of Different User Interfaces and Display Devices



Source: GAO.

XML Usage Complements Traditional Electronic Data Interchange Applications

XML does not represent the first attempt by IT developers—or the federal government—to standardize the process of data exchange. The EDI³ standards were also developed for this purpose, but their use has been limited. EDI has been implemented mostly by large organizations, which have the resources to buy the custom software generally required and to set up private communications networks. Another obstacle to implementing EDI is that it requires individuals with specialized knowledge to perform tasks such as converting an organization's business

³ EDI is the automated exchange of predefined and structured business data among information systems of two or more organizations.

data into the correct formats of the transmission standard, an often complex and time-consuming process. In contrast, XML has the potential to be more widely adopted, since it was designed to use the Internet's data communications infrastructure, which is already in place.

The EDI set of standards consists of electronic message formats for many business-related documents used in electronic transactions. Figure 4 is an example of an EDI-formatted "Request for Quotation" that adheres to the American National Standards Institute (ANSI) Accredited Standards Committee (ASC) X12 EDI standard. As the figure shows, data in an EDI-formatted document are cryptic. This is a major difference between EDI and XML, which uses simple text files and tags that are intended to convey readily understandable meaning (see figure 2). The cryptic format of EDI standards serves as an impediment to their broad adoption, because extensive, specialized knowledge is required to interpret EDI messages, troubleshoot problems, and adapt existing systems to conform to the standards.

Figure 4: A "Request for Quotation" Formatted as an EDI Message

```
ISA*00*      *00*      *ZZ*GATEC      *ZZ*PUBLIC      *960508*...
GS*RQ*GATEC*PUBLIC*960508*1237*000721330*X*003010
ST*840*000721331
BQT*00*F3360196T7174001*960508*106*960509
REF*IL*FM230061280242
PER*IC**EM*F33601@EC099.LLNL.GOV
DTM*002*960517
PO1*1*54*BX***FT*8940*SI*5499*FS*8940011728888*MF*SANDOZ ...
1*MF*SANDOZ NUTRITION*MG*NDE 00212-4580-01
PID*F****SUPPLEMENT, TOLEREX, DIETARY,
CTT*1
SE*16*000721331
GE*1*000721330
IEA*1*000721332
```

Source: Department of Defense.

EDI has been the primary data format used by large organizations to transfer business data among themselves, and it continues in widespread use. After an extensive effort to participate in and encourage the development of EDI standards, key federal government agencies such as the Department of Defense (DOD) and General Services Administration (GSA) adopted EDI as the standard format for data interchange for a

number of their business systems. However, smaller federal agencies generally have not made the same commitment to EDI. Lacking the necessary skills and resources, many small and midsize companies also have not adopted EDI. Accordingly, EDI-enabled organizations have been unable to conduct automated electronic business with those organizations that have not developed the same capability. As a result, EDI has not attained universal use as a data exchange standard.

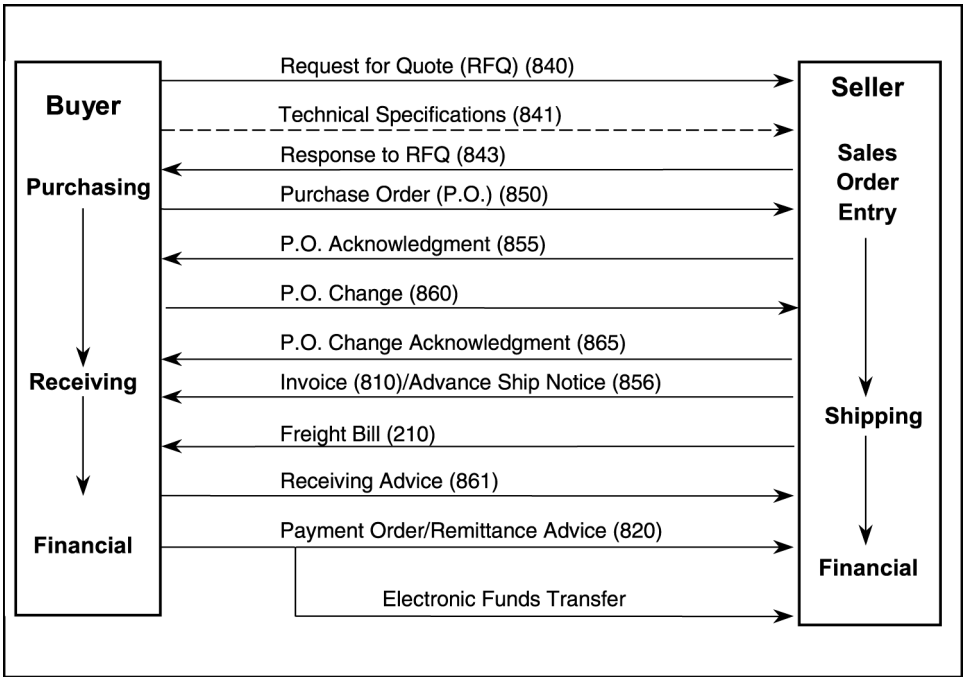
According to reports from Giga Information Group⁴ and the Logistics Management Institute,⁵ XML is not a replacement, but a complementary technology for EDI. Although both EDI and XML can be used to accomplish the same basic task—facilitating the transfer of business data from one system to another—each technology has advantages and disadvantages. Depending on business needs, the two can be used together, particularly if companies have already invested in EDI methodologies. The convergence of EDI and XML can provide a potentially lower cost alternative for small and midsize companies to conduct business with federal agencies that already have traditional EDI systems in place.

One advantage of EDI is that a full suite of standards is already in place to support business transactions. For example, figure 5 depicts the typical flow of electronic documents between a buyer and seller in an acquisition process using ANSI ASC X12 EDI transactions.

⁴ Giga Information Group, *XML's Role in the EDI World* (June 23, 2000).

⁵ Logistics Management Institute, *Open Buying on the Internet and Extensible Markup Language: Recommendations on Adoption by the Federal Government* (January 2000).

Figure 5: Typical Flow of Business Transactions Based on EDI Standards



Source: Department of Defense.

XML has the potential to lower costs for data exchange because it can take advantage of the Internet's communications infrastructure and protocols.⁶ EDI, on the other hand, was developed before the Internet became commonplace and thus has generally involved buying customized software and setting up expensive, private communications networks. These features have some advantages: the dedicated links associated with private communications networks are generally more reliable than a simple Internet connection, and the condensed format of EDI transactions makes it possible to transmit them much more efficiently than XML documents. However, the expense involved in attaining this capability is likely prohibitive for many applications. Table 3 provides a summary comparison of the major features of EDI and XML.

⁶ Widely used Internet protocols include Simple Mail Transfer Protocol (SMTP) for electronic mail, Hypertext Transfer Protocol (HTTP) for the World Wide Web, File Transfer Protocol (FTP) for file transfer, and others.

Table 3: Comparison of EDI and XML

	EDI	XML
Differences	Is based on industrywide EDI business standards, such as EDIFACT and ANSI X12, that are well-established, providing standard electronic formats for electronic transactions.	Lacks a complete set of business standards to support XML-based electronic transactions that are broadly agreed upon.
	Uses highly structured predefined formats that have specific, narrowly defined purposes.	Has the flexibility to allow new vocabularies to be defined to meet changing business needs.
	Originally designed to rely on private networks known as “value-added networks” for data exchange.	Designed to take advantage of the Internet’s capabilities and existing protocols for data exchange.
	Supports data exchange only.	In addition to data exchange, supports other data handling functions, such as content management and sophisticated Web searches.
Similarities	Both standards are freely available and nonproprietary.	
	Both facilitate data exchange between disparate computer applications.	
	Both allow developers to add proprietary extensions to their specific implementations.	

Federal XML Projects Vary in Size and Scope

XML is being broadly implemented, both commercially and within government. In the private sector, the Giga Information Group published the results of a survey to gauge the adoption of XML among its client base in April 2001.⁷ Based on responses from 80 businesses ranging from banking and insurance to health care and manufacturing, 81 percent said they had begun using XML in their organizations. Of the 18 percent of respondents who said they had not, 76 percent planned to use XML within the next year. The primary reported uses of XML were for enterprise application integration and business data exchange. Other areas of usage included data integration, publishing, content management, portals, and application development.

Federal XML projects undertaken to date have varied significantly in size and scope. In some cases, agencies have used XML to enhance data exchange within relatively narrow communities of interest with well-defined data exchange requirements. The Securities and Exchange Commission’s (SEC) Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system and Amtrak’s reservation system are two examples. In a few other cases, concerted efforts have been made to define XML-related data standards—or design a process for doing so—for larger communities

⁷ Giga Information Group, *Giga Survey: XML Achieving Mainstream Usage* (April 30, 2001).

of interest. Specifically, the Department of Justice has developed a set of definitions for basic data elements shared by several law enforcement information networks. Similarly, EPA has been working with state environmental agencies to develop XML data standards for a national network of environmental information. Several efforts are also under way within DOD to develop a common infrastructure to support the use of XML across the department.

Securities and Exchange Commission

In the SEC's case, agency officials made the decision to design their modernized EDGAR system to use XML for all external data exchanges as well as internal processing. However, as it is currently operating, EDGAR continues to use other more commonly known document formats because many external systems that interact with EDGAR are not yet XML-compliant.

According to agency officials, since 1992, the SEC has used EDGAR to electronically collect the financial and other business information that public companies are required by law to submit on a regular basis. As part of a larger modernization effort, the SEC in April 2001 began requiring that submissions be formatted with headers encoded in XML. The agency's EDGARLink client software, distributed to filers at no charge, uses a specialized vocabulary called the Extensible Forms Description Language to format headers in XML for transmission to the SEC. Although SEC officials have not quantified any cost savings associated with implementing XML, they believe its use has saved the agency software development expenses, because filers now use a commercial off-the-shelf product to format their submissions, instead of custom software, as had been previously required. According to SEC officials, third-party software developers should also be able to reduce costs by using commercial XML products to format submissions.

SEC officials stated that their use of XML to date has been limited to functions that did not require coordination with other government or private sector organizations. Because the SEC provides filers with copies of the XML-formatting software at no charge, it has been able to fully control how XML is implemented in the software and what specific vocabulary is used. The Extensible Forms Description Language that was used has been submitted to the W3C as a proposed standard but has not yet been approved.

SEC officials would like to broaden the use of XML to cover all the data in EDGAR filings rather than just header information. Doing so would take

much fuller advantage of XML's strengths and allow investors to better access financial data and automatically perform many kinds of analyses. However, to do so would require agreement on a complete vocabulary of data tags and schemas for describing financial statement information, which could require coordinating with other groups such as the XBRL.org consortium, which is also developing business vocabularies related to financial reporting. Further, in addition to agreeing upon a standardized vocabulary, developers would need to make software available to format financial information according to the standards so that it would not be burdensome for filers to conform to the standard vocabulary. Since none of this has yet happened, SEC officials believe it is not in the best interest of filers to levy an XML requirement at this time.

Amtrak

Amtrak, a federally chartered corporation, has successfully used XML to enhance its reservation system, according to Amtrak officials. However, in doing so, officials say they have consciously taken the risk that their self-defined data structures may not match industry standards that emerge in the future. According to Amtrak officials, the use of XML has streamlined software development, including reducing costs, and produced an easier set of specifications for travel agencies to address when developing or modifying their own systems. In moving to XML, Amtrak officials found that they were the first in the railroad industry to attempt to convert their data to XML format, and thus they were free to define data tags as they wished. They decided to base their definitions on specifications developed by the OpenTravel Alliance⁸ but found that those specifications were not sufficiently articulated to meet all of Amtrak's needs. As a result, Amtrak defined new tags for rail reservations purposes when none were available. Amtrak officials told us that they expect the OpenTravel Alliance to continue to develop its specifications, and tags may be standardized that are incompatible with Amtrak's. In that case, Amtrak will likely have to modify its system to meet the new industry standards.

Department of Justice

The Department of Justice reported in October 2001 that it had taken steps to move beyond single-system implementations of XML and facilitate broader information sharing and integration of justice information systems

⁸ The OpenTravel Alliance is a self-funded, nonprofit organization working to create and implement industrywide, open electronic business specifications. Membership in the alliance includes major airlines, hoteliers, car rental companies, travel agencies, and other interested parties.

nationwide.⁹ The need for effective data sharing among law enforcement agencies has been highlighted by the department's recent heightened efforts to combat the threat of terrorism. According to its October 2001 report, the department's experience to date shows that defining and implementing XML data standards across more than one organization is a complex process that requires a concerted effort.

Until recently, elements within the department had been working on three separate XML-related data standardization efforts: (1) a standard format for criminal histories, (2) a standard for law enforcement agencies to share criminal intelligence information, and (3) a data standard for electronic court filings. In June 2001, the department's working group on infrastructure and standards undertook an effort to reconcile the separate data tags and definitions that the three initiatives had developed. According to the department's lessons learned report, the reconciliation effort was an intense process that required the close cooperation of all participants. For example, in the beginning, the working group found that the three existing standards diverged in important ways for many basic data structures, such as how to represent individuals' names. Initially, representatives from the three different communities were reluctant to make changes in the existing definitions to accommodate a broader standard. However, ultimately the group was able to develop a draft "XML Justice Data Dictionary" containing 128 data elements.

Justice faces additional challenges in ensuring that its newly standardized data elements are broadly adopted. The department plans to establish an XML registry for these data elements but has not done so yet. Nor has a decision been made about working to integrate these elements into a developing commercial standard vocabulary, such as Legal XML. Both actions may be needed to promote the use of the department's data elements in law enforcement systems.

Environmental Protection Agency

Like Justice, EPA has attempted to work within its community of interest—state environmental protection agencies—to build an infrastructure for common access, both locally and nationally, to environmental information, according to EPA officials. EPA is required by law to collect a large volume of information from the states in order to

⁹ Lessons learned report of the XML subgroup of the Global Advisory Committee Infrastructure/Standards Working Group, Department of Justice, October 2001.

carry out its mandated functions, including oversight of state-level programs and administration of national programs. Since 1998, EPA and the states have been working on developing a National Environmental Information Exchange Network, using the Internet and standardized data templates, written in XML, to facilitate the exchange of data among participating partners. According to EPA officials, the network will be largely in place in fiscal year 2003, when templates are to be in place for priority data flows and a large number of the states are expected to be participating.

In addition, EPA officials report that they have taken steps to promote uniform internal implementation of XML. The agency established an XML technical advisory group as a forum for sharing advice and guidance about implementing XML. The group has focused on education and outreach. In addition, EPA officials said they are developing an XML registry to support the agency's Central Data Exchange facility, which they plan to have operational in April 2002.

Department of Defense

Officials in DOD foresee the potential use of XML in many of the department's systems and reported that they are taking action to promote interoperability of these systems and reuse of XML data components, both "vertically" within individual projects and "horizontally" across departmental organizations. Three major efforts—at the Defense Information Systems Agency (DISA), the Defense Logistics Agency, and the Department of the Navy—are focused on standardizing the implementation of XML.

DISA is promoting what officials call a "market-based" approach to standardizing the use of XML. According to this strategy, DISA will provide a central data clearinghouse—including an XML registry of standard data elements, definitions, and structures—where systems developers can come to share data elements and structures that they have developed or to locate existing ones that can meet their needs. The registry is designed to accommodate a number of different levels of compliance for different applications. DISA officials said they have created distinct domains within their clearinghouse where specific DOD communities of interest—such as personnel, finance and accounting, and military intelligence—can define their unique data structures. The agency has already established this data clearinghouse and has defined a management process for collecting, storing and disseminating XML components such as schemas, elements, attributes, DTDs, and style sheets. According to DISA officials, DOD is

considering adopting this clearinghouse, together with the processes for managing it, for use in all departmental systems.

The Defense Logistics Agency's Defense Logistics Information Service, which handles large quantities of information about military logistics, has been developing a repository of data structures related to logistics. According to agency officials, the service established an internal XML working group that initially identified the XML-based data exchange requirements of its customers and developed standard data definitions and structures based on those requirements. Officials said that the service is currently at work identifying its internal needs for an XML registry, evaluating commercial software tools, and assessing how it should interact with external systems, such as DISA's registry.

The Department of the Navy established an XML working group in August 2001 to provide leadership and guidance in maximizing the value of XML across the Navy. According to Navy officials, the group's initial activities have been to develop interim Navy XML policy and prepare an initial Navy XML developer's guide. The developer's guide is currently in draft form and is planned for official release in the first quarter of 2002. The group's goals for the developer's guide are to provide enough specific guidance to developers to ensure that they "move in the right direction," while being general enough to minimize the chance of conflict with future guidance.

Objectives, Scope, and Methodology

Our objectives were to assess (1) the overall development status of XML standards to determine whether they are ready for governmentwide use and (2) challenges faced by the federal government in optimizing its adoption of XML technology to promote broad information sharing and systems interoperability.

To address our objectives, we reviewed documentation and held discussions with representatives from the Chief Information Officers (CIO) Council's XML Working Group and key experts from the private sector, including KPMG, the Logistics Management Institute, and Microsoft Corporation. The XML Working Group is responsible for planning, accelerating, facilitating, and bringing about effective and appropriate implementation of XML technology in the information systems of the federal government. The key experts we contacted from the private sector are actively involved in one or more XML initiatives that may benefit the federal government.

To evaluate the maturity of XML standards for potential governmentwide adoption, we identified and assessed the progress of major nongovernmental standards activities, including those of the W3C, the Organization for the Advancement of Structured Information Standards (OASIS), the United Nations Center for the Facilitation of Procedures and Practices for Administration, Commerce, and Transport (UN/CEFACT), and RosettaNet.

We also held discussions with and reviewed documents from the XML Working Group, GSA, EPA, the National Archives and Records Administration (NARA), the National Institute of Standards and Technology (NIST), DOD, Justice, SEC, and Amtrak. These discussions and documents formed the basis for our assessment of the (1) progress of the federal government in planning and coordinating federal XML initiatives and (2) remaining challenges to be overcome in implementing XML technology throughout the government. In addition, we researched and reviewed documentation on XML prepared by the government of the United Kingdom, the National Electronic Commerce Coordinating Council, and the National Association of State Chief Information Officers.

We performed our review in accordance with generally accepted government auditing standards, working from April 2001 through January 2002, at various locations, including GSA Headquarters in Washington, D.C.; NARA Archives II in College Park, Maryland; and NIST Headquarters in Gaithersburg, Maryland.

Chapter 2: A Comprehensive Set of Standards for Implementing XML Is Only Partially in Place

Key technical standards for XML have been largely worked out under the auspices of the World Wide Web Consortium (W3C). These technical standards are focused on providing the generic structure and tools to tag data, transmit it over the Internet, and allow it to be processed by the computer systems that receive it.

Business standards, though equally important, are generally less well-developed, and reaching agreement on them is proving to be difficult when multiple communities of interest are involved. Business standards are needed to provide a more complete framework for conducting business over the Internet, including advertising products and services so that potential buyers and sellers can find each other, proposing and agreeing upon electronic transactions, and executing the agreed-upon transactions.

Business standards are also needed to define vocabularies for the specific data elements that are to be exchanged when transactions are conducted. These vocabularies, once fully developed, may also be useful to the government in certain cases. However, many of these potentially useful standard vocabularies are still in the initial stages of development and do not provide all the data structures needed to support current government needs.

XML Technical Standards Have Largely Been Defined

The W3C organization has completed development of a suite of core technical standards for XML, as well as a number of functional extensions. As table 4 shows, a number of core technical standards have been approved as official “recommendations” by the W3C.¹ In addition, various functional extensions are currently in development, such as XPointer, which defines how individual parts of a document are addressed; XQuery, which is a language for retrieving and interpreting information from diverse sources; and SOAP (Simple Object Access Protocol), which allows software programs to access and communicate with each other over a network such as the Internet.

¹ In the terminology used by the W3C, a standard is finalized when it is formally approved as a “recommendation.” Earlier versions are termed working drafts, candidate recommendations, and proposed recommendations.

Table 4: XML Technical Standards as of February 2002

Technical standards	Description	Comments
Extensible Markup Language (XML) 1.0	Core standard for XML language.	1st edition approved for implementation February 1998; 2nd edition approved October 2000.
Extensible Stylesheet Language (XSL)	Core standard for formatting XML documents.	V 1.0 approved for implementation, October 2001.
XML Schema	Core standard for specifying the structure, content, and semantics of XML documents.	Approved for implementation, May 2001.
XML Namespaces	Core standard for defining unique identifiers to qualify elements and attributes that may use the same name.	Approved for implementation, January 1999.
Document Object Model (DOM)	Generic method to dynamically access and update structure, content, and style of XML documents.	Level 1 approved October 1998; Level 2, November 2000. Work under way on Level 3.
XML Path Language (XPath)	Syntax to address specific parts of an XML document.	V 1.0 approved, November 1999.
XML Linking Language (XLink)	Language defining how one document links with another document.	V 1.0 approved, June 2001.
Associating Style Sheets with XML Documents	Specification providing a method for associating a style sheet with an XML document.	V 1.0 approved, June 1999.
Canonical XML	Specification describing a method to determine whether two XML documents are identical or whether an application has changed a document.	V 1.0 approved, March 2001.
XML Base	Syntax to define base locations that contain parts of XML documents.	V 1.0 approved, June 2001.
XML Information Set	Set of definitions for use by other specifications that need to refer to information in an XML document.	Approved, October 2001.
XML-Signature Syntax and Processing	Syntax and processing rules for creating and representing digital signatures in XML documents.	Approved, February 2002.

Based on progress to date, W3C technical standards for XML are relatively mature, even though work is still in progress on supplemental standards. Most of the core technical standards were approved within 2 years of being initially proposed, and the fact that commercial products are increasingly being made compatible with XML appears to indicate that the private sector is in general agreement with XML's basic technical infrastructure. For example, vendors providing XML-compatible products include such companies as Ariba, Commerce One, IBM, Mercator, Microsoft, Oracle, Sun, and WebMethods.

Additional Standards Have Been Proposed for Using XML to Conduct Electronic Business

According to industry experts, a suite of business standards beyond XML's technical standards is needed in order to enable organizations that do not have a previously established methodology for data exchange to conduct business and to tap information resources that are meant to be shared. Technical standards provide only the generic structure and tools to tag data and documents, transmit them over the Internet, and process them on the other end. Business standards, in contrast, are needed for two reasons. First, a group of standards is needed to address the overall process of (1) identifying potential business partners for transactions, (2) exchanging precise technical information about the nature of proposed transactions so that the partners can agree to them, and (3) executing agreed-upon transactions in a formal, legally binding manner. In addition to these business process standards, a second group of standards is needed to codify the precise types of data elements that are to be exchanged when a business transaction is conducted. This need is being answered by the development of data vocabularies (or languages) designed to meet the needs of specific businesses and professions.

Business process standards aim to capture electronically all the critical aspects of arranging and conducting a business transaction. For two organizations that have not made detailed arrangements in advance, conducting business transactions over the Internet requires a series of information exchanges that help define proposed transactions in precise terms and then reliably confirm that they have taken place. Individual companies first need to identify each other and share information about the products and services they offer. They must then agree upon which business processes and documents are necessary to carry out a proposed transaction, including determining how the exchange of information will take place and its contractual terms and conditions. Once all this is accomplished, they need to reliably exchange business information, products, and services according to these agreements.

Many of these processes can be captured generically for the activities of most businesses, although there will also be activities that are unique to certain kinds of businesses or certain specialized information exchanges. Examples of specifications that address generic business processes include the following:

- *Electronic business XML (ebXML)* provides a method for companies to exchange business messages and data, conduct transactions, and define and register business processes.
- *RosettaNet* provides vocabularies and business process models (e.g., inventory management and product review) for the electronics industry.

- *Universal Description, Discovery, and Integration (UDDI)* provides directories for Web services description and discovery. Using UDDI, companies can discover each other and define how they will interact and share information over the Internet.

In addition to business process standards, standard data vocabularies (or languages) will be needed for particular industries, professions, and other specific domains. Table 5 shows a representative sample of industry-specific efforts. Hundreds of such projects have been registered with the xml.org Web portal, which serves as a repository for industry XML information.

Table 5: Representative Industry-Specific XML Vocabularies

Vocabulary name	Description
Bioinformatic Sequence Markup Language (BSML)	Supports the encoding and display of DNA, RNA, and protein sequence information.
Chemical Markup Language (CML)	Addresses needs of the chemical industry, such as data tags that can be used to accurately represent chemical formulas.
Extensible Business Reporting Language (XBRL)	Supports financial information, reporting, and analysis.
Geography Markup Language (GML)	Supports the transport and storage of geographic information, including both the geometry and properties of geographic features
HR-XML	Supports human capital management functions such as exchange of staffing data and payroll transactions.
Legal XML	(In development) Will support the legal and legislative profession, especially for electronic court filings.
Mathematical Markup Language	(W3C standard) Facilitates the use and re-use of mathematical and scientific content on the Web.
OpenTravel Alliance	(In development) Will provide a commonly accepted communications process for the travel and transportation industry.
Spacecraft Markup Language (SML)	Provides standard definitions of XML tags and concepts of structure to allow the definition of spacecraft and other support data objects.
Wireless Markup Language (WML)	Facilitates the specification of content and user interface for electronic devices such as cellular phones and pagers.

Business Process Standards Are Less Well-Developed than Technical Standards

Ideally, a well-defined set of XML business process standards covering all key requirements of business data exchanges should be established and universally agreed upon. In conjunction with these basic business standards, individual industries would adopt standard vocabularies to express their unique data types. If agreement on this overall set of standards were achieved, systems developers would have the tools they need to build systems that capitalize on XML's potential to facilitate interoperability. Without such a universally agreed-upon set of standards,

however, XML's use could be limited to carefully prearranged data exchanges with well-established business partners.

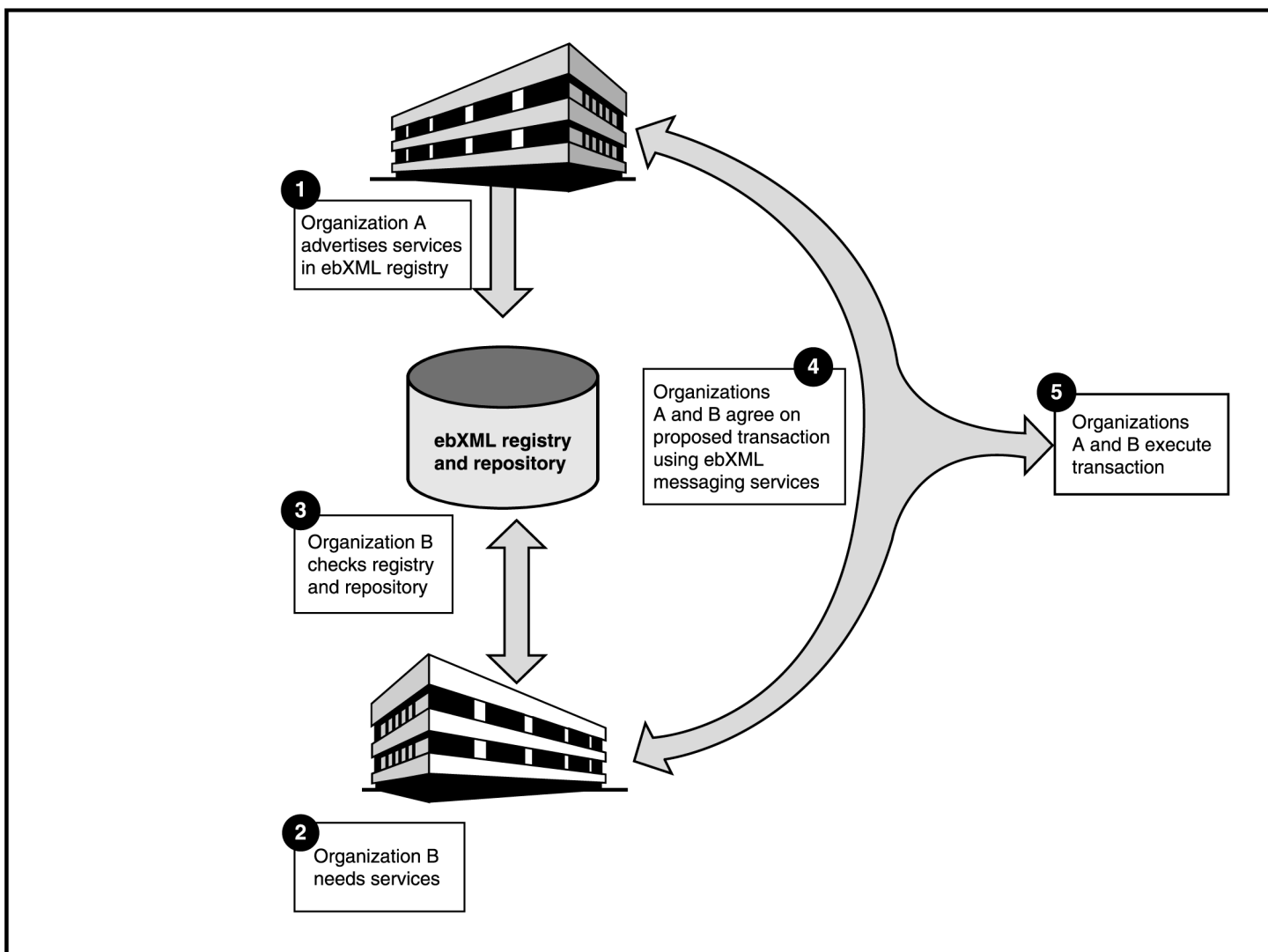
However, business standards are generally less well-developed and agreed upon than XML's core technical standards. Unlike XML technical standards, all of which are established and maintained by the W3C, business standards are developed by a variety of public and private sector organizations, including industry consortia, and are not always universally supported. For example, a number of different approaches to addressing the process of conducting business transactions have been proposed. Currently, at least three of them are vying for support and offer functionality that is in part overlapping and incompatible. These approaches include the following:

ebXML

UN/CEFACT and OASIS have approved a modular suite of ebXML specifications that enables the conduct of business over the Internet.² EbXML's goal is to allow any enterprise—of any size or in any industry—to conduct business electronically with any other entity anywhere in the world. Launched in November 1999, the ebXML project finished its initial development phase in May 2001. At that time, it established a set of design rules for data dictionaries as well as a number of significant reference documents, including a technical architecture, business process specification schema, registry information model, registry services specification, requirements specification, message service standard, and collaboration-protocol profile and agreement. Figure 6 shows a representative ebXML transaction involving two organizations that locate each other through an ebXML registry and then negotiate and carry out the transaction based on ebXML specifications.

² UN/CEFACT is the United Nations' Center for the Facilitation of Procedures and Practices for Administration, Commerce, and Transport. OASIS is the Organization for the Advancement of Structured Information Standards. OASIS is an international nonprofit consortium that promotes open, collaborative development of interoperability specifications to advance electronic business.

Figure 6: Representative ebXML Transaction



Source: GAO.

In public presentations, Office of Management and Budget (OMB) officials have expressed an interest in moving the federal government to greater use of ebXML. In October 2001, OMB defined standards for success in the area of expanding e-government, and ebXML was cited. Specifically, OMB called for federal agencies to “minimize burden on business by re-using

data previously collected or using ebXML or other open standards to receive transmissions.”³

Although many of ebXML’s specifications have been approved, specifications for “core components”—basic data elements and structures that are to serve as common building blocks for use across industries and business processes—are still incomplete. Because different industries often use different terms to refer to the same thing, exchanging information among them can be difficult. Using agreed-upon core components as basic elements for building electronic business messages could reduce the burden involved in getting these divergent systems to interoperate. Software designed to interpret business messages composed of standardized core components would then be able to operate more broadly across industries, thus increasing economies of scale and potentially lowering the cost for small businesses to conduct business electronically.

For example, one component would be an XML data tag structure for “bank account,” which might consist of an account holder’s name and an account number. Such a component would find many uses across a wide range of business activities and industries. Currently, ebXML has published technical reports on the core component methodology and framework, but complete specifications have not yet been defined.

Web Services

Several IT companies are supporting the use of a set of standards for implementing “Web services.” The concept of Web services is to allow businesses with on-line offerings to connect to other businesses to enhance their offerings with functions provided by those other businesses. For example, a company selling merchandise through a Web site could connect to a shipping company to automatically make shipping arrangements and calculate costs for customers. To form these connections, a set of four basic standards has been proposed: XML for representing data, UDDI for locating potential business partners on the Web and identifying services of interest, SOAP for allowing software programs to access and communicate with each other over a network such as the Internet, and Web Services Description Language (WSDL) for

³ Office of Management and Budget, Memorandum M-02-02, *Implementation of the President’s Management Agenda and Presentation of the FY 2003 Budget Request* (October 30, 2001).

describing what specific functions are available and how they can be accessed.

RosettaNet

Funded by a consortium of more than 400 companies, including corporations such as IBM, Cisco, and Dell, RosettaNet began as an effort to create XML standards for the IT supply chain but has expanded to include electronic components and semiconductor manufacturing. RosettaNet has developed three dictionaries: a business dictionary, e-commerce dictionary, and IT technical dictionary. Its business dictionary designates the properties used in basic business activities, and its technical dictionaries provide the properties for defining products. In addition, RosettaNet has developed electronic business guidelines in the form of partner interface process specifications, which include business models, impact and benefit analyses for implementing the business models, technical software designs, and implementation guides. RosettaNet has developed partner interface process specifications for administration, product and service review, product information, order management, inventory management, marketing information management, service and support, and manufacturing. Even though RosettaNet standards were designed for the electronics industry, they offer an approach for defining and modeling business processes that others may follow.

Based on discussions with industry experts and Web documentation, these standards are in different stages of development and acceptance. RosettaNet appears to be the most fully developed business standard, but it is not endorsed by any internationally recognized standards organization. EbXML has the advantage of the formal backing of UN/CEFACT and OASIS, but its suite of specifications is not yet complete. For example, the majority of ebXML's initial efforts focused on establishing the underlying rules for data dictionaries rather than developing the dictionaries themselves. Development began only in October 2001 for a common library of business documents for ebXML that will enable trading partners to unambiguously identify and exchange business information.⁴ Without these tools, data that are exchanged between organizations may not be interpreted and validated consistently.

⁴ In October 2001, OASIS formed the OASIS Universal Business Language (UBL) Technical Committee to define a common XML business document library.

Because uncertainty remains about which business standards will ultimately prevail, applications developed based on any of the current proposals may be at risk of being incompatible with future standards. In addition, without universally accepted standards, commercial IT vendors may use nonstandard XML extensions that could limit interoperability.

Potentially Useful XML Vocabularies Are Not Ready for Governmentwide Adoption

Within the business standards arena, XML is being used to create a variety of “standard” markup languages for particular industries and professions, and many of these languages, once fully developed, may be useful to the government as well. For example, in the future, federal agencies may be able to use HR-XML to exchange data related to human resources functions such as staffing exchange, payroll transactions, compensation, and background checking. Likewise, agencies may be able to use XBRL to format and develop financial statements in the future. And Legal XML could be used to create legal documents such as legislative and court documents. It is the policy of the federal government to use commercial standards whenever practical. However, many potentially useful standard vocabularies are still in the initial stages of development and do not provide all the data structures needed to support current needs. For example, although high-level specifications have been developed in HR-XML for several important human capital functions, very few specific data elements have been specified. Similarly, for XBRL, work has been completed on only one of six planned specifications. For Legal XML, no specifications have yet been completed.

HR-XML is being developed by the HR-XML consortium, a nonprofit group, to allow employers to reduce the ongoing costs of negotiating human-capital-related data exchanges on an ad-hoc basis. The consortium has focused its efforts on developing a suite of high-level specifications for a range of human capital functions, including recruiting and staffing, benefits enrollment, payroll, time and expense reporting, competencies, and background checking. To date, the specifications for all but payroll and background checking have been written. However, the consortium has not fully defined a vocabulary of data tags, DTDs, and schemas for these functions.

XBRL is being developed by XBRL.org, an industrywide consortium, and is intended to be a standards-based electronic language for financial information, reporting, and analysis. In particular, the consortium plans to adapt XBRL to a variety of specific applications, including financial statements, general ledger, regulatory filings, business event reporting, audit schedules, and tax filings. In addition, the consortium plans to

develop taxonomies (common vocabulary) for financial reporting across jurisdictions (e.g., United States, Canada, United Kingdom, and Germany) and taxonomies for specific industries (e.g., mutual funds, media and entertainment, and agriculture). As of this writing, the consortium has completed an XBRL specification for financial statements and a taxonomy for financial reporting of commercial and industrial companies that reflect the generally accepted accounting principles used in the United States. However, work on the other specifications and taxonomies has not been completed, and existing taxonomies for different communities of interest are not completely compatible.

Legal XML is being developed by a nonprofit organization of the same name, made up of volunteers from private industry, nonprofit organizations, government, and academia. The organization seeks to coordinate activities in both the “vertical” and “horizontal” domains of the legal profession. Vertical domains include court filings, transcripts, judicial decisions, and public law. Horizontal domains include general vocabulary and logical document structure. As of this writing, no standards have been completed.

The fact that many of these vocabularies are still in the early stages of development creates challenges for reaching agreement on their use for governmentwide or cross-agency functions. Accordingly, the governmentwide benefits that may be derived from using these standards will not be available in the near term. An apt example is the Human Resources Data Network, being developed by an interagency workgroup to capture essential workforce information to meet the needs of the Office of Personnel Management and other agencies. The planned network is intended to (1) replace the paper-based official personnel folders that are currently used to document pay, benefits, and work history of civilian employees, and (2) serve as a gateway to streamline the process by which agencies provide workforce information to the Office of Personnel Management. According to project officials, the workgroup would like to use commercial standards such as HR-XML to implement the planned network, and officials contacted the HR-XML consortium to assess the applicability of the standard. However, the HR-XML standard is still in early stages of development, with only two approved data definitions (for name and address) currently available. In contrast, the workgroup has completed a data modeling exercise that identified the need to define 984 critical data elements. Unable to wait for commercial standards to be developed, the workgroup defined its own data structure and vocabulary. Project officials noted that even if a fully developed HR-XML vocabulary were available, some of the data elements required by the Human

Resources Data Network likely would not be addressed because they reflect unique government needs.

Chapter 3: The Federal Government Faces Challenges in Realizing XML's Full Potential

Although XML offers the potential to greatly facilitate the identification, integration, and processing of complex information, a number of challenges face the federal government as it attempts to take best advantage of the technology's potential. XML system developers—both within the federal government and externally—must avoid several critical pitfalls when implementing XML, including the risk that data will not be adequately defined and that incompatible data definitions, vocabularies, and structures will proliferate; the potential for proprietary extensions to be built that would defeat XML's goal of broad interoperability; and the need to maintain adequate security.

In addition to these pitfalls, which all systems developers must address, the federal government faces additional challenges as it attempts to gain the most from XML's potential. Specifically, (1) no identifiable governmentwide strategy for XML adoption exists to guide agency implementation efforts and ensure that agency enterprise architectures address adoption of XML. Without agreement on such a strategy, agencies risk building and buying systems that will not work with each other in the future. (2) The needs of federal agencies have not been uniformly identified and consolidated so that they can be represented effectively before key standards-setting bodies. If federal requirements are not better understood and consolidated, the government may be unable to effectively provide input to commercial standards while they are still under development. (3) Although work has begun on a pilot, the government has not yet fully implemented a registry of government-unique XML data structures (such as data element tags) that system developers can consult when building or modifying XML-based systems. (4) Much also needs to be done to ensure that agencies address XML implementation through enterprise architectures so that they can maximize its benefits and forestall costly future reworking of their systems.

Implementing XML Presents Pitfalls

Although XML offers the potential to greatly facilitate the identification, integration, and processing of complex information—both within the federal government and externally—system developers face a number of pitfalls in implementing the technology, including the risk that markup languages, tags, DTDs, and schemas will proliferate; the potential for proprietary extensions to be built that would defeat XML's goal of broad interoperability; and the need to maintain adequate security. Regarding the risk that redundant markup languages, tags, DTDs, and schemas will proliferate, past experience with data interchange has shown that even if a specification such as the XML standard is as complete as possible, individual implementations can vary tremendously. As a result, it is

extremely difficult to get consensus on the definitions of data elements. For example, tags such as <PO_Number>, <PurchaseOrderNumber>, <PO_No>, and <purchase_order_number> could all be used by different applications to indicate a purchase order number. On the other hand, the different tag names could mean that different definitions of "Purchase Order Number" have been used. An XML processor cannot independently determine whether these tags all refer to the same thing. As a result, the processor must be given explicit instructions regarding what tags are equivalent or how to translate one set of tags to the format used by another system.

If diverging data structures and vocabularies proliferate among different organizations and user communities, XML's overarching promise of broad data interoperability could become more difficult to achieve. The use of incompatible data structures would require developers to devote resources to an expensive and error-prone process of defining and implementing translation schemes to exchange information among the incompatible systems.

The processing extensibility of XML can also have a downside, because it allows developers to add proprietary extensions to their specific implementations, which could defeat XML's goal of broad interoperability. It is easy to add elements to an XML document that place unique processing requirements and restrictions on the document, thus preventing other systems from being able to interpret it. An operating system vendor, for example, could add software "hooks" to XML documents that could be correctly processed only by machines running that vendor's particular operating system. The fact that the core XML standard is nonproprietary thus does not ensure that all applications built with it will also successfully interoperate.

Another important challenge in implementing XML is maintaining adequate security. XML's ability to facilitate the direct transfer of data between systems that automatically interpret and process that data has the potential to increase security risks. When XML is used, the direct transfer of data may bypass important security checks, such as those built into intermediate data processing software (virus checkers, for example). For instance, when a site's virus checker examines incoming messages for malicious code, it will not be able to check tagged data embedded in XML documents, unless these data are in American Standard Code for Information Interchange (ASCII) format. The application that then tries to interpret the unchecked XML tags and act on the information could be tricked into processing malicious code, such as a virus. Because XML is

still a relatively new technology, it is unclear how significant this potential vulnerability will be. We were unable to find documented examples of successful intrusions based on this potential vulnerability.

To mitigate this risk, system developers need to ensure that security is addressed when XML-based systems are implemented. For example, measures can be taken to check the integrity of the data received by a computer system, and software can be used to screen the incoming data for malicious code. Likewise, a local store of commonly used DTDs and schemas can be maintained as a check against the integrity of the corresponding DTDs and schemas that come with XML documents from outside sources.

These are a few of the more significant challenges facing XML system implementers. Table 6 summarizes these and other key strengths and pitfalls of XML.

Table 6: Strengths and Pitfalls of XML

Strengths	Pitfalls
XML's flexible, human-readable data tags and structures can be easily adapted to many different needs.	Defining unique data tags and structures can potentially lead to compatibility problems with other systems and defeat the goal of broad-based data exchange.
XML standards are freely available and nonproprietary.	It is easy for vendors and others to build nonstandard extensions to their products and systems, which also could inhibit broad-based data exchange. For example, incompatible business vocabularies have already been developed.
Information in XML documents can potentially be readily accessed and shared among disparate systems.	Increasing access to information that is tagged in human-readable form increases security concerns.
It is easy to search tagged XML data for specific information.	Data that are not highly structured—such as narrative text—may be difficult to convert to XML. Further, converting nontagged information to XML format may require a significant effort without prior agreements and established data dictionaries.
XML uses the nearly ubiquitous existing infrastructure of the Internet.	Using the Internet involves greater security and reliability risks than using private communications links.

The Intellor Group, Inc., conducted a survey on XML benefits and challenges in 2001 and collected 232 responses from many different industries and government agencies.¹ The respondents identified the major benefits of XML as (1) providing a common format that facilitates participation in business-to-business data exchanges, (2) establishing

¹ Intellor Group, Inc., *XML Adoption: Benefits and Challenges* (2001).

common data access techniques, (3) enabling integration of enterprise applications, and (4) achieving cost savings for data conversion. They identified XML's biggest challenges as (1) the immaturity of related standards, (2) the lack of IT staff qualified to develop and maintain XML-based systems, (3) choosing among competing standards, and (4) security for XML documents and XML-based transactions.

Governmentwide Actions to Promote XML Adoption Have Focused on Education and Outreach

To date, activities within the federal government to promote broad governmentwide adoption of XML technology have been limited. Neither OMB, which is responsible for developing and overseeing governmentwide policies and guidelines for agency IT management, nor NIST, which is responsible for developing federal information processing standards and guidelines, have defined an explicit governmentwide strategy for XML adoption to guide agency implementation efforts and ensure that agency enterprise architectures address incorporation of XML. Most governmentwide coordination activities have been performed by the XML Working Group, chartered by the federal CIO Council to facilitate effective and appropriate implementation of XML technology in the information systems of the federal government. The working group's activities have focused primarily on education and outreach. In addition, OMB officials told us that, as part of the annual budget preparation process, they have taken steps to encourage agencies to use XML consistently and share their development plans with other agencies.

Given that the greatest benefits of XML adoption to the government may derive from its promise of facilitating broad interoperability among systems in different organizations, it is important that an explicit strategy be developed for coordinating XML implementation across the federal government's many departments and agencies. However, most XML development within the federal government to date has been undertaken independently by separate federal organizations, with little or no coordination with other agencies. OMB has not issued explicit guidance regarding the use of XML, other than to cite ebXML in its October 2001 standards for success in expanding e-government, as previously discussed. Rather than formulating a specific strategy, OMB has relied on informal discussions with agency officials, as part of the budget preparation process, to encourage them to use XML consistently and share their development plans with other agencies. According to OMB officials, these actions, along with the XML Working Group's coordination activities, serve as the federal government's XML strategy. Further, NIST officials told us they are not planning to develop any federal information processing standards or other XML implementation guidance, which they

do not believe are necessary at this time. However, we believe that, without a well-defined strategy, the government runs the risk that incompatible data formats and standards will proliferate and prevent agencies from being able to take full advantage of XML to substantially improve governmentwide data sharing.

The XML Working Group was chartered by the CIO Council in September 2000 to (1) identify pertinent standards and best practices, (2) establish partnerships with industry and public interest groups, (3) establish partnerships with governmental communities of interest, and (4) promote education and outreach. In addition, in its strategic plan for fiscal year 2001–2002, the CIO Council tasked the working group to use its Web site—xml.gov—to lay out an evolving strategy with specific tasks for the working group to undertake to promote the effective and well-coordinated usage of XML to support governmental functions.

The XML Working Group has undertaken a number of education and outreach efforts, including (1) holding monthly meetings as a forum for presentations and discussions about XML-related topics, (2) establishing the xml.gov Web site for information sharing and dissemination, and (3) exploring opportunities for coordination with state governments.

As part of its effort to promote education and outreach, the working group holds monthly meetings to hear presentations and engage in discussions on XML-related topics. The meeting minutes, presentations, and information on other XML-related activities are shared and disseminated via the xml.gov Web site, as well as an electronic mailing list. In addition, agencies choosing to share information about their XML efforts can do so by registering with the working group, which then posts information about each effort on its Web site. To further promote their activities, working group officials met with state CIOs to explore opportunities to engage the states more effectively in the group's activities.

In an effort to identify best practices for XML adoption, the CIO Council issued, in January 2001, a call for all federal CIOs to participate in developing and improving the design and content of the xml.gov Web site. In addition, the CIOs were encouraged to register their agencies' XML-related activities, especially those that cut across communities of interest. As of December 2001, representatives from 24 projects and working groups at the federal, state, and nonprofit levels had registered their XML-related efforts. However, according to the co-chair of the XML Working Group, there were likely many other federal activities under way that had

not been registered. For example, the XML projects at Justice and SEC cited previously had not been registered at that time.

On its Web site, the XML Working Group noted that in developing an evolving strategy for the effective usage of XML, it faced a number of constraints and conditions, including very limited resources and the fact that it is not a policy-making body and has no operational responsibilities. According to a statement at xml.gov, the Web site itself is intended to be the embodiment of the working group's strategic plan. Because of the working group's constraints, the Web site does not provide specific guidance to agencies for implementing XML, participating in XML standards bodies, or incorporating XML requirements into enterprise architectures.

NIST, along with GSA, has developed a Web-based standards road map, to provide users with access to information regarding existing and emerging XML standards and activities related to electronic commerce. The standards road map allows users to identify standards information relevant to their individual projects and assess the applicability, maturity, and product availability associated with those activities. The tool can be accessed from the XML Working Group Web site or at www.nist.gov/roadmap. Although the standards road map has the potential to be a useful tool for promoting systems interoperability, it is still a work in progress because the standards are rapidly evolving. For example, technical specifications for UDDI are currently not in the standards road map.

OMB officials told us that, as part of the annual budget preparation process, they have taken steps to encourage agencies to use XML consistently and share their development plans with other agencies. Specifically, according to the OMB officials, federal agencies that request funding for XML-based initiatives are instructed to (1) determine whether an implementation approach has already been developed in private industry that can be emulated to meet the agency's needs, and (2) submit their activities for listing on the xml.gov Web site so that other agencies can be made aware of their plans. Further, OMB officials said they discuss with agency officials the importance of updating sections of the agency's enterprise architecture—specifically the standards profile and technical reference model—to reflect their XML plans. As discussed previously, OMB has established a standard for success in the area of expanding e-government by calling for agencies to minimize burden on business by re-using data previously collected or using ebXML or other open standards to

receive transmissions. The agency has also begun using XML for its own databases of federal IT management information.

Federal Government Needs Have Not Been Consolidated for Input to Standards- Setting Bodies

Several federal agencies are working individually with key industry and public interest groups to incorporate their unique requirements into standards and specifications as they are being developed. Specifically, officials from OMB, NIST, DISA, and GSA have each participated in one or more XML-related standards activities. However, no central focal point has been designated to identify cross-agency or governmentwide requirements for standard XML data structures or develop a dictionary of inherently governmental data tags. Further, no process has been implemented for consolidated collaboration with standards bodies on the development of XML standards and specifications to ensure that federal requirements are identified and incorporated. Past experience coordinating federal requirements for EDI suggests that one approach to resolving the problem would be to present a "single face to industry" through a single requirements coordinating committee.

Based on individual agency initiative, several federal agencies are participating in standards initiatives led by organizations such as the American National Standards Institute (ANSI),² UN/CEFACT, OASIS, and RosettaNet. For example, NIST is a member of OASIS and RosettaNet and has actively participated in the development of test suites to assess conformance with XML standards. NIST chairs several OASIS technical committees to influence the quality, correctness, and testability of ebXML specifications. In addition, NIST developed conformance test suites based on XML standards and submitted them to OASIS for the benefit of the entire community. Further, NIST co-sponsored a forum with ANSI in October 2001 to explore alternatives for using XML to improve ANSI's standards-setting process. GSA has also been active in standards setting by serving as a board member of the RosettaNet initiative. In addition, GSA officials, including the co-chair of the XML Working Group, have been actively participating in the development of ebXML standards at UN/CEFACT and OASIS. Also, OMB officials told us they were working with international organizations on trade-related standards.

² ANSI is a private, nonprofit organization that administers and coordinates the U.S. voluntary standardization and conformity assessment system.

DISA participates in various standards bodies and consortiums, including ANSI, UN/CEFACT, OASIS, W3C, the Internet Engineering Task Force, and others. The agency has contributed to the development of the ebXML standards suite and has applied ebXML to its own electronic business processes. In addition, DISA is a member of the W3C Advisory Committee and coordinates with the Defense Logistics Agency in the development of W3C XML standards.

Although these are valuable undertakings, none is specifically designed to serve the role of presenting unified federal requirements to standards bodies. The government's business processes are not necessarily the same as the private sector's, and in many cases government agencies may need to define unique data types and structures. The need for a defined set of inherently governmental data tags was highlighted in a recent study conducted by the Logistics Management Institute for GSA.³ The Institute was tasked to (1) identify the data elements associated with 22 commonly used government forms and (2) determine if those data elements were available in commercial registries. The study identified over 8,000 data elements in the 22 specified forms. The study's final report stated that an intensive review of a subset of these elements found that for a very large number of them, no corresponding entry in any of the commercial registries was found. The Logistics Management Institute concluded that because existing commercial registries did not focus on many of the government's business processes, the government would need to develop new dictionaries of data tags, in concert with industry and the public, to meet its needs.

Although similar needs for coordination have been successfully addressed in the past, the federal government does not have a process for providing consolidated input on XML to commercial standards bodies. Instead, OMB has allowed agencies to individually pursue participation in standards bodies to the extent that their interests and resources allow. As a result, participation has been limited and uncoordinated because it requires a commitment of staff resources that many agencies cannot afford, according to XML Working Group officials. OMB guidelines⁴ direct

³ Mark Crawford, Donald F. Egan, and Angela Jackson, *Federal Tag Standards for Extensible Markup Language*, Logistics Management Institute (June 2001).

⁴ Office of Management and Budget, Circular A-119, *Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities* (February 10, 1998).

agencies to use voluntary consensus standards in lieu of government-unique standards, except where inconsistent with law or otherwise impractical. The guidelines also address agency participation in voluntary consensus standards bodies and describe procedures for satisfying the reporting requirements of the National Technology Transfer and Advancement Act of 1995 (Public Law 104-113).

In the case of EDI, the federal government presented a “single face to industry” by chartering a Federal EDI Standards Management Coordinating Committee. The committee’s objectives were to (1) adopt governmentwide EDI standards for implementation, (2) coordinate federal agency participation in EDI standards bodies to ensure adequate consideration of the government’s business needs and to ensure consistency of position (thus presenting a “single face” to industry), and (3) share EDI information among agencies regarding current or planned implementations to avoid duplicate efforts and to streamline the process.⁵ As a result of the committee’s work, a number of larger federal agencies are now successfully using EDI to conduct electronic business with established business partners.

XML Interoperability across the Government Depends on an Effective Cross- Agency Registry

Systems developers in the federal government would benefit from the establishment of an XML registry, which they could consult to identify and obtain predefined data elements and structures that are already in use. The XML Working Group is in the process of building a pilot version of such a registry. However, the registry will be effective in supporting systems interoperability among federal agencies only if governmentwide policies are set, guidelines established, and a defined management and funding process put in place for operating the registry.

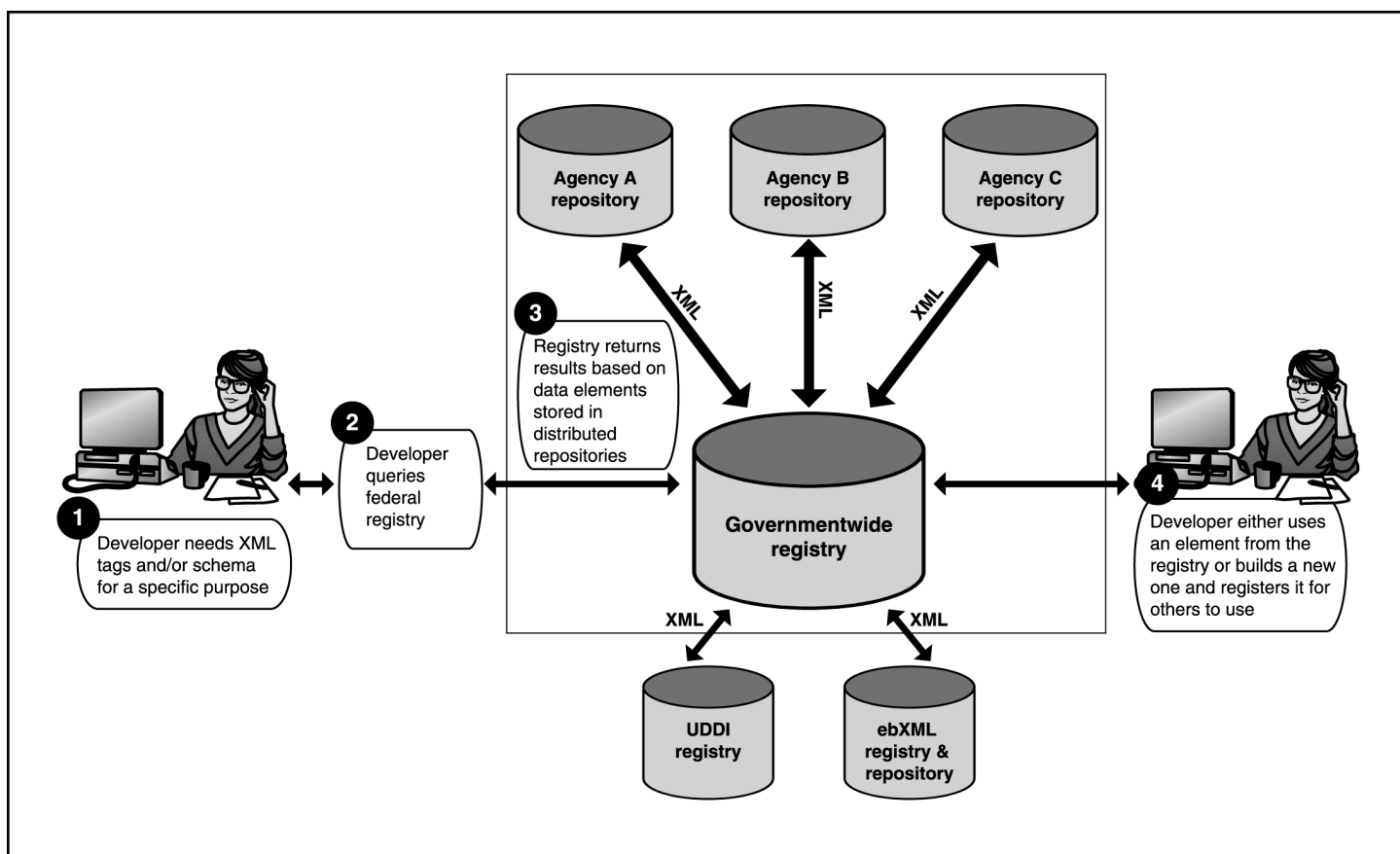
In contrast to the “top down” approach of defining and mandating the use of specific data structures or vocabularies, a “bottom up” approach is to establish a centralized registry of XML components—including data elements, DTDs, and schemas—and coordinate its use by XML systems developers. Under this arrangement, XML developers would be encouraged to submit data elements and structures used in their systems for inclusion in the registry. Other developers would then be able to look up these structures in the registry and incorporate them, as appropriate, into their own systems. Developers would have the incentive to reuse data

⁵ This process is described in FIPS Publication 161-2, *Electronic Data Interchange (EDI)*.

structures found in the registry because doing so would save costs and also bring about interoperability with other existing systems. The more widely specific data elements and structures were used, the closer they would come to becoming de facto standards.

A centralized registry would not necessarily include only a single option to address a specific business need. Overlapping variants of some types of tags, definitions, and data structures may be needed to address the needs of different communities. For example, a standard schema for military purchase orders might differ from a purchase order schema shared by a group of civilian agencies. Further, a government registry could link to a number of standard commercial variants defined for other communities of interest that may contain additional purchase order schemas used by specific industries. The chemical and automotive industries, for example, may use schemas that vary from each other as well as from the standard government version. A registry would provide access and information about all relevant predefined data definitions and structures, which would allow developers to make decisions about the extent they needed to adhere strictly to industry standards, government standards, or some combination. Figure 7 summarizes how an XML developer could hypothetically use an XML registry.

Figure 7: Using a Registry of XML Data Elements and Structures



Source: GAO.

Although no registry of “inherently governmental” XML components has yet been established, work is under way to create a pilot version of a registry. According to XML Working Group officials, NIST has developed a specification of the functional requirements for the pilot registry, and the working group’s leaders have determined that they can use a version of the system developed by the Defense Logistics Information Service to satisfy these requirements. No date has yet been set for putting the pilot registry into initial operation.

According to the co-chair of the XML Working Group, a governmentwide registry can provide users with the ability to (1) discover and use pertinent XML components and (2) register additional components that are “inherently governmental” in nature if those already specified in

commercial registries do not meet the users' requirements. With a registry in place, agencies could start using registered XML components, and de facto XML standards would thus begin to emerge within specific communities of interest. Under these circumstances, the CIO Council or OMB would be in a better position to define specific governmentwide standards at a later time, based in part on this activity.

However, a government XML registry will be effective in supporting systems interoperability among federal agencies only if governmentwide policies are set, guidelines established, and a defined management and funding process put in place to operate the registry. Work on defining exactly how an operational governmentwide registry—and the data repositories associated with it—should be administered and maintained is not yet complete. The XML Working Group has recently established a subgroup to define registry-related policies and procedures. However, it has not yet defined a management process that specifies (1) who is allowed to register new XML components, (2) how input to the registry is to be verified, (3) to what extent developers will be required to consult the registry when building new XML data structures, (4) classes of compliance for categorizing how rigorously organizations adhere to the standard data structures and definitions, or (5) a configuration management process to keep track of successive versions of each registered component. Members of the group drafted an XML Developer's Guide in December 2001 that includes a proposed requirement that agency XML developers make use of the federal registry, but the draft guide has not yet been approved and adopted.

Standard conventions for using XML's namespace feature and other rules for naming data elements, DTDs, and schemas in a consistent and unambiguous way have not yet been defined for the pilot registry. Without such a naming structure, different XML documents may use the same data tags for different definitions and structures. A standard use of the namespace feature would allow the tags in any given XML document to be traced back unambiguously to their proper definitions.

The registry's management framework would also need to include definitions of different classes of compliance with the registry's data structures. In some cases, individual agency implementations may not need to be integrated with other government systems, and agencies may have compelling reasons to develop nonstandard data structures. The establishment of different classes of compliance would define how loosely or tightly an XML implementation would be connected to the registry and would outline the operational implications associated with each class.

Once management policies and procedures are established, funding mechanisms will also be needed to support ongoing operation of the governmentwide registry. According to industry and XML Working Group officials, registry projects in the private sector to date have required significant commitments of resources. Thus it would be important to assess and plan for the expected costs of such an undertaking.

XML Implementations Can Be More Effective within the Context of an Enterprise Architecture

Planning the effective use of a standard such as XML to promote data interoperability is part of the larger process of establishing and implementing an enterprise architecture.⁶ According to the CIO Council,⁷ an enterprise architecture establishes an agencywide roadmap to achieve an agency's mission through optimal performance of its core business processes within an efficient IT environment. Data, as a corporate asset, are key to an agency's vision, mission, goals, and daily work routine. The more efficiently an agency gathers, stores, uses, and protects data, the more productive it is. Thus, one of the major goals in developing an architecture is to minimize the burden of data collection, streamline data storage, and enhance data access. Planning XML usage within the context of an agency's enterprise architecture can contribute significantly to achieving this objective.

A major component of an enterprise architecture is a standards profile, which defines the set of rules that governs systems implementation and operation. If agencies have a business need for XML, then the standards profile should be used to document the way in which XML standards and products will be used.

Without an effort to build an enterprise architecture, including the underlying data architecture, implementing XML is likely to provide only a patchwork solution to systems interoperability. Typically, if multiple systems have been developed independently and without an overall architecture, they are likely to use many data element definitions and structures that overlap in function or are completely redundant. In addition, secondary or tertiary data elements—data that do not represent

⁶ An enterprise architecture is an institutional systems blueprint that defines in both business and technology terms the organization's current and target operating environments and provides a road map for moving between the two.

⁷ Chief Information Officers Council, *A Practical Guide to Federal Enterprise Architecture*, Version 1.0 (February 2001).

discrete information but are merely the calculated derivatives of primary data elements—are also likely to proliferate. If XML is simply added on to “glue” these systems together, the organization will have to carry the burden of maintaining many more data elements and definitions than are necessary, as well as all the translations needed to effectively pass data among the systems.

We have recommended that an organization’s data needs be assessed as a whole and an architecture defined that includes a core set of critical data elements and structures. Redundant elements, as well as secondary and tertiary elements, can then be eliminated, saving the organization the expense of maintaining them. XML can then be implemented more efficiently, with fewer translations required between elements that have different names but refer to the same thing. The organization will also be better prepared to define interfaces to external systems and data sources. According to a National Electronic Commerce Coordinating Council report,⁸ applying XML within government can yield greater benefits if agencies take the initial step of inventorying common data exchanges.

As with any element of an IT infrastructure, security issues need to be identified and addressed when XML is being implemented. As previously discussed, XML documents potentially could be used to transport malicious code—such as viruses and worms—into an agency’s computer systems, because virus checkers do not always examine the content of XML documents. System design documents will need to include plans to compensate for this and other potential vulnerabilities.

⁸ National Electronic Commerce Coordinating Council, *An Introduction to XML's Potential Use within Government* (December 2000).

Chapter 4: Conclusions and Recommendations

Conclusions

XML has the potential to help the federal government significantly streamline the process of identifying, integrating, and processing information from widely dispersed systems and organizations. Many critical government functions depend on effective information sharing across organizational boundaries, yet the problem of overcoming obstacles to effective data sharing has never been satisfactorily resolved. Today, broad information sharing needs are at the forefront of national priorities. For example, identifying and countering a bioterrorist attack requires that important medical information be collected and integrated as rapidly and thoroughly as possible. Likewise, law enforcement information about known terrorists and their activities must also be integrated and shared at Internet speed. XML-based systems can play a valuable part in facilitating this kind of broad information exchange.

XML's greatest benefits accrue when organizations, such as government agencies, use standard data exchange procedures and agree on standard data definitions and structures. Effectively using XML as a means to share data among disparate systems across the federal government will require agencies to conform to a range of technical and business standards. While XML's technical standards are largely in place, important business standards—including many planned standard vocabularies—have not yet been completed, and in some cases, standards development to date has resulted in incompatibilities. To the extent that these business standards address government needs as they are developed, government agencies will likely have less of a need to develop their own nonstandard data vocabularies and structures.

Given that a complete set of XML-related standards is not yet available, system developers must be wary of several pitfalls associated with implementing XML that could limit its potential to facilitate broad information exchange or adversely affect interoperability, including (1) the risk that redundant data definitions, vocabularies, and structures will proliferate, (2) the potential for proprietary extensions to be built that would defeat XML's goal of broad interoperability, and (3) the need to maintain adequate security.

While education and outreach are important activities that are already under way in the federal government, an explicit strategy for adopting XML across the government has not yet been defined. Such a strategy is an important foundation for promoting standardization across agencies and facilitating broad information exchange while at the same time reserving the flexibility for agencies to tailor their use of XML to best meet their needs. Without a well-defined strategy, the government runs the risk that

incompatible data formats and standards will proliferate and prevent agencies from being able to take full advantage of XML to substantially improve governmentwide data sharing.

The federal government, which is committed to adopting commercial standards wherever possible, still has the opportunity to have its needs considered in the process of developing these standards. However, federal requirements have not yet been identified and consolidated so that they can be clearly communicated to the standards bodies that are currently at work on XML business standards.

Given that XML is still in the early stages of its development and implementation, a top down strategy of predefining XML data structures and designating specific commercial standards, such as ebXML, as universal solutions for addressing interoperability is not likely to be effective. Instead, to be effective, the government's strategy must balance top down guidance with bottom up incentives that encourage agency initiative and provide leeway for agencies to develop implementations that best meet their needs. Specifically, establishing an operational registry for XML data elements and structures with incentives for agencies to make use of it could encourage a bottom up development of de facto standards. As elements of a government XML vocabulary became standardized through this registry on a de facto basis, the government would be in a better position at a later date to revisit the question of what commercial standards and vocabularies to officially endorse. The XML Working Group is developing a pilot registry along these lines, but it is not yet operational and lacks an agreed-upon set of policies and guidelines to promote the broadest possible use.

XML's larger promise of facilitating data exchange across broad domains (such as an entire agency, a group of agencies, or a set of external stakeholders and client organizations) will be difficult to realize until critical data elements and structures are identified and standardized across entire agencies and communities of interest. This task of identifying and standardizing critical data elements and structures is part of an agency's larger task of developing an enterprise architecture. Well-planned enterprise architectures can also promote the adoption of flexible implementations that can be modified in the future to conform to commercial standards that become established over time. Thus, agency enterprise architectures are key building blocks to effective governmentwide adoption of XML.

Recommendations for Executive Action

Given the statutory responsibility of OMB to develop and oversee governmentwide policies and guidelines for agency IT management, we recommend that the director of OMB, working in concert with the federal CIO Council and NIST, develop a strategy for governmentwide adoption of XML to guide agency implementation efforts and ensure that the technology is addressed in agency enterprise architectures. This strategy should, at a minimum, address how the federal government will address the following tasks:

- Developing a process with defined roles, responsibilities, and accountability for identifying and coordinating government-unique requirements and presenting consolidated, focused input to private sector standards-setting bodies during the development of XML standards. This process could be patterned after the current process that is in place for EDI coordination among federal agencies, or OMB might consider adapting the EDI process to cover XML as well. Guiding the overall process should be the presumption that mature, agreed-upon commercial standards will be adopted by the government whenever possible.
- Developing a project plan for transitioning the CIO Council's pilot XML registry effort into an operational governmentwide resource. This plan should include identifying time frames and resources needed to implement and maintain an operational registry linked to agency repositories of standard data structures.
- Setting policies and guidelines for managing and participating in the governmentwide XML registry, once it is operational, to ensure its effectiveness in promoting data sharing capabilities among federal agencies. These policies should clarify the roles and responsibilities of specific agencies and should consider including definitions of classes of compliance, which could be used to categorize how rigorously organizations adhere to the policies. Further, these policies should promote the consistent use of XML namespaces to resolve potential ambiguity in data references across XML documents.

In addition, as part of its ongoing process for reviewing agency IT architectures and annual budget requests, we recommend that OMB ensure that agencies' business needs for XML technology are defined in their enterprise architectures. Specifically, OMB should specify requirements for documenting the usage of XML standards and products in the standards profile section of the architecture—the section that defines the set of rules governing systems implementation and operation.

Agency Comments and Our Evaluation

In oral comments on a draft of this report, officials from OMB's Office of Information and Regulatory Affairs, including the Information Policy and Technology Branch chief, generally agreed with our findings and conclusions and stated that they would consider our recommendations. The officials also provided information on recent OMB actions aimed at promoting the adoption of XML by federal agencies. We have incorporated this updated information in the report. We view these recent OMB actions as positive steps. Nevertheless, we also believe that OMB can improve on these actions by implementing the recommendations in this report.

We received oral comments from the co-chairmen of the XML Working Group; officials of NIST's Information Technology Laboratory; and the deputy associate administrator, Office of Electronic Commerce, GSA. We also received written comments from the chief information officer, National Aeronautics and Space Administration; and the director for policy and communications staff, National Archives and Records Administration. Letters from these latter two agencies are reprinted in appendixes I and II. All of the agency officials who reviewed the draft agreed with the overall content of the report. Officials from the XML Working Group and the National Archives and Records Administration expressed concern that the draft overemphasized the value of a "top down" XML implementation strategy that emphasizes executive direction and guidance as opposed to a "bottom up" approach relying on individual initiative at lower management levels. We believe that it is important to strike a balance between the two approaches. In response to this concern, we are including language in the final report to emphasize that a balance between the bottom up and top down approaches is needed. In addition, each agency provided technical comments, which have been addressed where appropriate in the final report.

Appendix I: Comments from the National Aeronautics and Space Administration

National Aeronautics and
Space Administration
Office of the Administrator
Washington, DC 20546-0001



March 18, 2002

Mr. John A. de Ferrari
Assistant Director
U.S. General Accounting Office
441 G Street, NW, Room 4T21
Washington, DC 20548

Dear Mr. De Ferrari:

Thank you for the opportunity to comment on the draft GAO report, "Electronic Government: Challenges to Effective Adoption of the Extensible Markup Language." The report is quite comprehensive and effectively communicates the history, potential benefits, and challenges of adopting the Extensible Markup Language (XML).

The draft report clearly demonstrates that XML, as contrasted with other emerging technologies, presents virtually unique challenges in that its effective use requires the convergence of both technical and business standards, and the business standards span virtually all segments of the private sector and government. In the case of most other technologies, the standards battles are usually fought at the technical level and are much less dependent on the vocabulary and business processes of potential industry and government users. In the case of XML, the World Wide Web Consortium (W3C) has worked out the technical standards, but each segment of the private sector is struggling through the process of developing its XML business standards. Since this process requires the cooperation of competitors, the final products are difficult to achieve and long in coming. In some areas key to the performance of the government, because the private sector is proceeding slowly or does not have requirements, the government, working cooperatively with the private sector, should take the lead in defining the government-unique business standards.

To date, individual government departments and agencies (as documented in your draft report) have begun using XML based on a tradeoff of the benefits of its use in an incomplete business standards environment, versus the risk that their implementations will have to be redone to conform to business standards that are eventually finalized by the private sector segments with whom they interact. Given the current status of XML standards, this seems to be a rational approach. Therefore, while there is benefit to formalizing a government-wide strategy for adoption of XML along the lines described in the draft report, until XML business standards are much further along towards finalization, for the foreseeable future individual government entities will likely have to

2

continue with the same risk assessment and trade-off approach in their implementations of XML.

Implementing the elements of the XML strategy described in the draft report would help drive successful adoption of this technology across the government, but, to be effective, would require a significant commitment of new resources to groups such as the CIO Council XML Working Group, and should not be undertaken unless those resources are provided.

Please contact Mr. Robert Benedict at (202) 358-1475 or at robert.benedict@hq.nasa.gov for questions on or clarification of these comments.

Cordially yours,



Lee B. Holcomb
Chief Information Officer

Appendix II: Comments from the National Archives and Records Administration



National Archives and Records Administration

8601 Adelphi Road
College Park, Maryland 20740-6001

March 14, 2002

John A. de Ferrari
Assistant Director
U.S. General Accounting Office
441 G Street, N.W. Room 4T21
Washington, D.C. 20548

Dear Mr. De Ferrari:

The National Archives and Records Administration (NARA) appreciates the opportunity to review and comment on the draft GAO report, "Electronic Government: Challenges to Effective Adoption of the Extensible Markup Language." We believe that the report accurately describes the present state of XML in the Federal Government.

NARA strongly supports use of XML by the Federal Government. Indeed, our Electronic Records Archives (ERA) project will have XML as one of the building blocks to provide a dynamic solution that incorporates the expectation of continuing change in information technology and in the records it produces. We suggest that you include NARA's ERA project in your examples of agencies that are using XML in the section that begins on page 32. Beyond the ERA project, we suggest that GAO could also emphasize the use of XML in records management and recordkeeping for agencies.

We appreciate GAO's recognition that there will be multiple Government XML registries/repositories. NARA will be working with the XML Working Group on the development of their pilot registry and will have a robust registry/repository as part of ERA. We may be the appropriate agency to host the cross-agency centralized registry.

Finally, we strongly support the "bottom up" communities of interest approach to implementing XML in the Government. Although the draft report discusses this approach, it appears that you prefer the "top down" approach used to develop EDI. We believe that agencies should determine which approach better addresses their needs.

Thank you for the opportunity to provide these comments. If you have any questions about our comments, please contact me.

Sincerely,

A handwritten signature in black ink, appearing to read "Lori A. Lisowski".

Lori A. Lisowski
Director
Policy and Communications Staff

NARA's web site is <http://www.nara.gov>

Glossary

Application Programming Interface	The interface between the application software and the application platform (i.e., operating system), across which all services are provided.
Attribute	A property associated with a specific data element in an XML document.
Business Process	A collection of related, structured activities—a chain of events—that produce a specific service or product for a particular customer or customers.
Collaboration Protocol Agreement	Information that identifies or describes the specific collaboration protocol that two (or more) parties have agreed to use.
Collaboration Protocol Profile	Information about a party that describes one or more business processes and associated protocols that the party supports for purposes of collaboration.
Data Type	A description of the attributes of a specific set of data, such as whether it represents integers or text strings.
Document Type Definition (DTD)	A file that describes the structure of XML documents and defines how markup tags should be interpreted. A DTD can be used to automatically interpret multiple documents in a uniform way.
Electronic Business	The exchange of information within or among enterprises by electronic means for the purpose of conducting business transactions or other related activities.
Electronic Commerce	Business done electronically, including the sharing of standardized unstructured or structured business information by any electronic means.
Electronic Data Interchange (EDI)	The automated exchange of predefined and structured business data among information systems of two or more organizations. Federal government use of EDI is governed by Federal Information Processing Standard 161-2.

Electronic Government	Government's use of technology, particularly Web-based applications, to enhance the access to and delivery of government information and services to citizens, business partners, employees, other agencies, and government entities.
Encryption	Cryptographic transformation of data (called "plaintext") into a form (called "ciphertext") that conceals the data's original meaning to prevent it from being known or used.
Enterprise Architecture	An institutional systems blueprint that defines in both business and technology terms an organization's current and target operating environments and provides a road map for moving between the two.
Extensible Markup Language (XML)	A flexible, nonproprietary set of standards for tagging information so that it can be transmitted using Internet protocols and readily interpreted by disparate computer systems.
Extensible Stylesheet Language (XSL)	A language used to transform XML-based data into HTML or other presentation formats for display in a variety of media.
Hypertext Markup Language (HTML)	The standard markup language used to display information on the Web. It uses tags embedded in text files to encode instructions for formatting and displaying the information.
Interoperability	The ability of two or more systems or components to exchange information and to use the information that has been exchanged.
Markup	The addition of tags or labels to data elements in a document to provide processing instructions or to indicate structure or meaning.
Metadata	Data containing descriptive information about other data. For example, a block of numerical data might be identified in metadata as representing unit cost in dollars.

Namespace	A unique identifier, such as a Web address, referenced at the start of an XML document as a source for definitions of the tags and other data structures used in the document. An XML document can reference more than one namespace.
Parser	Software that reads an XML document and determines the structure and properties of the data in the document.
Registry	An electronic listing of specifications—such as DTDs, XML schemas, and the metadata about them—as well as pointers to their locations (called repositories).
Repository	A location or set of distributed locations where registry items reside and from which they can be retrieved and used in conjunction with marked up documents, such as XML documents.
Schema	A set of custom tags and attributes that defines the permissible tagging structure for an XML document and conforms to the W3C Schema specification.
Search Engine	A program that searches documents for specified keywords and returns a list of the documents where the keywords are found.
Style Sheet	A text file that provides instructions for formatting and displaying the information in XML documents. Style sheets can include variations depending on the type of device used to access the document. For example, the same XML document could be displayed differently on a handheld wireless computer or a desktop computer, based on different style sheets.
Valid XML Document	An XML document that has an associated document type declaration and that complies with the specifications expressed in it.

Well-formed XML Document	An XML document that conforms to the W3C XML specification.
XML Document	A text document marked up with hierarchically arranged descriptive tags and attributes. An XML document can also begin with declarations that refer to other files providing further instructions for interpreting and displaying data elements.
XML Path Language (XPath)	A language for referencing specific parts of an XML document.
XML Processor	A software module used to read XML documents and give applications access to their content and structure. Validating processors also identify discrepancies with the XML 1.0 standard and the constraints expressed in DTDs and external entities referenced in an XML document.
XSL Transformation (XSLT)	An extension to the XSL standard that provides commands to transform one XML document into either another XML document or a different format, such as HTML.

GAO's Mission

The General Accounting Office, the investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents is through the Internet. GAO's Web site (www.gao.gov) contains abstracts and full-text files of current reports and testimony and an expanding archive of older products. The Web site features a search engine to help you locate documents using key words and phrases. You can print these documents in their entirety, including charts and other graphics.

Each day, GAO issues a list of newly released reports, testimony, and correspondence. GAO posts this list, known as "Today's Reports," on its Web site daily. The list contains links to the full-text document files. To have GAO e-mail this list to you every afternoon, go to www.gao.gov and select "Subscribe to daily e-mail alert for newly released products" under the GAO Reports heading.

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are \$2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. General Accounting Office
P.O. Box 37050
Washington, D.C. 20013

To order by Phone: Voice: (202) 512-6000
 TDD: (202) 512-2537
 Fax: (202) 512-6061

Visit GAO's Document Distribution Center

GAO Building
Room 1100, 700 4th Street, NW (corner of 4th and G Streets, NW)
Washington, D.C. 20013

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm,
E-mail: fraudnet@gao.gov, or
1-800-424-5454 or (202) 512-7470 (automated answering system).

Public Affairs

Jeff Nelligan, Managing Director, NelliganJ@gao.gov (202) 512-4800
U.S. General Accounting Office, 441 G. Street NW, Room 7149,
Washington, D.C. 20548