Domain: Information

DESCRIPTION

The Information Domain addresses standards and guidelines for:

- Data Interoperability
- Data Management
- Data Formats
- Records Management (TBD)

A process-independent, enterprise view of government information enables data sharing where appropriate within the bounds of security and privacy considerations. Service oriented architectures promote information and service reuse through open standards.

To help the Commonwealth achieve the enormous benefits of information and service reuse, the Information Domain emphasizes standards for data interoperability among diverse internal and external platforms and applications. By promoting the ubiquitous use of XML standards, the ETRM specifications insure that all new development initiatives result in interoperable services that can be reused across the enterprise, as well as with external business partners and governments where appropriate.

Given the level of complexity of integration projects, especially with multiple developers and teams collaborating on the development of services, data models should be explicitly visible to all architects, developers, and project managers as a coherent set of XML schemas, in a Commonwealth Registry, and service development should be driven by those schemas.

Initiatives such as Homeland Security rely upon all parties adhering to Community of Interest XML specifications, defined by open standards bodies comprised of representatives from Government, Business and Technology Communities. Open formats for data files ensure that government records remain independent of underlying systems and applications thereby preserving their accessibility over very long periods of time.

STRATEGIC IMPORTANCE

Return on investment in IT assets is greatly improved by the ability to reuse information and services based on open standards. When information and data is viewed as a Commonwealth strategic asset and resource, it can improve state government's ability to serve its constituents, to improve its stewardship of public records currently and in the future, and to consistently apply appropriate privacy and security protections to information no matter where that information is held. Better data interoperability and management will foster better IT governance, while also improving the quality and accessibility of information and services.

RELATED TRENDS

- Customer-centric approaches to information management leverage data across organizational boundaries to give a comprehensive view of the organization's interactions with that customer

- Information classification is being used at the enterprise level to assign appropriate and consistent levels of sensitivity and security across the various organizational boundaries

- Data that is common to many business processes are being shared and re-used within the constraints of privacy and security considerations

- As records move from paper to electronic formats there is an increasing need for electronic records management and conservation policies and systems.

VISION

Information is no longer viewed as an exclusive agency asset but is leveraged and re-used throughout the enterprise while observing appropriate privacy and security protections. Electronic records are preserved in open formats that allow for optimal electronic records conservation and availability to the public over long periods of time.

ROADMAP

Current State

- Data is collected and managed by individual agencies often on a program-specific basis.

- The same constituent data is often collected by more than one agency and kept in redundant data stores.

- There is no standard information classification system to assign consistent and appropriate protections for data as it travel within and outside the enterprise.

- Electronic records are stored by agencies most often in proprietary formats that jeopardize the long-term accessibility of those records.
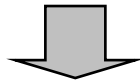
Target State

- Data is categorized at the Executive Office or Community of Interest level to identify data that may be reusable or that can support multiple business processes

- XML data standards are adopted for all new development projects

- Data that can be used by multiple applications is collected once and encapsulated as service components that can be reused by those applications

- All data is classified for sensitivity according to a standard enterprise classification system. Data classification is captured as metadata that travels with the information

- Electronic records are stored in standard open formats with associated metadata and are managed using enterprise Records Management Applications (RMAs)

BOUNDARY

The Information Domain addresses specifications for Data Interoperability, Data Management, Data Formats, and Records Management. Inclusion of these specifications in the development of service oriented applications is addressed in the Application Domain.

RELATED POLICIES

- Enterprise Open Standards Policy
- Enterprise Information Classification Policy (TBD)

ASSOCIATED DISCIPLINES

- Data Interoperability
- Data Management
- Data Formats
- Records Management (TBD)
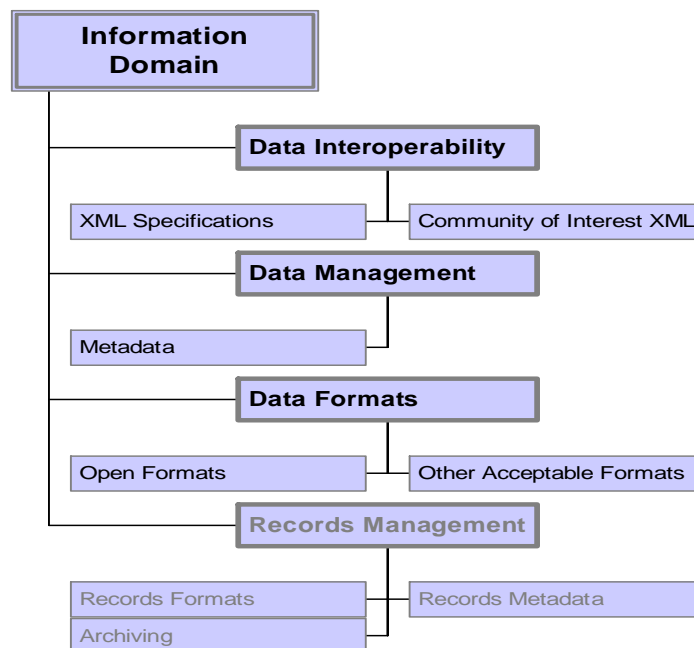
Domain: Information
Discipline: Data Interoperability

DESCRIPTION

One of the most critical SOA decisions for the Commonwealth is the adoption of XML as the primary standard for Data Interoperability. XML has become the lingua franca of application integration, facilitating application interoperability, regardless of platform or programming language. The adoption of XML is the cornerstone of the Commonwealth's Service Oriented Architecture (SOA) vision of a unified enterprise information environment.

Agencies should consider the use of XML for all projects, and should implement XML, unless there are compelling business reasons not to do so. XML should always be considered when undertaking new work or when beginning a major overhaul of an existing system. Agencies should always consider the fact that an XML solution will result in greater long-term benefits for the agency and the enterprise as a whole.

RELEVANT STANDARDS ORGANIZATIONS

- **W3C** - The World Wide Web Consortium was created in October 1994 to lead the World Wide Web to its full potential by developing common protocols that promote its evolution and ensure its interoperability. W3C has around 400 Member organizations from all over the world and has earned international recognition for its contributions to the growth of the Web. More information about W3C can be found at http://www.w3.org

- **WS-Interoperability —** The Web Services Interoperability Organization is an open industry effort chartered to promote Web Services interoperability across platforms, applications, and programming languages. More information about WS-I can be found at http://www.ws-i.org.

STAKEHOLDERS/ROLES

- designers and implementers of Commonwealth information services
- external and internal users of government information
- enterprise application and data architects
- software development service providers
- business strategists and analysts
- system owners
- project managers

ROADMAP

Currently XML is just beginning to be used by agencies to create XML-aware applications. The Mass.gov portal content management solution uses XML to separate content from presentation. The Enterprise Open Standards policy requires compliance with open

standards for prospective IT acquisitions however government records are currently captured in a variety of proprietary and open formats. The target state includes the ubiquitous use of XML for Data Interoperability in application development and content management as well as the use of open formats for displaying and storing data files.

ENTERPRISE TECHNOLOGY SOLUTION

Not applicable

ASSOCIATED TECHNOLOGY AREAS

- XML Specifications
- Community of Interest XML

Domain: Information
Discipline: Data Interoperability
Technology Area: XML Specifications

DESCRIPTION

What is commonly referred to as "XML" is actually a large collection of specifications that rely on XML-encoded packets or instructions. The set of specifications includes: XML Schema, XSLT, XPath, and XQuery to name a few. But all have one requirement in common: all of these XML specifications require an SOA infrastructure that can parse, transform and process XML at network speeds.

Being text-based, XML more readily supports incremental development, debugging, and logging. Other XML benefits include:

- Long-term reuse of data, with no lock-in to proprietary tools or undocumented formats

- The use of inexpensive off-the-shelf tools to process data

- Reduced training and development costs by having a single format for a wide range of uses

- Increase reliability, because applications can automate more processing of documents

- Businesses and governments can now define platform-independent protocols for the exchange of data

- Information presentation flexibility, under style sheet control

TECHNOLOGY SPECIFICATION: EXTENSIBLE MARKUP LANGUAGE (XML)

Description – XML is a self-describing, extensible markup language that encodes the description of a document's storage layout and logical structure. XML provides a mechanism to impose constraints on this logical structure. XML is text-based, so XML fragments are easily created, edited, and managed using common utilities. Originally designed to meet the challenges of large-scale electronic publishing, XML is playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere. XML is a meta-language, which enables interchange of information with any kind of application, in various presentations, for different target groups and different purposes.

Guidelines -

- *Stay with open standards:* To insure maximum interoperability it is recommended that proprietary extensions to any XML specifications be avoided.

- *Partner with industry and other government jurisdictions:* There is a tremendous amount of work being done on vertical specific vocabularies and there are additional initiatives that tend to be more horizontal in their approach. Many government agencies have begun working with these initiatives and they are helping to create a standard they can use with their industry partners.

- *Publish the work that is being developed:* This is a tremendous step toward interoperability and also allows other organizations to share in the benefits. This can lower costs and accelerate usage of the specification.

- *Maintain extensibility:* XML design can be a complicated task but can allow agencies to model a process to gain efficiencies. Creating an extensible architecture can allow schemas to be versatile and dynamic by design.

- *Start small:* Look for a specific area that you can begin in and then expand the scope. Starting with the entire framework of an organization's data can be overwhelming and prohibitively expensive. A smaller pilot project can get XML introduced in a production setting and it will grow as the opportunity and resources are available.

Standards and Specifications –

- XML v. 1.0: XML 1.0 has been fully ratified by the W3C, and is included in the WS-Interoperability Basic Profile 1.0. XML 1.0 can be validated against the formal definition of the protocol specification.

  Refer to: http://www.w3.org/xml

Migration Strategy - Agencies should begin to use XML for Data Interoperability requirements. Agency or Secretariat-specific XML specifications and policies must be compliant with the enterprise XML specifications detailed in the ETRM.

TECHNOLOGY SPECIFICATION: XML SCHEMA

Description – The purpose of an XML Schema is to define the valid structure of an XML document. An XML Schema:

- defines elements that can appear in a document

- defines attributes that can appear in a document

- defines which elements are child elements

- defines the order of child elements

- defines the number of child elements

- defines whether an element is empty or can include text

- defines data types for elements and attributes

- defines default and fixed values for elements and attributes

Schemas express shared vocabularies and provide a means for defining the structure, content and semantics of XML documents.

Guidelines – All schema need to be compliant with the WS-Interoperability Basic Profile, to insure interoperability with SOAP, WSDL and UDDI.

Standards and Specifications –

- XML Schema Part 1: Structures and XML Schema Part 2: Data types – These XML Schema specifications have been published as open standards by the W3C, and are included in the WS-Interoperability Basic Profile 1.0.

  Refer to: http://www.w3.org/TR/2004/REC-xmlschema-1-

20041028/structures.html for Part 1: Structures and
http://www.w3.org/TR/2004/REC-xmlschema-2-20041028/datatypes.html for
Part 2: Data types.

Migration Strategy - XML Schemas should be used, in most Web applications, as a migration strategy away from DTDs.

TECHNOLOGY SPECIFICATION: EXTENSIBLE STYLESHEET LANGUAGE (XSL)

Description – This specification defines the features and syntax for the Extensible Style Sheet Language (XSL), a language for expressing style sheets. It consists of two parts:

1. a language for transforming XML documents, and

2. an XML vocabulary for specifying formatting semantics.

An XSL style sheet specifies the presentation of a class of XML documents by describing how an instance of the class is transformed into an XML document that uses the formatting vocabulary.

Guidelines – Given a class of arbitrarily structured XML documents or data files, designers use an XSL style sheet to express their intentions about how that structured content should be presented; that is, how the source content should be styled, laid out, and paginated onto some presentation medium, such as a window in a Web browser or a hand-held device, or a set of physical pages in a catalog, report, pamphlet, or book.

Standards and Specifications –

- XSL v. 1.0: XSL v. 1.0 is an open standard ratified by W3C that defines a language for expressing style sheets.

  Refer to: http://www.w3.org/TR/xsl/

Migration Strategy – While CSS can be used to style HTML documents XSL, is able to transform documents. For example, XSL can be used to transform XML data into HTML/CSS documents on the Web server. This way, the two languages complement each other and can be used together. Both languages can be used to style XML documents.

TECHNOLOGY SPECIFICATION: XML QUERY LANGUAGE (XQUERY)

*Description –*XQuery for XML is like SQL for relational databases. Compared to SQL, it is designed to be a language in which queries are concise and easily understood. It is also flexible enough to query a broad spectrum of XML information sources, including both databases and documents. XQuery 1.0 uses the structure of XML to express queries across all these kinds of data, whether physically stored in XML or viewed as XML via middleware. XQuery operates on the abstract, logical structure of an XML document, rather than its surface syntax. This logical structure is known as the data model.

Guidelines – XQuery should be used for integration and transformations. With transformation powers that rival XSLT, XQuery not only provides query results, but can also prepare those results for presentation. XQuery is more efficient than XSLT when transforming the results of a database query. Use XQuery when you have requirements to search multiple back-end systems and combine results, effectively integrating multiple sources of information.

Standards and Specifications – XQuery v. 1.0 is not yet ratified as an open standard, i.e. XQuery 1.0 is still a W3C Working Draft. However, XQuery 1.0 is being widely implemented so it is included in the ETRM as an emerging standard. The specification describes a query language called XQuery, which is designed to be broadly applicable across many types of XML data sources. XQuery 1.0 has been defined jointly by the XML Query Working Group and the XSL Working Group. The XPath 2.0 and XQuery 1.0 Working Drafts are generated from a common source. These languages are closely related, sharing much of the same expression syntax and semantics, and much of the text found in the two Working Drafts is identical. For more information go to http://www.w3.org/TR/xquery/

Migration Strategy – When evaluating XML products, agencies should include XQuery support in selection criteria. It is expected that XQuery will soon be ratified as a W3C open standard.

Domain: Information
Discipline: Data Interoperability
Technology Area: Community of Interest XML

DESCRIPTION

Extensible Markup Language (XML) and XML-based schema languages provide a strong, yet easy to adopt, set of technologies for achieving service interoperability within specific communities of interest, e.g. justice, health, finance, education. Standardized Community of Interest XML specifications enable the exchange of structured information between different applications, agencies and/or business partners in a platform-independent way.

As a result, Community of Interest groups and standards bodies have started to adopt XML to specify both their vocabularies and schema. These schemas are becoming widely published and implemented to facilitate communication between both governments and businesses. Wide support of XML has also resulted in independent solution providers developing solutions that enable the exchange of XML-based information with other third-party or custom-developed applications.

To avoid confusion, please be aware that other documents and the trade press may also refer to Community of Interest XML as:

- Domain specific XML
- Industry specific XML
- Vertical XML standards

TECHNOLOGY SPECIFICATION: GLOBAL JUSTICE XML DATA MODEL (GLOBAL JXDM)

*Description –*The Global JXDM is a comprehensive product that includes a data model, a data dictionary, and an XML schema. The Global JXDM is sponsored by the Federal Government's OJP (Office of Justice Programs), with development supported by the Global XML Structure Task Force (XSTF), which works closely with researchers at the Georgia Tech Research Institute. The XSTF is composed of government and industry-domain experts, technical managers, and engineers.

The Global JXDM is an XML standard to be used specifically for criminal justice information exchanges, providing law enforcement, public safety agencies, prosecutors, public defenders, and the judicial branch with a tool to effectively share data and information in a timely manner. The Global JXDM removes the burden from agencies to independently create exchange standards, and because of its extensibility, there is more flexibility to deal with unique agency requirements and changes. Global JXDM is endorsed by the Federal Government, the National Association of State CIO's (NASCIO) and the National Governor's Association (NGA) among others.

Guidelines -

- All instances must be validated against the Global JXDM reference schema.

- If the appropriate component (type, element, or attribute) required for the application exists in the Global JXDM, use that component (i.e., do not create a duplicate of one that already exists).
- Be semantically consistent. Use Global JXDM components in accordance with their definitions. Do not use a Global JXDM element to represent data other than what its definition describes.
- Apply XML Schema rules correctly and consistently.

Standards and Specifications –

- Global JXDM v. 3.0.2: This latest release of the Global JXDM series is enhanced to increase the ability of justice and public safety communities to share justice information at all levels, laying the foundation for local, state, and national justice interoperability.

  Refer to: http://www.it.ojp.gov/topic.jsp?topic_id=43

Migration Strategy – Agencies engaged in criminal justice information exchanges should migrate to XML that utilizes the Global JXDM data model.

Domain: Information
Discipline: Data Management

DESCRIPTION

Data Management standards for the Commonwealth are intended to improve data:

- Conformity - What data is stored in a non-standard format?
- Consistency - What data values give conflicting information?
- Accuracy - Does the data accurately represent reality or a verifiable source?
- Duplication - What data records are duplicated?
- Integrity - What data is missing important relationship linkages?

Data Management problems can occur in many different ways. The most common include:

- A lack of enterprise standards and policies
- Inadequate data entry procedures
- Errors in the migration process from one system
- Data coming from outside may not adhere to standards
- Data received may be of dubious quality

Agencies need to share information visibility across the Commonwealth, regardless of how far along they are in their plans to implement a Service-Oriented Architecture (SOA). Without visibility into the workings of the systems, applications, and other elements of their IT infrastructure, agencies are unable to manage or improve their IT environment, eliminate stove pipes, and most importantly, meet their business requirements.

A key to the enterprise visibility issue is metadata: information about shared services. To provide adequate IT visibility, agencies must follow basic metadata best practices for discovering and organizing metadata, encapsulating business logic in metadata, managing with metadata, and modeling with metadata.

A significantly underused mechanism for working with Web services is the services metadata repository. At present, these repositories primarily store only the interfaces for services. However, for Web services to be supportive of fusion, additional metadata is necessary. Service metadata includes sequencing information to properly order service execution, parameters and exception handling information for the process model, and data to manage services into usable assemblies. Content metadata, such as user interface elements, and the connection of Web services to multiple portlets must be stored in metadata to allow modification of the system without code changes. For Services to be searchable across applications they must be versioned and represent processes that are independent of a single-application model.

RELEVANT STANDARDS ORGANIZATIONS

- **W3C** - The World Wide Web Consortium was created in October 1994 to lead the World Wide Web to its full potential by developing common protocols that promote its evolution and ensure its interoperability. W3C has around 400 Member organizations from all over the world and has earned international recognition for its contributions to the growth of the Web. More information about W3C can be found at http://www.w3.org

- **WS-Interoperability —** The Web Services Interoperability Organization is an open industry effort chartered to promote Web Services interoperability across platforms, applications, and programming languages. More information about WS-I can be found at http://www.ws-i.org.

STAKEHOLDERS/ROLES

- designers and implementers of Commonwealth online services
- external and internal users of government information
- enterprise application and data architects
- external software development service providers
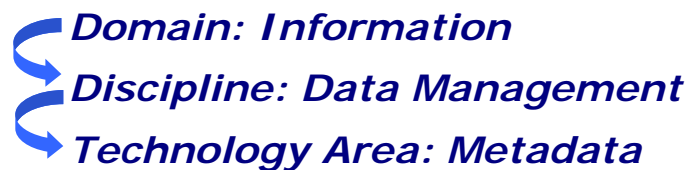- business strategists, system owners, and project managers

ROADMAP

Currently there is a lack of cross-agency data management standards. As the need for cross-agency interoperability increases, the need for metadata visibility becomes critical. The target state is a profusion of metadata design artifacts, such as XML Schemas and Web Services Definition Language (WSDL) documents as well as an Enterprise Web Service Registry to help discover and manage schema, policies, WSDLs, etc.

ENTERPRISE TECHNOLOGY SOLUTION

- Web Service Registry (see Integration Domain)

ASSOCIATED TECHNOLOGY AREAS

- Metadata

Commonwealth of Massachusetts
Enterprise Information Technology Architecture

Domain: Information

Discipline: Data Management

Technology Area: Metadata

DESCRIPTION

Web services use metadata to describe what other endpoints need to know to interact with them. Specifically, WSDL describes abstract message operations, concrete network protocols, and endpoint addresses used by Web services; XML Schema describes the structure and contents of XML-based messages received by and sent by Web services.

Metadata provides a means for defining, obtaining and organizing the data obtained from endpoints, as well as propagating data to endpoints. A Registry can actively pull metadata from endpoint services, and the endpoints (or hosting environments) can actively pull the metadata from the Registry.

Using metadata provides the following advantages:

- It provides a mechanism for locating reusable components when they need to be reused.

- The taxonomy in metadata will help the Commonwealth create a reference model of the services provided

- It facilitates good governance via well-defined processes that identify and maintain high-quality information and services

- Leads to having a team responsible for the management of the service metadata repository

It is also essential to have a standards-based development framework (SODA) that encourages reuse of these metadata.

TECHNOLOGY SPECIFICATION: WEB SERVICE DESCRIPTION LANGUAGE (WSDL)

Description – The Web Service Description Language (WSDL) can be used to describe a service so that individuals and businesses can provide or consume those services electronically. A WSDL (pronounced wiz-dill) is a document written in XML that describes a Web Service. It specifies the location of the service and the operations (or methods) the service exposes.

Guidelines – There is a clear process that developers need to follow to effectively develop an interoperable Web Service. The WSDL interface should be generated first, before the functionality of the Web Service is written. There are two major ways to generate a WSDL. The "WSDL First" practice consists of writing the WSDL by hand and then creating the service code from the WSDL. This practice affords the most flexibility in WSDL design and is best for creating interoperable WSDLs because it allows language-independent development, leverages the power of XML, and leverages standard markup languages to define types.

The alternative way to create a WSDL is to have a Web Service toolkit automatically generate the WSDL from the service code. Using this method, it is important to choose a

good interoperable toolkit and render the WSDL from a skeletal interface. The business functionality should not be written until after the interface is stable, and the WSDL is determined to be free of interoperability problems. This way, business functionality will not have to be reworked when it is discovered that the interface is not interoperable. Checking the WSDL is the first line of defense in preventing interoperability problems. By generating the WSDL as soon as possible, problems can be caught early, saving time and money. It is important to keep the structure of the data being passed between the Web Service Consumer and Provider as simple as possible. Not every toolkit will handle all of the XML Schema data types, and keeping the data structures simple will increase interoperability.

Standards and Specifications –

- WSDL v. 1.1 - This specification has been ratified by the W3C as an open standard, and is included in the WS-Interoperability Basic Profile 1.0. To address any interoperability concerns, the Web Services Interoperability Organization (WS-I) has recommended using a restricted subset of WSDL, in the Basic Profile 1.0, which allows the Commonwealth to focus on fewer issues, for greater compatibility. By restricting possible interpretations, the WS-I provides a greater assurance of interoperability.

    Refer to: http://www.w3.org/TR/wsdl.html

Migration Strategy – Use of WSDL is an essential part of any migration to Web Services. Initially WSDL is typically used for static use cases, with eventual migration to more dynamic use cases. All Web Service Providers and Consumers must migrate to WSDL standards. Service Providers should be developed based on these standards and Service Consumers must be able to understand WSDL.

Domain: Information
Discipline: Data Formats

DESCRIPTION

Information can be captured and exposed via a variety of data types. For example, information can be captured as text, numbers, maps, graphics, video and audio. The software used to create data files stores these files in different data formats. These formats can be proprietary and therefore controlled and supported by just one software developer. Formats can also be non-proprietary or open.

The Commonwealth defines **open formats** as specifications for data file formats that are based on an underlying open standard, developed by an open community, affirmed and maintained by a standards body and are fully documented and publicly available. **It is the policy of the Commonwealth of Massachusetts that all official records of the Commonwealth be created and saved in an acceptable format as detailed below.**

BOUNDARY

The Data Formats Discipline addresses the acceptable formats in which data can be presented and captured. Data formats for the long term conservation of files will be addressed in the Records Management Discipline.

STAKEHOLDERS/ROLES

- application developers
- content developers
- end users of government information and services

ROADMAP

Information that traditionally has been presented in text form is increasingly being enriched through the use of multimedia data types such as graphics, audio and video. The variety of data formats used however raises concerns regarding interoperability and accessibility. Given that XML is the cornerstone of the Commonwealth's Service Oriented Architecture (SOA) vision of a unified enterprise information environment, it is crucial that the schema used to create XML files meet the open format definition as well. The target state is the ubiquitous use of open formats to capture and store data within applications and in individual data files.

ENTERPRISE TECHNOLOGY SOLUTION

Not applicable

RELEVANT STANDARDS ORGANIZATIONS

- **IETF** - The Internet Engineering Task Force (IETF) is a large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. It is open to any interested individual. The actual technical work of the IETF is done in its working groups, which are organized by topic into several areas (e.g., routing, transport, security, etc.). More information on the IETF can be found at http://www.ietf.org/home.html.

- **ISO** - In 1946, delegates from 25 countries decided to create an international organization "to facilitate the international coordination and unification of industrial standards". The new organization, ISO, officially began operations on 23 February 1947. The International Standards Organization is a network of the national standards institutes of 148 countries with a Central Secretariat in Geneva, Switzerland, that coordinates the system. More information about ISO can be found at http://www.iso.org.

- **OASIS** – Organization for advancement of structured information standards is a not-for-profit, international consortium that drives the development, convergence and adoption of e-business standards. Members themselves set the OASIS technical agenda, using a lightweight, open process expressly designed to promote industry consensus and unite disparate efforts. OASIS produces worldwide standards for security, Web services, conformance, business transactions, supply chain, public sector, and interoperability within and between marketplaces. More information about OASIS can be found at http://www.oasis-open.org.

- **W3C** - The World Wide Web Consortium was created in October 1994 to lead the World Wide Web to its full potential by developing common protocols that promote its evolution and ensure its interoperability. W3C has around 400 Member organizations from all over the world and has earned international recognition for its contributions to the growth of the Web. More information about W3C can be found at http://www.w3.org.

ASSOCIATED TECHNOLOGY AREAS

- Open Formats
- Other Acceptable Formats

Domain: Information

Discipline: Data Formats

Technology Area: Open Formats
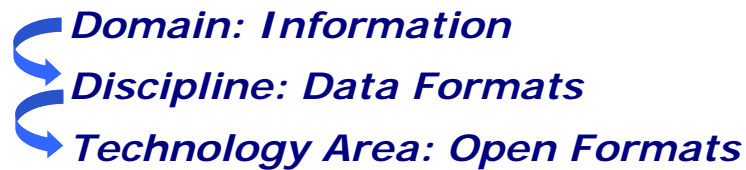
DESCRIPTION

The Open Formats Technology Area addresses open standards and specifications for the presentation of data as office documents, text, numbers, maps, graphics, video and audio. The selection of format must consider the access channel being used (Web, PDA, cell phone), the nature of the data and structure (legal requirements that address preservation of document structure), and ease of accessibility for users.

The open formats identified below do not yet address all data types. Future versions of the ETRM will address open formats for map, graphics, video and audio data.

TECHNOLOGY SPECIFICATION: OASIS OPEN DOCUMENT FORMAT FOR OFFICE APPLICATIONS (OPENDOCUMENT)

Description – The OASIS Open Document Format for Office Applications (OpenDocument) is a standardized XML-based file format specification suitable for office applications. It covers the features required by text, spreadsheets, charts, and graphical documents. The specification was recently approved by OASIS as an open standard. OASIS has also submitted the standard to ISO for consideration as an international standard for office document formats.

Guidelines – The OpenDocument format must be used for office documents such as text documents (**.odt**), spreadsheets (**.ods**), and presentations (**.odp**). The OpenDocument format is currently supported by a variety of office applications including OpenOffice.org, StarOffice, KOffice, and IBM Workplace.

Standards and Specifications –

- OpenDocument v. 1.0 – Defines an XML schema for office applications and its semantics. The schema is suitable for office documents, including text documents, spreadsheets, charts and graphical documents like drawings and presentations, but is not restricted to these kinds of documents.

   Refer to: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office

Migration – Given the majority of Executive Department agencies currently use office applications such as MS Office, Lotus Notes and WordPerfect that produce documents in proprietary formats, the magnitude of the migration effort to this new open standard is considerable. Agencies will need to develop phased migration plans allowing them to configure existing applications to save office documents by default in the OpenDocument format with an implementation date of January 1, 2007. Any acquisition of new office applications must support the OpenDocument format natively.

Agencies should begin to evaluate office applications that support the OpenDocument specification to migrate from applications that use proprietary document formats. As of January 1, 2007 all agencies within the Executive Department will be required to[1]:

1. Use office applications that provide conformance with the OpenDocument format, and

2. Configure the applications to save office documents in OpenDocument format by default.

TECHNOLOGY SPECIFICATION: PLAIN TEXT FORMAT

Description – Plain Text refers to textual data in American Standard Code for Information Exchange (ASCII) format. Plain text is the most portable format because it is supported by nearly every application on every machine. It is quite limited, however, because it cannot contain any formatting commands.

Guidelines – Because of its limitations, Plain Text should not be used for documents where formatting is important or is part of the official record. Use of Plain Text for formatting email messages reduces the likelihood of email client interoperability issues and reduces download time for clients with dial-up connections.

Standards and Specifications –

- Plain Text Format – Documents are presented as **.txt** files

Migration Strategy – Documents created in proprietary document formats can be saved as .txt files when formatting is not important.

TECHNOLOGY SPECIFICATION: HYPERTEXT DOCUMENT FORMAT

Description – Hypertext documents contain links to other documents and data files that allow the reader to easily move from one document/data file to another with the aid of an interactive browser program.

Guidelines – Hypertext document format is the preferred format for documents that will be accessed through the Internet/Intranet or using a web browser.

***Standards and* Specifications** –

- Hypertext Document Format – Hypertext authoring or conversion software must support HTML v. 4.01. Documents are presented as **.html** files.

Migration Strategy – Many documents created in proprietary formats can be saved as **.html** files.

---

[1] Agencies are permitted to take actions for the purpose of ensuring that the rights of persons with disabilities are respected. ITD, in consultation with the community of persons with disabilities, will develop a plan to address any accessibility concerns.

Domain: Information

Discipline: Data Formats

Technology Area: Other Acceptable Formats
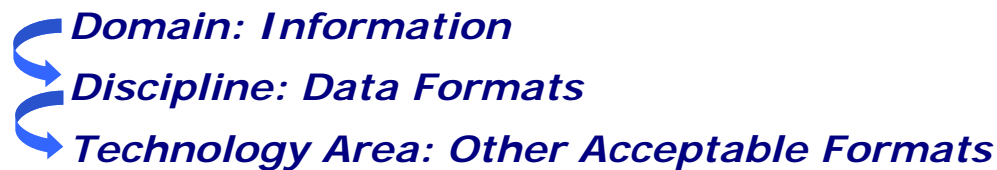
DESCRIPTION

The Other Acceptable Formats Technology Area addresses de facto formats and specifications for the presentation of data as text, numbers, maps, graphics, video and audio that are also acceptable for use with official records of the Commonwealth. These formats, while not affirmed by a standards body, meet the other criteria of openness and are therefore considered acceptable at this time.

The acceptable formats identified below do not yet address all data types. Future versions of the ETRM will address acceptable formats for map, graphics, video and audio data.

TECHNOLOGY SPECIFICATION: PORTABLE DOCUMENT FORMAT

Description – Portable Document Format (PDF) is a file format specification developed by Adobe Systems. PDF is a universal file format that preserves the fonts, images, graphics, and layout of any source document, regardless of the application and platform used to create it.

Guidelines – The PDF format may be used for documents whose content and structure will not undergo further modifications and need to be preserved. Agencies can use a number of proprietary and open source products to create PDF files. Application developers can also build in PDF creation functionality into their applications using the latest reference specification published by Adobe. PDF readers are freely available for download.

Standards and Specifications –

- Portable Document Format – Must be based on, at a minimum, PDF Reference v. 1.5 that supports XML functionality. Documents are presented as **.pdf** files.

  Refer to: http://www.adobe.com/products/acrobat/adobepdf.html