

Gilbane Beacon

Guidance on Content Strategies, Practices and Technologies

Content Management for the Defense Intelligence Enterprise

How XML and the Digital Production Process Transform
Information Sharing across the Intelligence Community

By Geoffrey Bock and Barry Schaeffer, Senior Analysts

May 2009

Sponsored by



Facing the Content Management Challenge

As described by the Defense Intelligence Agency (DIA) *Strategic Plan 2007 – 2012*, the Defense Intelligence Enterprise thrives on producing insightful information to deliver “the right intelligence for the right customer at the right time.”¹ “DIA believes collaboration between collectors and analysts is critical to tearing down walls that limit the ability to collect, process, and share information quickly. DIA will accomplish enhanced collaboration through . . . implementation of interoperable systems and creation of an information sharing culture.”² The success of this effort depends on the ability to create, access, share, and distribute actionable information across the intelligence community, in a systematic manner.

Actionable Information and Interoperable Systems

There are many steps to managing the actionable information for the multiple stakeholders of the Defense Intelligence Enterprise.³ In the digital age and with the advent of global networks, information should flow seamlessly yet securely from the content creators to the content consumers – from the collectors and analysts who author, edit, review, approve, and publish reports through digital production processes, to the task teams and decision makers who act upon them.

Globally distributed across various organizations, stakeholders should easily communicate, coordinate, and collaborate with one another over the network. Integral to the stakeholders of the Defense Intelligence Enterprise are the interoperable systems that are based on common standards.

Leveraging an XML-based Infrastructure

Over the past decade, DIA has invested heavily in XML to drive interoperability standards for defining, managing, and publishing community-accessible content. Mandated in 2002 as Intelligence Community Metadata System for Publications (IC-MSP), required to be in use by 2003, and most recently specified as Implementation Profile for Intelligence Publications (XML Encoding) (PUBS-XML),⁴ DIA defines the XML schemas and data dictionaries for the major types of

¹ Defense Intelligence Agency, *Strategic Plan 2007 – 2012: Leading the Defense Intelligence Enterprise*, p. 12.

² *Strategic Plan*, p. 14.

³ As defined by the *Strategic Plan* (p. 3), the “Defense Intelligence Enterprise [encompasses] the Defense Intelligence Community (organizations within the Department of Defense (DoD) that have an intelligence mission and/or function), plus all of their stakeholders. Stakeholders [include] entities and interests that are involved in creating, sustaining, and enhancing the mission capacity, either voluntarily or involuntarily, of the enterprise. Enterprise [is] the nodal element within a network of interrelated stakeholders.”

⁴ See http://www.dni.gov/ICIS/dsca/pubs/xml/pubs_xml.htm

documents produced within the intelligence community. Furthermore, the Intelligence Community Standard for Publication Metadata (ICS PUBS) relies on PUBS-XML to describe the structural components of documents and provides the foundations for “the successful implementation of a wide range of advanced automated tools that will help find, organize, analyze, and manage information products.”⁵

With the adoption of content management technologies that leverage XML, it is now possible for stakeholders to develop and deploy interoperable systems. But XML as the underlying tagging standard is only the means to an end -- the overall DIA objective of ensuring interoperability and community-wide information sharing, including the mandate for stakeholders to submit reports to the Library of National Intelligence (LNI).

In short, the prospect of end-to-end information flows based on an XML infrastructure is fast becoming a reality. It is altogether feasible for analysts to collaborate around content creation activities, tag documents using predefined XML schemas, and submit them to a seamless review and approval process. Once approved and published, these documents can be aggregated and distributed to the multiple stakeholders, in various renditions and formats, formally catalogued in a library system, and retained as records in a records management system. Building the information sharing culture for the Defense Intelligence Enterprise depends, in large part, on having the right content management capabilities in place.

Key Capabilities for Managing Content

To develop and deploy interoperable systems, it is important to focus on how documents tagged as PUBS-XML are created, indexed, stored, accessed, published, and retained as records for subsequent research and reporting. Stakeholders across the Defense Intelligence Enterprise need to share content in a secure and structured fashion – both when developing reports as work-in-progress documents and then as published documents when approved for release. When developing and publishing content, it is important to manage PUBS-XML documents throughout their life cycles, both at a granular level and in a systematic manner.

Specifically, stakeholders that create the actionable information need to ensure content security, workflow, collaboration, and integration with disparate sources. At a minimum, stakeholders need to exploit the powerful content and process services provided by an enterprise content management (ECM) platform to manage work-in-progress documents and the publication of final form documents. In addition, for DIA to achieve its objective of interoperable systems, it is also important

⁵ See http://www.dni.gov/ICIS/dsca/pubs/ics_pubs.htm.

to focus on how the XML documents (tagged through PUBS-XML and its predecessor standards) are stored, published across the intelligence community, retained as records, and submitted to the LNI.

XML adds structure to enterprise content. For system interoperability and management, it is essential to be able to combine the capabilities for XML-enabled ECM together with those of ECM-enabled XML.

A Native XML Database

Thus, to achieve the capabilities of interoperable systems and XML as described by DIA, it's important to manage XML documents in their full fidelity: storing the tag-sets in their sibling order as well as preserving the information about document identity, key internal content structures, processing instructions, whitespace, and so forth. This task is best accomplished by maintaining XML documents in a native XML database.

An XML database has a number of features that are useful for working with documents, once they are tagged as PUBS-XML (and its predecessors). The most important include the flexibility of access to the full XML data model, XML-aware full-text searches, and support for XQuery -- all without requiring explicit (or predefined) componentization when items are stored within the database. These features allow documents to be stored and queried in a single repository. Other useful features include node-level updates (which reduce the cost of updating large documents), links, versioning, and the flexibility of extending a schema on-the-fly. An XML database can also exploit various XML standards—such as DOM, XPath, XSL, XProc, and XLink—to provide a powerful environment for processing and rendering content.

With an XML database, there is no need to lock in a predefined schema at design time. A native XML database can easily extend the schema to add new tag sets and rapidly incorporate the changes and additions that occur in practice.

An XML Content Platform: Combining XML and ECM

An XML content platform delivers the combined benefits of XML and ECM. An XML content platform leverages the resources of a native XML database and exposes them as XML services. This platform also includes ECM services for content security, workflow, lifecycle management, and collaboration -- services that can drive the digital production processes across the Defense Intelligence Enterprise.

XML Management Capabilities

Specifically, an integrated XML content platform provides resources for storing and accessing XML content in its native (or hierarchical) form. These resources enable full access to the content's logical components, values, attributes, and internal structures without fragmenting (or chunking) an XML document into physical components. The XML content platform can be configured to automatically parse, validate, and transform incoming XML documents into discrete components according to a specified DTD or schema (such as one based on PUBS-XML) before storing them. This componentization is typically implemented to support the logical functions of authoring, management, and collaboration. Granular access, though, remains independent of the componentization required by the underlying storage environment.

For stakeholders, an XML content platform provides important capabilities for managing documents within a dynamic environment. As the demand for greater granularity in storage and access grows, so too does the granularity and complexity of the required componentization. This creates a significant management challenge in the sheer number of components as well as in the need for constantly changing content management capabilities to handle the resulting changes.

Once captured and stored within the XML content platform, XML documents can be maintained in their entirety, or as virtual documents composed of discrete components, and managed through predefined document lifecycles. The componentization can occur automatically, and can be configured to define the various object types, where to store them, what metadata to assign, and the level within a document at which the decomposition into components should occur. At the same time, new and revised content can be imported into the XML database, making it available at an infinite level of granularity.

The processing of XML documents occurs within a high performance and highly scalable environment. Each component is maintained as a distinct object for review, revision, collaboration, and participation in designated approval processes. At the same time, the details of the content may be accessed instantly and completely, at any desired level of granularity, via searches against the native XML repository into which it has been imported.

ECM Capabilities

An XML content platform provides the enterprise-scale services and the overall framework for managing any type of content, including XML documents, in a consistent fashion. The platform manages efficiently designed content components in their full fidelity, provides full granular access to the content, and also supports the range of ECM capabilities to maintain these components within interoperable systems.

- **Security.** An XML content platform manages authentication and access controls at a fine-grained, content component level. Platform-level security services, leveraging the native XML access provided by the system, determine how stakeholders (or the applications they run) are authenticated and authorized to access and use XML at any predefined level of granularity in a consistent and auditable manner.
- **Workflow.** An XML content platform with ECM can structure the flow of XML-tagged documents, as work-in-progress from creators to editors, reviewers, and approvers. The platform can manage the workflow status, links, and events at both document and sub-document levels of granularity.
- **Lifecycle management.** With an XML content platform, ECM lifecycle policies can be extended to XML. The platform defines, maps, and implements flexible content lifecycle rules according to the business policies established by the stakeholders. These can include the business rules for managing changes that apply to XML as it moves through predefined stages such as "draft," "in review," "active," and "obsolete." An application can use XQuery to analyze XML documents, determine the lifecycle stage, and return only components identified as being at a predefined stage.
- **Collaboration.** With an XML content platform, analysts in various stakeholder organizations can collaborate and share information as they author, edit, review, and approve documents. Frequently they can find all essential work items in sets of shared folders, which only predefined groups of analysts can access. Analysts can rely on the platform to organize and manage the tasks required to publish documents to other stakeholders.
- **Information Rights Management (IRM).** The XML content platform enables stakeholders to control, protect, and secure information that is distributed both within their organizations and with outside stakeholders. IRM allows stakeholders who publish documents to define and change user permissions, recall reports, automatically expire information, set security policies, and dynamically control information across the intelligence community. These stakeholders set the IRM policies to control who can and cannot view, copy, print, and forward the published reports.

Once published, the XML documents can be automatically defined and categorized as records, and, with little or no human intervention, ingested into a records management system or other repositories within the Defense Intelligence Enterprise, such as the library system maintained by the LNI.

Enabling the Digital Production Process

In short, an XML content platform enables the digital production process for multiple stakeholders across the intelligence community. In a digitized work environment, analysts within an organization can easily establish an information-sharing framework for producing actionable information. Common artifacts -- containing processes, tasks, data, and documents -- are readily at hand and maintained through secure workspaces and interoperable systems.

Moreover, as requirements for actionable information change, stakeholders can rapidly respond and adapt their digital production processes to new situations. They can configure and enhance their workspaces and systems, without coding. As a result, stakeholders can leverage the interoperability standards, embedded within an XML content platform, to ensure the seamless and secure flow of intelligence reports to those organizations that need them.

Digital Production Processes in Operation

How does an XML content platform expand the horizons for interoperable systems and digital production processes? With XML-enabled ECM plus ECM-enabled XML, stakeholders can finally manage granular components in their full fidelity, analyze large content collections, recombine items that meet predefined criteria, render documents in multiple formats, and publish them to various other organizations across the intelligence community.

To highlight the power of digital production processes, it is important to assess the impact of an XML content platform on stakeholders across the intelligence community, both when managing work-in-progress documents and when producing approved publications. Let's consider two use cases that illustrate an XML content platform in operation.

Managing Work-in-Progress Documents

A stakeholder within the Defense Intelligence Enterprise, responsible for creating intelligence reports, needs to manage the development, review, and approval of documents. To achieve the DIA objective of system interoperability, the organization develops and implements an XML publishing process.

- Content creators (the collectors and analysts who author, edit, and review documents) use an XML editing environment with templates containing PUBS-XML tags to author, edit, and review documents. The XML editing environment ensures real-time validation of the resulting XML structure and provides error-correction support where necessary.
- Content creators manage their work-in-progress documents through an ECM system, which provides workflow, lifecycle management, version control, role-based permissions, security support, information rights management, and XML-based metadata management capabilities. Content managed within the ECM system meets DIA formatting and metadata standards, including the CAPCO security classification markings at the sub-document (or content component) level.

Content creators rely on template-based editing capabilities to define the document structure up front. When they submit documents to the ECM system, pre-defined XML elements are automatically extracted and incorporated into metadata maintained by the system for use in workflow and document lifecycle management. These work-in-progress documents are often routed internally within the organization, searched by other analysts for key concepts, citations, and

references, associated through links and references, and revised and extended with additional analysis through digital production processes. The ECM system maintains control of this process, and supports the productivity and accuracy of the critical analysis phase.

Building on an XML content platform is essential for system extensibility and adapting to new DIA mandates. The platform approach provides full visibility into the structure of XML documents at all times and phases in their life cycles. When designing and implementing the system, there is no need to predefine the metadata required for searching or content management. XML documents can be managed at various sub-document or content component levels.

The XML content platform assures proper association, collection, and reassembly of content components for revision and output processing. The platform also supports integrated search and retrieval capabilities, by combining XQuery with full-text searching. With an XML content platform in place, document structure can evolve over time without substantially reconfiguring the underlying system. XML documents belonging to different versions of a schema can be seamlessly merged into a single search result.

Publishing Documents to External Stakeholders

Once a stakeholder organization completes its analysis and publishes its results, the final form reports then flow to other external stakeholders across the intelligence community, and into their own repositories for storage, retrieval, and stakeholder-specific decision support applications. Tagged as PUBS-XML documents, these intelligence reports support system interoperability at the content (or data) level; they can be automatically ingested and catalogued by the repositories that various stakeholders maintain. For example, a published report can be automatically incorporated into a records management system, maintained by an external stakeholder – the PUBS-XML tags can determine the entry catalog in the records management file plan. Once the report is ingested, it can be processed and automatically analyzed according to predefined XML rule sets.

By managing documents as PUBS-XML and storing them in their native formats, the XML content platform supports the dynamic publishing and distribution processes. With XQuery and XSLT support and the use of output style sheets, the platform can automatically assemble multiple renditions or versions of the final form reports that are targeted at various audiences. Thus one version of the report might be an executive summary for decision makers, listing key action items, while another version might be for area specialists or subject matter experts, detailing newly identified trends. These resulting documents can use and include embedded classification tags to assure compliance with the access rights of intended recipients.

Decision Criteria for XML Content Management

DIA faces the continuing challenge of developing and deploying interoperable systems for the Defense Intelligence Enterprise. Basing systems on common standards, including PUBS-XML and ICS PUBS, is a critical first step; these standards define the tags for content and data interoperability. But it is also essential to pay attention to the ways in which the content and data, contained within documents, is managed across the intelligence community.

An XML content platform, one that combines ECM and native XML access, provides the required infrastructure to create, access, share, and distribute actionable information to multiple stakeholders. Content creators can manage their work-in-progress documents in a collaborative yet controlled fashion. When they publish their results, documents produced by one organization can be easily redistributed and maintained by the repository of another.

Moreover, within the intelligence community, stakeholders must continually adapt to a changing environment. An XML content platform provides the flexibility and extensibility to ensure that stakeholders can respond rapidly to the new situations facing the Defense Intelligence Enterprise.