Standardizing IRS EPSS by Applying SCORM and S1000D Lessons Learned

Steven A. Kerschenbaum
VERTEX Solutions, Inc.
Falls Church, VA
skersch@vertexsolutions.com

Tim Tate
DoD Job Performance Technology Center
Alexandria, VA
tatet@adlnet.org

ABSTRACT

Because of the nature of its business and the authoritative and complex nature of the Internal Revenue Manual (the IRM), many Internal Revenue Service (IRS) business units have found themselves developing ad-hoc Electronic Performance Support Systems (EPSS) to help navigate and apply its directives effectively. Unfortunately, this has led to static Web systems that are decoupled from the trusted source content (the IRM), and that cannot be leveraged or even discovered by other business units with potentially similar needs.

For these reasons, the IRS began to apply the standardization principles behind the Sharable Content Object Reference Model (SCORM) to EPSS. In conjunction with the ADL Co-Lab and other industry participation, the IRS began the arduous task of applying a SCORM-like, object-oriented rigor to EPSS content development. Using these principles, the IRS defined several Performance Support Objects (SPOs), along with their corresponding metadata elements based on existing standards such as the Dublin Core Metadata Initiative.[1]

While the IRS demonstrated facilitated EPSS successfully using these structures and a collaborative authoring tool (an LCMS), electronic access to the IRM itself has emerged as the principle challenge and ultimate goal for enterprise-wide use at the IRS. Traditionally published in hardcopy, the IRM requires both structural and technological transformations for it to become a "trusted electronic source" for all IRS needs. This led the IRS to examine S1000D and how it might help facilitate the transformation of the IRM into a truly reusable and current electronic content source.

This paper will review the process followed by the IRS to decompose its Performance Support Objects using a SCORM-like approach, its successes in standardizing facilitated EPSS, its challenges in maintaining and leveraging the predominantly print-based trusted source (the IRM), and its goals for applying S1000D principles for transforming the IRM for use with EPSS, WBT, as well as traditional hardcopy.

ABOUT THE AUTHORS

Steven A. Kerschenbaum is Chief Technology Officer and General Counsel for VERTEX Solutions Inc. He focuses on the best ways to use technology to help organizations improve performance. His specific program experience includes strategic planning, requirements development and analysis, program and project management, LMS acquisition and support, and information system design, development, and testing. Steve has published and presented at the 2003 and 2004 I/ITSEC.

Tim Tate has been the Director of the DoD Job Performance Technology Center since 2003, and leads the DoD ADL initiative to link SCORM with the International standard for Technical Data, S1000D. He is also Technology Advisor to the DoD Training Transformation initiative and Chief DoD Advocate for Job Performance Technology. In his roughly twenty-seven years in instructional technology, Tim has served as Training Technology Advisor to OPNAV N00T, represented the Navy as part of the DoD Total Force Advanced Distributed Learning Action Team and DoD Training Transformation initiatives, and conducted the first Human Performance project to use the "Revolution in Navy Training" 4 quadrant model. Tim was also the lead for the Tri-Service MIL-STD 1379D ICW development, and Chief Architect for the Navy's Electronic Classrooms and Distance Learning classroom projects while at Naval Sea Systems Command Headquarters.

[1] **The Dublin Core Metadata Initiative** is an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models. See http://dublincore.org for more information.

Standardizing IRS EPSS by Applying SCORM and S1000D Lessons Learned

Steven A. Kerschenbaum
VERTEX Solutions, Inc.
Falls Church, VA 22202
skersch@vertexsolutions.com

Tim Tate
DoD Job Performance Technology Center
Alexandria, VA
tatet@adlnet.org

INTRODUCTION

Background

The Internal Revenue Service (IRS) has been following the latest developments in Electronic Performance Support Systems (EPSS) since roughly 1997.[2] The IRS, more specifically the Human Capitol Office and Leadership and Education (L&E), has long since recognized the potential importance and significance of EPSS to its customer-centered mission.

In late 2003, L&E recognized a growing proliferation of static Web pages and sites being created and used throughout the Service for performance support. To help evaluate the content types, data types, and technologies and approaches employed, L&E sponsored a meeting at the Department of Defense's Advanced Distributed Learning (ADL) Co-Lab to discuss the latest advances in learning and performance content management and delivery, and to examine a wide variety of IRS-specific examples in use at the time.

With the assistance of the ADL Co-Lab and representatives from across many IRS business units, the EPSS Team concluded the following:

- Industry "best practices," most notably the ADL Sharable Content Object Reference Model (SCORM), had come to show the utility and value in the reusability of courseware content. It seemed that a similar, structured approach for IRS performance content was also possible based on other organization's experiences.[3]
- While many business units and other organizations had explored the use of procedure guides in this way, it was also clear that there was little guidance available for standardizing, maintaining, synchronizing, or developing such content.

Initiative Purpose and Goals

For the reasons mentioned earlier, L&E decided to prototype an e-Guide approach for an end-user tool for customer service representatives addressing Failure-to-File and Failure-to-Pay Penalty issues. In addition to this very tangible goal, several additional objectives were also defined for this initiative:

1. Facilitate and simplify EPSS content management and maintenance for the IRS.

2. Apply an IRS-specific technical approach based on a COTS test platform and relevant content (e.g., a LCMS application and e-ARG FTF/FTP content) and evaluate the results and possibilities. This includes assessing the test platform's ability to facilitate content development, its ability to create sequenced tasking, its ease of maintainability, and its ability to support users with various needs.

Based on these findings, L&E decided to move forward with an initiative to further the progress identified in the meeting, and work towards a way to facilitate the institutionalization of EPSS across the Service.

1. Develop a methodology and specification, based on industry standards and best practices, for defining classes of Sharable Support Objects (SPOs) for the IRS (e.g., e-Guide, calculator, etc.). Definition for SPOs would be based on Shareable Content Object definition and experience.

2. Develop support templates and guides for developing EPSS for the IRS.

[2] An EPSS is defined in the industry as electronic systems or tools that focus on providing just-in-time support and other content to users at the time of need to improve their business performance.

[3] Most notably, these organizations included the U.S. military, European Association for Aerospace Industries (AECMA), and the Co-Lab itself.

Purpose of this Paper

The purpose of this paper is to review the process the IRS undertook to standardize their approach to applying SCORM-like content reference models to their Electronic Performance Support Systems, such as the European Association for Aerospace Industries (AECMA) S1000D Standard for Technical Publications utilizing a Common Source Data Base (CSDB). At a high-level, this paper examines the challenges encountered and the success achieved using this approach, and includes an in-depth discussion of the Performance Support Object (SPOs) model, and the applied methodology for determining SPOs.

Definitions

For the purposes of this publication, it is important to provide specific definitions and clarity to some of the appropriate terminology:

Activity Package – a collection of reusable tasks and their assets described by an XML Manifest that provides a standardized way to exchange these resources between different e-Guides, systems, or tools.

ADL – The Advanced Distributed Learning Initiative, sponsored by the Office of the Secretary of Defense, is a collaborative effort between government, industry and academia to establish a new distributed learning environment that permits the interoperability of learning tools and course content on a global scale. The purpose of the ADL Initiative is to ensure access to high-quality education and training materials that can be tailored to individual learner needs and made available whenever and wherever they are required.

Asset – A single file that may be text, audio, image, web page, or other file type.

Common Source Database – An information store and management tool for all objects required to produce the technical publications within projects using the S1000D specification.

CORDRA – The Content Object Repository Discovery and Registration/Resolution

e-ARG – The e-Accounts Resolution Guide, used for the EPSS prototype

e-Guide – The electronic Guide is one of the IRS' most successful alternatives to traditional training methods. It consists of three key elements: Activity Package, Sharable Performance Object, and Assets.

EPSS – An electronic system or tool that provides the right information, to the right person, at just the right time to improve job performance.

HTML – Hypertext Markup Language is the coded format language used for creating hypertext documents on the Web, and controlling how Web pages appear.

SCORM – The Sharable Content Object Reference Model (SCORM) aims to foster creation of reusable learning content as "instructional objects" within a common technical framework for computer and Web-based learning. SCORM describes this technical framework by providing a harmonized set of guidelines, specification and standards. Borrowing from work of other specification and standards bodies, ADL developed a model for creating and deploying e-Learning.

SPO – Sharable Performance Support Objects are reusable, sharable software entities that can be discretely identified, uniquely distinguishable, and contains consistent attributes and behaviors to provide the necessary information, data, and references to support human performance. For e-Guides within the IRS, these will be task oriented that enable the activities and deliver a specific performance objective.

XML – Originally designed to meet the challenges of large-scale electronic publishing, eXtensible Markup Language plays an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere.

MODELING PERFORMANCE SUPPORT OBJECTS

Defining the Challenge

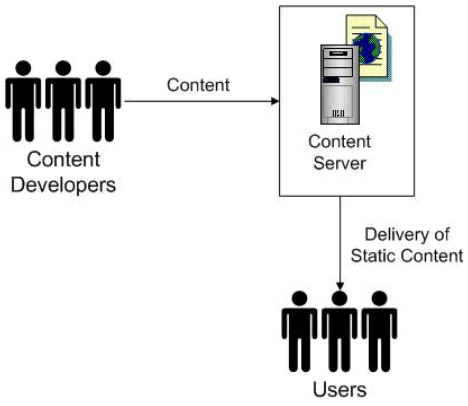

Figure 1 – Traditional HTML Development Model

Industry "best practices," most notably the ADL Sharable Content Object Reference Model (SCORM), had come to show the utility and value in the reusability of courseware content, and that it was also possible for performance (EPSS) content. While Hyper Text Markup Language (HTML) and related technology had made developing Web-based content easier, creating and maintaining the files manually presented many challenges for the IRS:

- HTML combines three (3) logical components into a single, making them extremely difficult to maintain separately:
    - File Presentation (using the markup language)
    - Business Rules and Logic (using .ASP or JavaScript)
    - The raw content itself (text, graphics, reference links, etc.)
- The subcomponents of each performance Web site failed to include a level of granularity that could be indexed, referenced, maintained, or reused.
- Pre-dates the use of style sheets, which are becoming more common today in the IRS. [4]

Using HTML on a small scale was manageable for several years. From a Service-wide perspective, however, the shortcomings of these approaches become painfully clear:

- Large, heavily-used sites required constant attention and maintenance.
- Users were unable to leverage the work already performed across the Service.
- Uninitiated and inexperienced users had little guidance regarding how to begin the process of developing EPSS content in an effective manner.

To address these challenges, the IRS began examining the object-oriented, well-defined reusability principles presented in the ADL Sharable Content Object Reference Model (SCORM). It was thought that applying these disciplines to smaller EPSS units, similar to a SCO (Sharable Content Object), could help address many of the reusability issues. Furthermore, by coupling the development and maintenance of the EPSS content via a Learning Content Management System (LCMS), the IRS could avoid many of the maintenance challenges it was increasingly facing. This was the genesis of the IRS "SPO."

A Sharable Performance Object (SPO) is defined as "a reusable, sharable software entity that can be discretely identified, uniquely distinguishable, and contains consistent attributes and behaviors to provide the necessary information, data, and references to support human performance." Similar to SCORM, SPOs are then contained in Activity Packages that provide a collection of reusable tasks and assets in a standard way to exchange." These Activity Packages and SPOs were then used as the templates for facilitated EPSS development using the IRS Prototype LCMS.

While the concept was defined, the IRS needed an actual EPSS model to examine and decompose into specific SPOs. For this purpose, they deconstructed the e-ARG and e-ACSG static support pages. These are currently in use within the IRS for supporting call center operators. These support pages are basically "electronic documents" that help navigate the call center operator quickly to a requested response. This "eGuide" became the initial EPSS model to decompose into SPOs.

In examining the e-ARG and e-ACSG, common lists of activites were identified, defined, and later formed the core objects for the SPOs (See below):

Table 1 – List of EPSS Activities

- Action	- Decision Table	- Information
- Activity List	- Explanation	- Reference
- Alert	- Fill-in Form	- Selection
- Calculator	- FAQs	- Toolbar
- Checklist	- Illustration	
- Checklist	- Index	

Meeting the Challenge

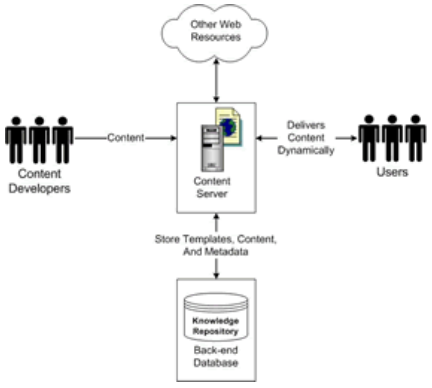

Figure 2 – A Facilitated Approach to EPSS Content

[4] Cascading Style Sheets (CSS) are a simple mechanism for adding style (e.g. fonts, colors, spacing) to Web documents, giving both Web site developers and users more control over how pages are displayed.

Figure 2 (page 4), illustrates a more modern approach to developing and maintaining Web content. It shows the business rules, data, and its presentation stored centrally and separately, and can therefore be maintained once and used in any number of ways via a Relational Database Management System (RDBMS) such as Oracle or MS SQL.

THE INTERNAL REVENUE MANUAL: THE TRUSTED SOURCE

For the IRS, aside from the U.S. Tax Code, the Internal Revenue Manual (IRM) is the most authoritative source for IRS policies and procedures. The IRM consists of the following:

- Policy statements that direct personnel in the administration of the tax laws.
- Procedures and guidelines for how best to serve taxpayers in administering the tax laws.

The IRS adheres to strict procedures for adding to, updating, and communicating the IRM. This print-based publication is overseen by the Service-wide Policy, Directives and Electronic Research organization. Since it has been in use for some time, the IRM is maintained in Standard Generalized Markup Language (SGML) for primarily print purposes. While the Service is now moving towards XML rather than SGML, the EPSS initiative was forced to perform small, static exports for purposes of the Prototype.

With the static extracts exported, the primary focus of the EPSS prototype was to demonstrate a viable solution with a Commercial off the Shelf (COTS) LCMS product, while minimizing development and customization. To help scope the initiative, the EPSS Team focused on the Failure to File/Failure to Pay (FTF/FTP) penalty portion of the IRM (Section 20). This was identified as a common function from the Electronic Accounts Resolution Guide (e-ARG) e-Guide that was easily identifiable and provided the best fit to the SPO definition.

FACILITATED EPSS CONTENT DEVELOPMENT

As mentioned earlier, the FTF/FTP content was decomposed into discrete components and then mapped to SPOs and Activity Packages. The designated SPOs were then mapped to a comparable LCMS object that could be demonstrated within the application. Finally, the e-ARG content was reassembled inside the LCMS using the newly formulated SPO-to-LCMS object mappings.

The flow of information presented within the existing E-ARG e-Guide was reproduced within the LCMS by manipulating the customized object hierarchy. Further customization was added to certain SPOs to reproduce the decision branching and application navigation reflected in the e-ARG e-Guide.

Once development was complete, the content and its inherent informational flow were exported from the LCMS in a static HTML package. This first version of the prototype contained all of the FTF/FTP content within the content frame of the prototype. Navigation within this content frame was performed by using the linking functionality currently available within the LCMS.

Table 2 (below) illustrates the SPO definition mappings to their corresponding LCMS objects:

Table 2 – SPO Mapping

Shareable Performance Object (SPO)	>> MAPS TO >>	LCMS OBJECT
Activity	>> MAPS TO >>	Course or Topic
Decision	>> MAPS TO >>	Description
Decision Table	>> MAPS TO >>	Table
Explanation	>> MAPS TO >>	Summary
Reference	>> MAPS TO >>	Reference Object
Task	>> MAPS TO >>	Module

Content Development

This effort required modifying the LCMS application environment within the Prototype instance. This also required the modification of the e-ARG graphics to map into the LCMS template. The LCMS application required modifications to its LCMS objects and themes to mimic the look and feel of the e-ARG accurately.

In an effort to mimic the E-ARG e-Guide more closely, a second version of the prototype was also generated. Additional customizations to the exported package were required to reproduce the same hyper-linking used in the original e-ARG. This did not change the functionality of the prototype, but was intended to demonstrate how the prototype could better mimic the look and feel of the e-ARG e-Guide. Although the product did mimic e-ARG more accurately, the team recognized the benefit of the first choice because of the significantly better Return on Investment (ROI) of avoiding the mostly cosmetic alterations that had to be performed manually.

Prototype Architecture

The prototype architecture used for the development and demonstration was comprised of the following standard COTS software components:

- File Server
- Web Server
- Learning Content Management System
- Database Server

Potential Production Environment Architecture

Based on the Team's experience, a potential production/demonstration environment based on the prototype might consist of a dedicated application and database server containing the LCMS application and an RDBMS (such as Oracle). End users would access the environment through a Web browser to access the e-Guide created by the LCMS. Developers would access the application using their browsers to modify content or the LCMS application as appropriate (See Figure 3, below).

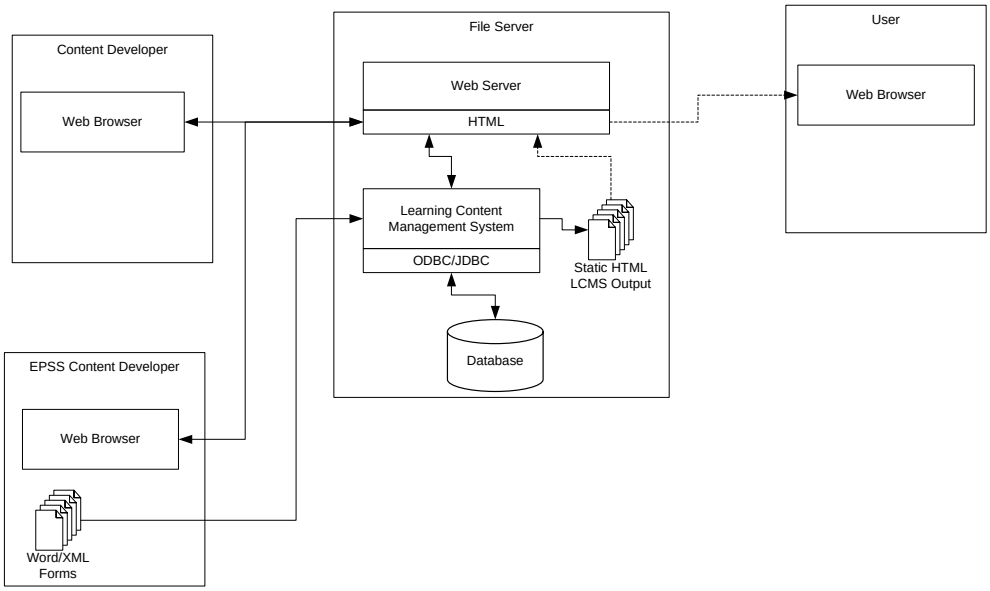

Figure 3 – Potential EPSS Architecture

Content development and maintenance could also be accomplished using a standard MS Word template to import data into the LCMS. Eventually, this could be done via XML using the IRM as a "trusted source" for content updates. These updates could then be performed by non-technical personnel making updates to the IRM, or to an applicable e-Guide. LCMS developers would require more specialized training for specific updates to the LCMS environment to enhance the look and feel of the application as necessary.

Throughout the initiative, various members of the team investigated a variety of technologies and software packages regarding maintaining the necessary content structure and meta-data Service-wide. The IRS currently has a structured authoring process in place for developing new IRM content. This process can be used to create an IRM as a "trusted source" to distribute IRM content if it can follow certain interoperability standards such as Extensible Markup Language (XML).[5]

There were also significant discussions related to following standards for the meta-data associated with SPO definitions and using the IRM as a "trusted source." An initial meta-data schema was defined by the working group during prior working sessions, using Metasoft. Part of the discussion related to the use of other existing meta-data schemas such as S1000D to facilitate interoperability and limit re-work for the EPSS effort in a future state environment.

The prototype was able to successfully demonstrate that a meta-data model can be used to define sections of the IRM and utilized by third party tools. This was accomplished by using the initial meta-data schema and is limited to information related to the FTF/FTP Section 20 of the IRM. The sections below detail the results of the prototype effort related to the standards and technologies discussed by the working group.

S1000D

S1000D is an international specification for the procurement and production of technical publications.[6] Although its stated use applies to technical publications, it has been found through application that the principles of the specification can be applied to non-technical publications. This specification has been initially developed by the Aerospace and Defense Industries Association of Europe (ASD) [former European Association for Aerospace Industries (AECMA)]. This issue has been jointly produced by ASD and the Aerospace Industries Association of America (AIA), who form the Technical Publications Specification Maintenance Group (TPSMG) to establish standards for documentation agreed by the participating nations.

Figure 4 (below) shows the intended correlation among an Object-Oriented training type, an EPSS content type, and an S1000D content type. The idea behind all three is to create reusable core content broken into the smallest logical unit of instruction, support, or data.

Figure 4 – Object-Oriented Content Type Mapping

A key concept of S1000D is the Common Source Database (CSDB) which manages the all the information necessary to produce a publication. The CSDB is designed to:

- Support the technical publication process
- Support the controlled authoring
- Support the QA process
- Support the data exchange with partners, suppliers and customers
- Support delivery of technical publications on various media independent from the source storage format[7]

Output from the CSDB is intended for print or for interactive electronic presentation (IETP). Output can be generated in either SGML or XML, depending on the destination media. The CSDB and data module concept closely aligned with the SCO/SPO concepts used by the IRS.

After researching S1000D, it was determined that using this standard was possible, but additional research was required in the next phase of the project. This is especially true that both SCORM and S1000D have

[5] Extensible Markup Language (XML) is a subset of Standard Generalized Markup Language (SGML), a standard for how to specify a document markup language or tag set (used extensively by the IRM Team for document creation), constituting a particular text markup language for interchange of structured data.

[6] See http://www.s1000d.org for more information.

[7] From the S1000D specification, Chapter 4.2

agreed to develop the two respective standards in a cooperative fashion from this point forward.

Technologies and Conclusions

Throughout the initiative, the IRS Modernization & Technology Services (MITS) strongly encouraged cooperation between the EPSS effort and the chosen Service-wide XML repository package – Digital Concepts' Metasoft Semantic Registry. Since the Prototype was standards-based rather than product-specific, future EPSS initiatives within the IRS will be implemented using the approved enterprise technologies.

In addition, the initiative also identified the importance of the "Web Service" model as a best practice in Web design. This helps to move the Service away from ad-hoc Web site development, saving maintenance time and resources. Similarly, and in cooperation with the Co-Lab, efforts such as the Content Object Repository Discovery and Registration/Resolution Architecture (CORDRA) will remain important to ongoing EPSS and related initiatives at the IRS. Designed to be an enabling model to bridge the worlds of learning content management, delivery, content repositories, and digital libraries, CORDRA aims to identify and specify (not develop) appropriate technologies and existing interoperability standards that can be combined into an enterprise-wide reference and content discovery model. Regardless of the technology, these themes will continue to be important to the effectiveness and value of EPSS and Knowledge Management within the IRS.