The Bridge Project: Open Source Transformation from S1000D™ to SCORM®

Wayne Gafford
Advanced Distributed Learning Initiative
Alexandria, VA
wayne.gafford@adlnet.gov

Tyler Shumaker
Concurrent Technologies Corporation
Johnstown, PA
shumaket@ctc.com

ABSTRACT

S1000D is an international specification for the production and management of technical publications. The specification is based on XML and is primarily used for representing information for operating and maintaining air, land and sea systems (military and commercial). Over the past two years, an OSD-sponsored Reduction of Total Ownership Cost (RTOC) program managed by Advanced Distributed Learning (ADL) has defined a standards-based approach for bridging the technical data and the learning communities. Known as the S1000D Transformation Toolkit, it leverages S1000D to improve learning content lifecycle management for the DoD. One component of the Bridge Project is the transformation of S1000D learning content into Sharable Content Object Reference Model (SCORM) conforming content packages.

This transformation has taken the form of an open-source toolkit available on Source Forge. The significance of this shift is that the management of learning data is refocused from SCORM content packages to authoritative information modules managed within a common source database (CSDB). The CSDB contains all technical information that supports an air, land or sea project. The open-source toolkit can therefore transform both technical information and original learning content from S1000D to SCORM. This significant development creates a clear distinction between SCORM content as an output format and authoritative source information in XML-based structured markup.

This paper discusses the open-source toolkit's architecture and functionality. The paper finishes with a discussion of the lessons learned from the project, the future direction of the toolkit and possible impacts on distributed learning.

ABOUT THE AUTHORS

Mr. Wayne Gafford combines a background in teaching and education with his experience in XML-based standards that has resulted in innovative ideas for e-learning content management and data interoperability. For the last five years, Mr. Gafford has led subcommittees, studies and projects that explore how learning content can benefit from structured markup and lifecycle support. Results have led to an increased awareness that standardized metadata and XML structure can unify diverse, but related content that supports common systems, procedures and products. Mr. Gafford has taken his research to the Advanced Distributed Learning Initiative where he leads the Lifecycle and Configuration Team. He is the government co-chair of the S1000D Learning Standards Harmonization Task Team, is an active public speaker at S1000D and ADL events, and is a supporter of developing XML schemas that model instructional development and learning content to improve knowledge management and distributed learning. He has a Bachelor's degree in English from San Diego State University and a Masters in Education from Marymount University in Arlington, Virginia.

Mr. Tyler Shumaker is a Software Engineer at Concurrent Technologies Corporation (*CTC*) where he is responsible for research and development in the learning, education, training and performance support domain. He has been supporting this domain for over six years. Mr. Shumaker is the technical lead of the S1000D Transformation Toolkit project. He has a Bachelor's degree in Computer Science from the University of Pittsburgh.

The Bridge Project: Open Source Transformation from S1000D™ to SCORM®

Wayne Gafford
Advanced Distributed Learning Initiative
Alexandria, VA
wayne.gafford@adlnet.gov

Tyler Shumaker
Concurrent Technologies Corporation
Johnstown, PA
shumaket@ctc.com

INTRODUCTION

Elegant technical content management strategies make clear distinctions between content in authoring formats and the same content delivered in published formats. Managing content during development is distinct from managing content during distribution. Recognizing the separation of the two states of content enables content managers to choose the right format for production and the right format for distribution. The question then becomes how to transform the "source" format into the "distribution" format at the time of publishing while also supporting Department of Defense (DoD) data lifecycle management requirements.

The answers are not with Microsoft Word, Powerpoint and HTML formats. They provide single, all-purpose data formats while not supporting DoD lifecycle data requirements, such as mapping learning content to equipment. The answer is in the selection of data formats that link training content to the systems it supports, then building tools that enable management and publishing processes that contribute to readiness.

In today's DoD environments where, for example in the Navy, two-thirds of curricula maintenance is driven by equipment changes (Naval Education Training Command, 2009), formats must include metadata that associate technical training content with the systems they support. Without these links, changes and deliveries are protracted up to 18 months after systems are fielded. Microsoft Word, PowerPoint and HTML fall short of lifecycle management objectives and capabilities.

The Bridge Project and the S1000D Transformation Toolkit

System-aware data formats are directly linked to successful lifecycle management practices. The Bridge Project, an Office of the Secretary of Defense (OSD)-funded project managed by the Advanced Distributed Learning Initiative, endeavors to improve lifecycle management of technical training content. A 2009 Navy Inspector General Report on the State of Computer-based Training (CBT) said, "We found minimal governance or standardization for the acquisition, design and development for the lifecycle management of CBT curricula." (Callahan, et al., 2009)

The Bridge Project introduces standardization for CBT production and lifecycle management by using the S1000D technical data specification to manage learning content. S1000D is a modular, XML-based technical data development and management format that also supports technical training content. The S1000D Transformation Toolkit is one component in the Bridge Project. It is an open source application that transforms S1000D content into SCORM, mobile devices and PDF. The significance is that the management of technical learning data is refocused away from SCORM content packages and other distribution formats to authoritative source formats replete with system metadata managed within a common source database (CSDB). Necessary training content changes can be readily identified as technical systems rapidly evolve.

There are other Bridge Project tasks that target gaps preventing enterprise-based training content management. They include connecting learning content development tools to authoritative source databases, developing S1000D support in training development systems, and discovering all modular learning data linked to system design changes.

PROBLEM STATEMENT

There is no interoperable publishing activity that supports the building of multiple delivery platforms from lifecycle-aware learning content formats.

The Toolkit is an open source design that specifically targets developers of technical learning content who have the following requirements: 1) to manage lifecycle and configuration of technical training content; 2) to output that content to multiple outputs (e.g., SCORM and mobile devices; and 3) to use open source technologies.

WHY A TOOLKIT?

The Toolkit provides an optimal development opportunity for a suite of data transformation functions to be shared within DoD. It can contain layers of operational support shared by diverse programs that must lifecycle-manage training content. It also provides support for programs to choose S1000D as a format for technical content of any kind for any project. Most important, the Toolkit design is based on the fidelity, of the S1000D specification making it a truly interoperable application.

Toolkit Applicability

The initial Toolkit version targets the publically available S1000D sample bike course located at www.adlnet.gov as the proof of concept. S1000D-converted legacy training content from the Littoral Combat Ship Mission Module (LCS MM) program has also been transformed to SCORM output successfully through the Toolkit.

All S1000D 4.0 learning data modules schema objects are supported, including learning plans, curricula and assessments. The toolkit was created as an open source framework and reference implementation so the training community can contribute to its functionality through enhanced formatting, support for multiple delivery outputs and support for multiple versions of S1000D. The intent of the Toolkit is to:

- Help make the use of S1000D for technical training content more viable.
- Optimize the production process for content reuse.
- Transform S1000D learning and technical data modules into SCORM, mobile web application and PDF formats.
- Assist with lowering the barrier of adoption of the integration of S1000D, SCORM and mobile platforms.

As more training content is identified, broader Toolkit support will be introduced. The broadest talking point about the Toolkit fits into an overall philosophy about technical data management: technical training data can be part of acquisition strategies, and product support can be designed with a shared specification in mind.

S1000D TRANSFORMATION TOOLKIT

The Toolkit is a black box that consumes S1000D input files and graphics from a Common Source Database (CSDB) and produces multiple outputs. The input files consist of two different types of S1000D information, i.e., organizing content and technical content. The organizing content takes the form of a S1000D SCORM Content Package Module (SCPM). The SCPM organizes training information into clusters of modules that represent the structure of a course or learning event. The technical content takes the form of a Learning Data Module (LDM). The LDM contains a reusable piece of content that can be tracked and aligned with deployed systems.

S1000D SCORM Content Package Module

The SCPM is the primary input to the Toolkit's transformation processes. During these processes, the SCPM is mapped and transformed to the IMS Manifest in the SCORM content package. The advantage to having a single source content organization file in the form of the SCPM is the ability to apply critical configuration and lifecycle metadata to clusters of file references. These clusters can be mapped to learning objectives and to learner profiles.

S1000D Learning Data Module

During the transformation processes, the LDMs referenced in the SCPM are mapped and transformed to HTML and Adobe Flash. The advantage to having a single originating source content format in the form of the S1000D LDM is the ability to apply the same metadata objects, referencing mechanisms and file naming conventions to all related technical content in a project.

S1000D and the Toolkit offer a consistent method for systems-based training developers to create traceable

and configurable technical learning content grouped into learning modules.

Multiple Platform Support

The Toolkit publishes to multiple platforms. This enables instructional designers to develop a greater range of learning and support strategies and allows authoring tools, CSDB environments and other applications a wider array of output options. In turn, this makes it attractive for organizations to fully adopt S1000D as a governing content management strategy.

The Toolkit is written as a Java-based library. Because Java is portable, the Toolkit is able to run on any supported hardware/operating-system platform. Java's portability allows for any authoring tool, CSDB environment or independent application to easily integrate the Toolkit. A Java-based Toolkit provides the most widely supported solution across all platforms to lower the S1000D adoption barrier.

The Toolkit's mobile web application output is wrapped in the jQuery mobile framework. It is designed to provide a web application that will work on most Smartphone and tablet platforms. The iPhone, iPad and Android-based platforms are the platforms currently supported.

Extensibility

As the Bridge Project output requirements evolved, extensibility became an important architectural feature of the Toolkit. With extensibility, any organization that implements the Toolkit will have the ability to add new features to support specific needs. As the technical learning community evolves on how S1000D data is utilized in learning, education and training contexts, the Toolkit will be poised to evolve as well.

The Toolkit supports extensibility by utilizing the Apache Commons Chain Library. It provides an Application Programming Interface (API) that facilitates the Chain of Responsibility design pattern. The Chain of Responsibility design pattern decouples the sender and receiver of a request to maximize programmatic adaptability.

The Toolkit's architecture utilizes the Chain API by building a series of "commands" that are combined into a "chain". **Figure 1** depicts at a high level how each desired output from the Toolkit is achieved by executing a specific "chain" that contains a set of unique "commands". Each rectangle in the figure is an object that shares a common interface handling request. A utility package is provided to supply classes and functions used across various commands.

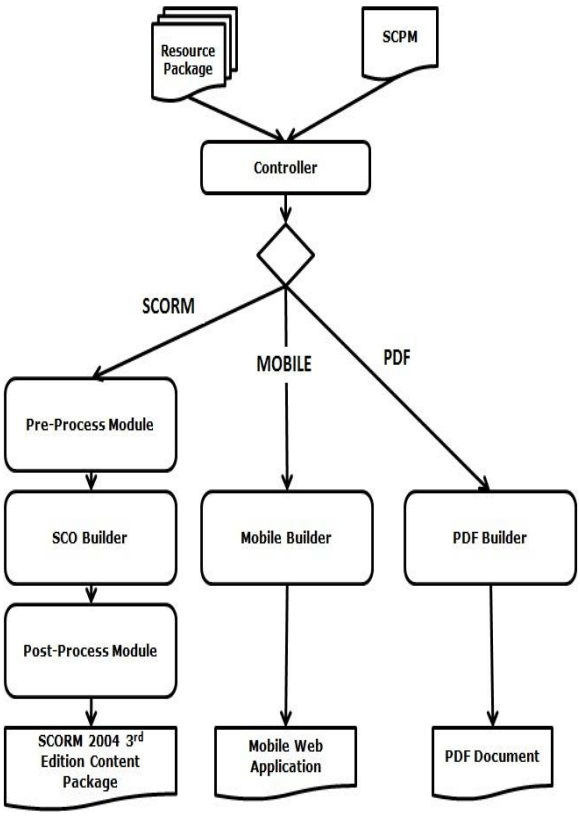

Figure 1. Conceptual Toolkit Architecture

Each command includes an execute() method that is passed a "context" object. The context contains the parameters required between the commands. The execute() methods in each command return a Boolean value (true or false) that determines whether or not the processing of the current chain is complete (true) or if the next command in the chain should be executed (false). The Chains provide an elegant way to handle exceptions that may occur in the Toolkit. If an exception occurs causing the Toolkit to fail in the current chain, then the execute() method for that command returns a "true". This halts the remaining commands in the chain from executing and display the appropriate information about the exception to the Toolkit user. Chains provide design advantages for emerging technologies by making the pattern type flexible for user-defined, context-specific request definitions.

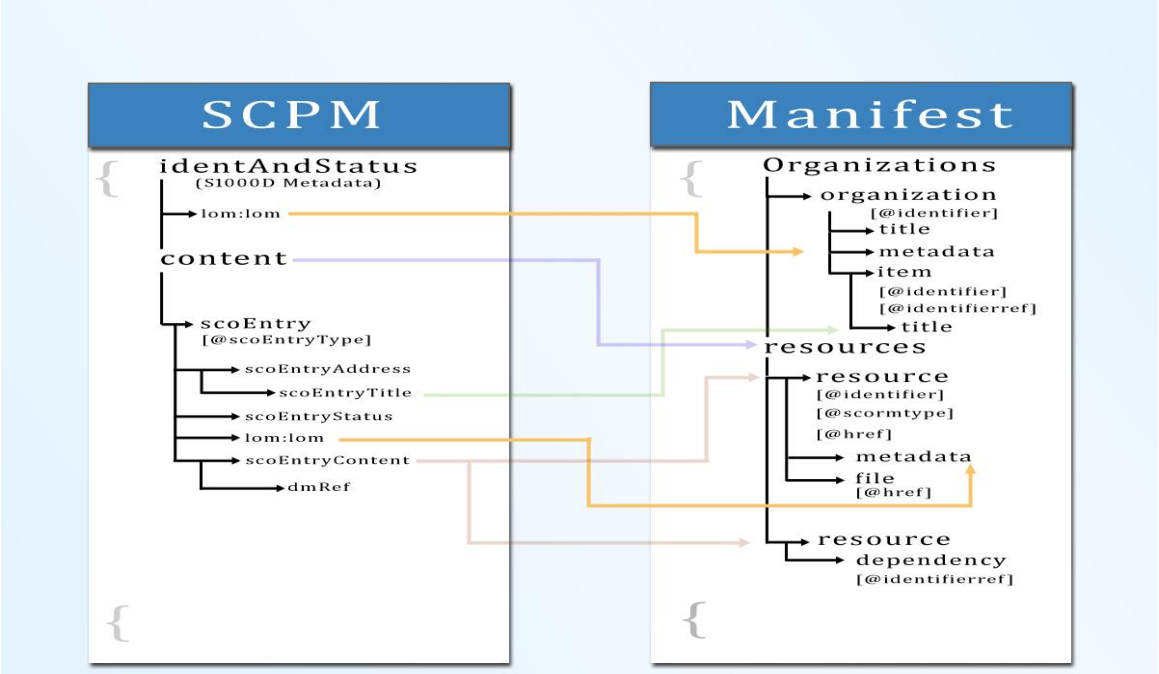

Figure 2: SCPM Mapped to the IMS Manifest

The Toolkit Transformation Process

The transformation from the S1000D SCPM to the IMS Manifest file in the SCORM content package and the transformation from a SCPM to a mobile web application menu page are executed using Extensible Stylesheet Language Transformations (XSLT) files. **Figure 2** depicts the XSLT mapping from SCPM objects such as titles, metadata and referenced files to specific structures in the IMS Manifest file. The XSLT files provide one implementation for the SCPM transformation. The XSLT files also provide an intersection point for the SCPM transformation to be customized.

The toolkit supports all XML objects declared in the SCPM and LDM schemas. Additionally, those technical and procedural data module objects reused in the sample S1000D bike course are also supported. **Figure 3** depicts the sample Toolkit interface. It includes ability to choose an SCPM, an S1000D resource package and an output format. During transformation process, runtime and error logs are displayed in a viewing window. If desired, the sample interface can be customized and integrated with any existing system.

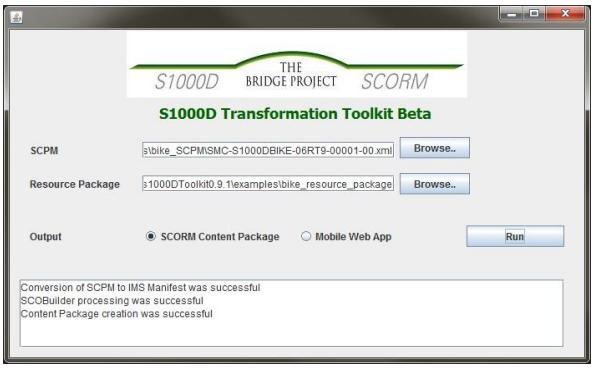

Figure 3: S1000D Toolkit Interface

Viewing S1000D Data Modules in a SCORM LMS

After the Toolkit completes the transformation processes, the S1000D output can be viewed through a Learning Management System (LMS), iPhone, iPad and Android-based device. The Toolkit provides a Viewer Application that contains a rendering engine used for SCORM output. This rendering engine provides an example of how S1000D learning content can be rendered in a Web browser. The Toolkit is designed to allow other Viewer Applications to replace the one initially provided. Along with the XSLT files, the Viewer Application includes:

- a SCORM JavaScript API scripting wrapper file,

- a JavaScript file used to support navigation,
- HTML files to provide the viewing page's header, content, and footer,
- basic image files for navigation buttons, headers and footer,
- Flash files for the handling of interactive assessments, and
- Cascading Style Sheets (CSS) used for look-and-feel and formatting of the content.

Figure 4 demonstrates an example SCORM output.

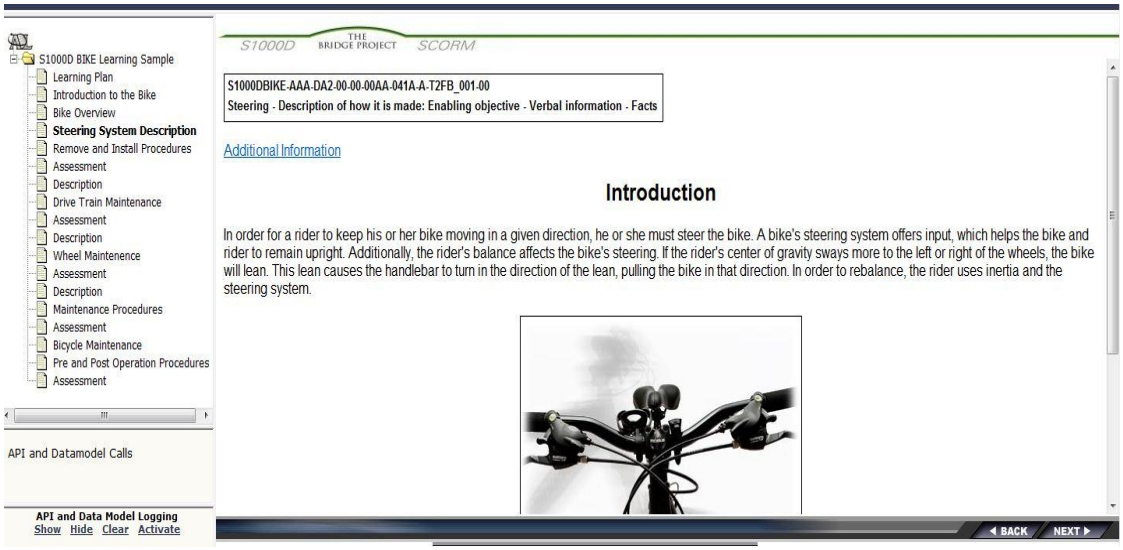

Figure 4: Sample S1000D Bike Courseware in a SCORM LMS

All of the files included in the Viewer Application are available to be modified by Toolkit implementers. This gives organizations the ability to add their own graphics in the headers and footers. Developers can also modify the CSS and XSLT files to change the content views and styling. Organizations will then be able to use the Toolkit to produce project-specific SCORM Sharable Content Object (SCO) files in the SCORM Content Package output.

Viewing S1000D Data Modules in Mobile Devices

The mobile web application produced by the Toolkit, reuses some of the same XSLT files that are used for the SCORM output. A few mobile-specific XSLT files are used for handling the graphics and references when generating the mobile output. All of the HTML pages generated for the mobile web application output are wrapped in the jQuery mobile framework. The jQuery mobile framework offers some different options to provide customizable themes to the mobile web pages. CSS3 properties, color "swatches" and simple icon sets can be used to create many different designs. By modifying the XSLT files that generate the jQuery mobile output, a Toolkit implementer can create custom-designed mobile outputs for their organization. **Figure 5** and **Figure 6** depict S1000D output to iPhone and Android-based devices.

Figure 5: S1000D Bike Content on iPhone

Figure 6: S1000D Bike Menu on Android

WHY OPEN SOURCE?

To leverage community input and to lower the barrier to adoption of integrated S1000D learning content, the Toolkit was developed as an open source project. SourceForge.net was chosen as the source code repository to host the Toolkit project in order to manage the development process,

The open source project gives implementers the greatest degree of freedom for Toolkit adaptability. Because the integration of S1000D and SCORM is new to technical publication and computer based training (CBT) communities, the open source project provides a way for peer-to-peer quality assurance and feedback. Peers in these communities can validate the quality of the Toolkit and source code while providing bug fixes and recommended features.

An open source solution offers organizations that manage and produce S1000D learning data more control over how quickly bugs are corrected and how quickly features are added. A proprietary S1000D learning data transformation product would require the original developers to make bug fixes, add features and add costs. This could take a long time and be expensive for organizations that implement a proprietary system. SourceForge.net offers bug tracking and feature request tracking that organizations can utilize to identify any bugs and collaboratively establish new requirements.

Open Source License

The open source licenses of all the source material allow for derivation and are free to use in any product that manages or produces S1000D data. The Toolkit source code and associated files are available to be customized by anyone. Organizations that maintain these products can choose to implement the entire Toolkit source code or choose to utilize specific pieces. With the Toolkit being open source, it allows for easy translation of all code. S1000D is an international specification and the ability to have multiple language support is important.

As with the acclimation to any new technology or philosophy, learning by example is one of the most effective ways to comprehend something new. Organizations in the technical publications and e-learning communities will be able to utilize the Toolkit as a reference implementation of the integration of S1000D and SCORM. The Toolkit provides example mappings between S1000D and SCORM, S1000D SCPM and learning data modules, plus how to view the S1000D learning projects in various outputs.

LESSONS LEARNED AND LIMITATIONS

Several lessons were learned about S1000D and SCORM integration as the Toolkit development progressed.

Limited Hierarchical Content Structure

The S1000D 4.0 bike learning data modules were used as the base use case for the Toolkit development. The content package amounted to 17 sharable content objects (SCOs). A S1000D 4.0 SCPM limitation is a limited ability to create hierarchical content structures. The S1000D 4.0 SCPM offers a flat structure for the breakdown of the nested modules that are mapped to the IMS Manifest files as SCOs. This creates no interoperable way to define a S1000D 4.0 SCPM the can be mapped to an IMS Manifest file with a parent/child structure that is common in SCORM 2004 Content Packages. This deficiency has since been corrected in S1000D 4.1.

In S1000D 4.1, the SCPM is divided into two schemas to promote extended SCO reuse and nesting. To put it simply, the equivalent of the IMS manifest “resources” portion of the SCPM is now a separate schema from the equivalent of the IMS manifest “organization” portion of the SCPM. The “organization” schema keeps the name “SCPM”. The “resources” schema is named “SCO DM”, or SCO data module. The SCPM structure ends with a reference to a SCO DM identifier. SCO DMs will be managed as separate, reusable objects.

Limited Data Set

The lack of S1000D 4.0 learning data modules that were available for testing also caused an issue when developing the Toolkit. The only test data that was available was the S1000D bike sample data and some S1000D-converted legacy training content from the LCS MM program. All of these learning data modules were developed by the same authoring group and limited the scope of support that the Toolkit could initially provide. As more S1000D training and learning content is developed and available for testing, additional support could be integrated into the Toolkit by the open source community.

The Value of S1000D File Naming

As the mapping of the S1000D SCPM and learning data modules to the SCORM Content Packaging design was being analyzed, it became apparent that the data module and file naming structure presented much value to the Toolkit. Each data module is named by a specific structured identifier known as a data module code (DMC). The DMC consists of configuration information about the system being supported, basic maintenance activity to be performed on the system, and the type of instructional design and human performance strategy used in the development of the content. This information provides the ability for the Toolkit to identify content by the information type and filter content based on DMC information in support of any organization's needs.

Tracking Performance in Mobile Devices

The Toolkit development team determined that the mobile output is better suited for performance support material. Assessments from the S1000D bike data could not be included in the mobile web application output because there is a solution for tracking scores and completion status consistently across mobile devices does not currently exist. The mobile output simply provides the material needed to learn how to maintain and operate a bike. As interoperable learning tracking solutions become available for mobile devices, the Toolkit will be poised to support assessments.

CONCLUSION

An S1000D approach to enterprise learning content management is a key strategy to improving governance and standardization for the acquisition, design and development for the lifecycle management of CBT curricula. New editing, publishing and search functions will require flexible applications that serve multiple content types and delivery platforms. Most important, source content and source collections will be configured to the systems they support. The Toolkit is a key publishing ingredient for programs that discriminate between formats that manage content and formats that deliver content. It is an interoperable publishing activity that supports the building of multiple delivery platforms from a lifecycle-aware learning content format.

The Toolkit is an emerging technology in support of an emerging philosophy that technical learning content should be managed as a business asset integrated with other information products in an industry-based, technical data specification.

REFERENCES

Apache Commons, (2010). http://commons.apache.org/chain/

Bango R., et al., *jQuery Mobile Framework,* http://jquerymobile.com/

Callahan K, et al., (2009). *NAVIGSGEN Computer Based Training Study*.

Hammerly M., Shumaker T. (2011). *S1000D Transformation Toolkit Documentation.*

Naval Education Training Command, (2009). *Computer-Based Training & Personal Computer-Simulation Prioritization and Cost Estimation Assessment for NETC COO Supporting FY10/11 Spend Plan and POM12 Submittal*.

Wilson J., (2011). *Benefits of open source code,* http://www.oss-watch.ac.uk/resources/whoneedssource.xml