

Balisage 2021 Conference Program

Contents

Vision.....	4
Mission.....	5
8-02-10:00. Compatibility	6
8-02-11:00. XQuery	7
8-02-12:00. Element Order	8
8-02-13:00. Conversation	9
8-02-14:00. SGML.....	10
8-02-15:00. SGML & TEI-XML	11
8-02-16:00. Presentational Markup.....	12
8-03-10:00. XSLT.....	13
8-03-11:00. Text-As-Graph	14
8-03-12:00. JATS.....	15
8-03-13:00. Docugami	16
8-03-14:00. Query Language.....	17
8-03-15:00. Mathematics Curriculum.....	18
8-03-16:00. Change Tracking	19
8-04-10:00. Models.....	20
8-04-11:00. String Comparison.....	21
8-04-12:00. Literary Texts	22
8-04-13:00. Editing	23
8-04-14:00. PubMed	24
8-04-15:00. Citations	25
8-04-16:00. Scriptural Markup	26
8-05-10:00. XProc & XSLT.....	27
8-05-11:00. Manuscripts	28
8-05-12:00. XML & CSS.....	29
8-05-13:00. Balisage Bard	30
8-05-14:00. XForms.....	31
8-05-15:00. STAR	32
8-05-16:00. Web Semantics	33
8-06-10:00. TEI Text	34
8-06-11:00. XML Strings.....	35
8-06-12:00. CSS & XSLT.....	36
8-06-13:00. Sponsor Presentation	37
8-06-14:00. SVG Text.....	38
8-06-15:00. Complexity.....	39
Administrative Information.....	39

DEMONSTRATION ONLY



Balisage 2021: The Markup Conference (BS2021)

Description:

Balisage is an annual conference devoted to the theory and practice of descriptive markup and related technologies for structuring and managing information.

Stakeholder(s):

Markup Practitioners :

Balisage: where serious markup practitioners and theoreticians meet every summer. If you are happy to be the person in your project who understands the angle brackets and stuff, then you are a markup geek and Balisage is the place for you. Even if you are NOT a markup geek, if you find it instructive to spend time with them now and then, you will enjoy Balisage.

Balisage 2021 Conference Committee

B. Tommie Usdin :

Chair / Mulberry Technologies

Deborah A. Lapeyre :

Co-Chair / Mulberry Technologies

James David Mason :

Co-Chair

C. M. Sperberg-McQueen :

Co-Chair / Black Mesa Technologies

Norman Walsh :

Co-Chair / Saxonica

Balisage 2021 Advisory Board

Syd Bauman :

Northeastern University

Jeff Beck :

National Library of Medicine

David J Birnbaum :

University of Pittsburgh

Jon Bosak

Robin Cover :

OASIS

Steve DeRose :

Independent Consultant

Bob DuCharme :

CCRI

Patrick Durusau

Eric Freese :

TopQuadrant

Eduardo Gutentag

G. Ken Holman :

Crane Softwrights

Sam Hunting

Michael Kay :

Saxonica

David A. Lee :

Nexstra, Inc

Chris Lilley :

World Wide Web Consortium

Yves Marcoux :

Université de Montréal

Sean McGrath :

Propylon

Mary McRae :

IQ Solutions

Steven R. Newcomb :

Coolheads Consulting

Wendell Piez :

Piez Consulting Services

Ari Nordström :

Creative Words

Liam Quin :

Delightful Computing

Allen H. Renear :

University of Illinois at Urbana-Champaign

Bruce Rosenblum :

Inera

— continued next page

Stakeholders (continued)

Jeni Tennison :
Jeni Tennison Consulting

Henry S. Thompson :
University of Edinburgh

Fabio Vitali :
University of Bologna

Lauren Wood

Balisage 2021 Sponsors

Docugami :
Gold Sponsor

Antennahouse :
Silver Sponsor

Oxygen XML Editor :
Sponsor

Text Encoding Initiative :
Co-Sponsor

Washington Area SGML/XML Users Group :
Co-Sponsor

The XML Guild :
Co-Sponsor

Balisage 2021 Peer Reviewers

Syd Bauman

Geert Bormans

John Cowan

Angelo Mario Del Grosso

Patrick Durusau

David Farmer

Peter Flynn

Amanda Galtman

Tony Graham

Joel Kalvesmaki

Jirka Kosek

Deborah A. Lapeyre

James David Mason

David Maus

Sheila Morrissey

Ari Nordström

Maria Papadopoulou

Dave Pawson

Wendell Piez

Liam Quin

Georg Rehm

C. M. Sperberg-McQueen

Bethan Tovey-Walsh

B. Tommie Usdin

Norman Walsh

Sam Wilmott

Lauren Wood

Balisage 2021 Sister/Related Conferences

Declarative Amsterdam

XML Prague

Markup UK

NISO Plus Conference

Mulberry :
Producer

Vision

Good, practical theories

Mission

To document and share the agenda for the Balisage 2021 conference.

DEMONSTRATION ONLY

8-02-10:00. Compatibility

Address the problem of making XML backwards compatible as software is upgraded

Stakeholder(s)

B. Tommie Usdin :
Mulberry Technologies

Monday 10:00- 10:30 EDT (+ Q&A 10:30 - 10:45) ~ The (unspoken) XML gotcha ~ XML is a platform-neutral way to exchange, share, and manipulate information. But what persuades many to use XML is the claim that XML provides a long-term way to store information, independent of tools (both hardware and software) with their short life spans. Projects spend significant resources on XML setup and then settle into doing the real work, using that XML infrastructure to compile, write, analyze, or whatever it is they do. Until, one day — something doesn't work. Hardware is retired; software is upgraded; specifications go into new releases. Users get stuck. And when they complain, we respond that "of course that doesn't work any more, you have been accumulating technical debt for years! It is time to reinvest." They thought they had committed to a one-time cost, and now we tell them that it is an ongoing expense. If the user had put documents into their favorite spreadsheet, they complain, they could still import them into the current version. How do we answer that complaint? We (the XMLers) think we described the values of XML plainly and fairly. We (the XML users) think that the claim that XML documents last a long time is relying on a specious technicality, and we have been trapped dishonestly. I live on both sides of this: as a user I want to invest in infrastructure once and have it last; as a developer I want to be able to improve my product without the limitations imposed by backwards compatibility. We as a community often complain that not enough people are using XML. If we really want XML use to grow, we need to address the gotcha that too many XML users are feeling.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-02-11:00. XQuery

Build libraries of useful XQuery algorithms

Stakeholder(s)

Mary Holstege

XQuery Developers

Monday 11:00 - 11:30 EDT (+ Q&A 11:30 - 11:45) ~ Fast bulk string matching ~ XQuery developers deserve to have access to libraries of implementations of useful algorithms. Programmers using other programming languages have such libraries — why not us? For example, it would be nice to have a library implementation of the classic Aho/Corasick algorithm which shows how to search for multiple words — an arbitrarily large set of words in fact — in a single linear pass over the document. It's a useful algorithm with interesting applications. But why wait for someone else to build the libraries we want? Let's build them ourselves. And let's start now.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-02-12:00. Element Order

Consider the significance of element order.

Stakeholder(s)

Robin La Fontaine :
DeltaXML

Monday 12:00 - 12:30 EDT (+ Q&A 12:30 - 12:45) ~ Element order is always important in XML, except when it isn't ~ "Which came first," begins an old joke. But the more interesting question might be, "does it even matter?" There are many obvious and several not-so-obvious ways in which the order of items (be they XML elements or attributes, or JSON maps or arrays) can be understood to be significant or insignificant. These are not new questions and how they're answered plays out across vocabulary design, schema design, and individual documents. They are important questions when it comes deciding if two documents are "the same" or "different" and to what extent.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-02-13:00. Conversation

Meet new people and/or chat with people outside your usual orbit.

Monday 13:00 - 13:45 EDT ~ Small Group Conversation ~ During the mid-day break today is a good time to meet some new people and/or chat with some people who are outside your usual orbit. Everyone who is in this "room" 5 minutes after the start of the mid-day break today will be invited to a breakout room. Groups will be small and randomly assigned. Note: this is an opportunity, not an obligation. Participate if the idea appeals to you; or take a break from your screen if that seems better.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-02-14:00. SGML

Break SGML down into topic-like units.

Stakeholder(s)

Ari Nordström

Monday 14:00 - 14:30 EDT (+ Q&A 14:30 - 14:45) ~ Topic-based SGML? Really? ~ Topic-based applications like DITA are all the rage these days, and DITA is an XML application. But what if your customer's products are tied to industry standards that were created in SGML days? And what if the customer uses a content-management system that the vendor swears will support SGML, so the customer has little impetus for switching to XML? What, then, if the customer wants to publish HTML, and the best way of getting there is through DITA? The answer may be to look for ways to break SGML down into topic-like units. That's a great idea, but what if their SGML is full of features that most XML software vendors have never heard of, such as graphics stored in entities? There are many tricks to pulling this off. The key to our approach is XProc pipelines running dozens of incremental XSLT steps. Climb on for a wild ride!

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-02-15:00. SGML & TEI-XML

Convert SGML hybrids to TEI-XML.

Stakeholder(s)

Tracey El Hajj :
University of Victoria

Janelle Jenstad :
University of Victoria

Monday 15:00 - 15:30 EDT (+ Q&A 15:30 - 15:45) ~ Converting SGML hybrids to TEI-XML: The case of the internet Shakespeare editions ~ In late 2018, the Internet Shakespeare Editions (ISE) experienced catastrophic code failure. In an attempt to preserve the data for future use, a project was launched to convert the ISE from its boutique markup and bespoke workflows to TEI and standard workflows. In this paper, we describe the markup language used by the ISE (known as IML for ISE Markup Language), various fundamental differences between IML and TEI, and the challenging work of converting and remediating the ISE's IML-encoded files. These challenges include not only mechanical issues such as unclosed tags but also logical challenges in finding ways in the new tag set to encode concepts easily encoded in the original. Our central question is how to do this work in a principled, efficient, well-documented, replicable, and transferable way. We conclude with recommendations for re-encoding legacy projects.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-02-16:00. Presentational Markup

Make the recognition of content objects more efficient and reliable

Stakeholder(s)

Allen H Renear :

School of Information Sciences, University of Illinois at Urbana-Champaign

Bonnie Mak :

School of Information Sciences, University of Illinois at Urbana-Champaign

Monday 16:00 - 16:30 EDT (+ Q&A 16:30 - 16:45) ~ Presentational Markup: What's going on? (LB) ~ Presentational markup, the addition of rendering features that identify and differentiate between content objects, is familiar and ubiquitous, with origins possibly coeval with human communication. But it is not at all clear exactly what presentational markup is doing. Exactly how does presentational markup make the recognition of content objects more efficient and reliable? What are the connections with other nonlinguistic contributions to textual communication, ranging from rhetorical style to punctuation?

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-03-10:00. XSLT

Enable users to perform operations on their data, securely, on their own systems.

Stakeholder(s)

Wendell Piez

Tuesday 10:00 - 10:30 EDT (+ Q&A 10:30 - 10:45) ~ Client-side XSLT, validation and data security ~ Client-side XSLT (CSX) is often used in scenarios where data (in XML) from a remote server is provided to a user who processes it in some way, for example rendering it locally for display. That is, the server provides the data and the client does the work on that data to make it useful. However, that is not the only scenario in which CSX is useful. In an environment in which the user already has, or is in the process of creating, XML, CSX can be a convenient and powerful tool, enabling users to perform operations on their data, securely, on their own systems. The potential for this use of CSX is illustrated with uses of Saxon-JS for several security-related applications.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-03-11:00. Text-As-Graph

Support translation of TAGML to XML for those purposes.

Stakeholder(s)

Elli Bleeker :

Huygens Institute for the History of the Netherlands

Bram Buitendijk :

Royal Netherlands Academy for Arts and Sciences

Ronald Haentjens Dekker :

Huygens Institute for the History of the Netherlands

Tuesday 11:00 - 11:30 EDT (+ Q&A 11:30 - 11:45) ~ Hyper, multi, or single? Thinking about text in graphs and trees
~ There are many reasons to choose XML as a vehicle for encoding cultural heritage texts (such as the availability of applications and supporting tools), but there is simultaneously a risk of being boxed in by its model of text. Students of text are all too familiar with issues such as concurrent or overlapping hierarchies, discontinuous text, and non-linear structures that are not well served by the Ordered Hierarchy of Content Objects model. Our work on the Text-As-Graph (TAG) model supports these alternative patterns, and we have working editors and repositories for TAG. Nonetheless, it is useful to be able to export XML to take advantage of the many tools that support analysis and processing. Accordingly, we are building tools to support translation of TAGML to XML for those purposes.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-03-12:00. JATS

Identify commonly used elements and attributes of the JATS Publishing schema

Stakeholder(s)

Gerrit Imsieke :
le-tex publishing services

Nina Linn Reinhardt :
Leipzig University of Applied Sciences (HTWK)

Tuesday 12:00 - 12:30 EDT (+ Q&A 12:30 - 12:45) ~ JATS Blue Lite: The Quest for a Compact Consensus Customization (LB) ~ JATS (the Journal Article Tag Suite ANSI/NISO z39_96-2019) is a tag set of elements and attributes describing the content and metadata of journal articles. Publishers and archives worldwide have used the base JATS schemas as well as made extensive supersets and subsets. For some time, the JATS community has been requesting a new “official” JATS subset that is smaller and simpler than the existing schemas (including both journal metadata and article metadata) and that has been reduced for ease of editorial tool use. This work identifies commonly used elements and attributes of the JATS Publishing schema, through usage statistics of articles from major publishers and repositories, and builds such a minimal subset. While the authors identify a naive minimal customization based on usage, the minimal subset schema has been enhanced to include strategic structures considered necessary. For example, JATS is revised every few years, and the authors have taken into account the fact that items introduced in more recent JATS versions might be strategically critical, even though less numerous. Other functional aspects (such as accessibility, open access, and machine processability) have suggested structures to be retained, despite infrequent current usage. The resulting minimal subset was produced as a JATS DTD customization, significantly smaller than the JATS publishing tag set.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-03-13:00. Docugami

Deconstruct the content and create highly functional semantic XML representations documents

Stakeholder(s)

Jean Paoli :
Docugami

Tuesday 13:00 - 13:45 EDT ~ Sponsor Presentation: Docugami ~ We described in our 2019 Balisage presentation our recognition of what we called “Document Dysfunction” and five principles that can lead the industry to more effective solutions. In this session, we will present and, for the first time, publicly demonstrate Docugami, our answer to “Document Dysfunction”. Docugami is an AI Document Software as a Service that is designed for Business Users. Docugami enables users to point to business documents in PDF (scanned or digital) or doc[x] formats, and without lengthy setup or training, start building reports from existing documents or getting help from Docugami when creating new documents. In the background, Docugami automatically deconstructs the content and creates a highly functional semantic XML representation of each document. This semantic XML representation enables Docugami to share information with line-of-business systems, create new documents coherent with previously created documents, present the information using multiple views and start process automation across an organization.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-03-14:00. Query Language

Design a query language that offers more power than a “bag of words in a search box” without imposing a syntax.

Stakeholder(s)

C. M. Sperberg-McQueen :
Black Mesa Technologies

Tuesday 14:00 - 14:30 EDT (+ Q&A 14:30 - 14:45) ~ Ariadne’s thread: A design for a user-facing query language for texts and documents ~ It is likely that most Balisage attendees are familiar with at least one query language for searching structured documents, for example XPath, and perhaps several others. Domain experts are often much less familiar with query languages of this kind, may find the syntactic requirements awkward, and have been trained by common web search interfaces to think they are unnecessary. This paper explores the design of a query language, Ariadne, that offers much more power than a “bag of words in a search box” without imposing a syntax that’s so unfamiliar it is likely to drive away new users. Ariadne takes inspiration from Arras and DynaText (two interactive search and retrieval systems of the 1980s and 1990s). It has a simple but expressive grammar that is amenable not only to trees but also to other models such as concurrent hierarchies, Goddag structures, multi-colored trees, and even, with restrictions, systems like LMNL.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-03-15:00. Mathematics Curriculum

Atomize digital mathematics learning materials into Learning Objects

Stakeholder(s)

Alan Edward Bickel :
Big Ideas Learning, LLC / Larson Texts, Inc.

Tim Larson :
Big Ideas Learning, LLC / Larson Texts, Inc.

Elisa E. Beshero-Bondar :
Penn State Erie, the Behrend College

Tuesday 15:00 - 15:30 EDT (+ Q&A 15:30 - 15:45) ~ A Linked-Data Method to Organize an XML Database for Mathematics Education (LB) ~ The authors are designing a content-delivery system for a mixture of print and digitized materials (textbooks, teacher materials, tutorials, and assessments) and born-digital mathematics learning materials, all in multiple media formats. The goal is to atomize these resources into Learning Objects, to allow for rapid curriculum customization to suit varied learning contexts. Because there is a fundamentally progressive way in which mathematics is taught, and certain competencies require other prior-knowledge competencies, we propose a Competency Graph. Essentially, a competency graph is a low-level knowledge framework that underpins a state standards set, or a system of mathematical learning objectives classification. Although the competency graph solution is semantic by nature, it is not relational. We have chosen an XML representation of Resource Descriptive Framework (RDF) data. Our plan is to design an XML database storing a set of RDF associations and data pointers that correlate resources, standards, topics, related topics, competency assessment, and usage tracking. The authors have begun drafting RDF ontologies in XML, and are organizing this using eXist-dB. We use XPath and XQuery for fine-grained searching, retrieving, and visualizing networked data. We'd like to share what we have started; we'd like your feedback.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-03-16:00. Change Tracking

Track changes before final XML serialization in a write-only log.

Stakeholder(s)

Patrick Durusau

Tuesday 16:00 - 16:30 EDT (+ Q&A 16:30 - 16:45) ~ Pre-XML document change tracking: Change.log, collaboration, immutability, XML, UUIDs ~ Change tracking in XML documents runs into some of the thorniest and most intractable problems in markup languages, notably those that hover around concurrent and overlapping structures. The tracking mechanisms in the popular open-source packages Apache Office and Libre Office avoid the problems by being lossy. But it doesn't have to be that way! We can solve many of the issues if we push change tracking to just before final XML serialization and track the changes in a write-only change log, mediated with immutable IDs. UUIDs allow us to make those IDs globally unique.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-04-10:00. Models

Challenge preconceived notions about derived data models

Stakeholder(s)

Joshua Lubell :

National Institute of Standards and Technology

Wednesday 10:00 - 10:30 EDT (+ Q&A 10:30 - 10:45) ~ The Model Made Me Do It! A Cautionary Tale From a Security Control Baseline Tool Developer (LB) ~ Even the best written specifications can be complicated documents to read and understand. Normative prose is often supported by tables and diagrams intended to clarify the specification. What happens when those clarifying features can be interpreted as implying a different model than the normative prose intends? What does this say about relying on derived data models in the tools that support the specification? Expect your preconceived notions to be challenged!

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-04-11:00. String Comparison

Develop a new XSLT function built on a non-traditional approach to finding the longest common subsequence or substring.

Stakeholder(s)

Joel Kalvesmaki

Wednesday 11:00 - 11:30 EDT (+ Q&A 11:30 - 11:45) ~ String comparison in XSLT: The tan:diff() function ~ Classical models of string comparison are difficult to implement in XSLT, partly because they are designed for imperative, stateful programming. I propose a new XSLT function, tan:diff(), which is built on a non-traditional approach to finding the longest common subsequence or substring (LCS). For testing tan:diff(), I have assumed two strings of arbitrary length that are roughly similar to each other, typically because one is the result of editorial changes to the other. (Note: Strings can be in any language because the differences are expressed at the character level, necessary since in languages such as Chinese and Thai one cannot assume the use of the space to differentiate words.) Rather than tokenizing (on spaces, for example), gigantic strings are subdivided into large strings, with the segments processed pairwise. The tan:diff(), function extracts series of progressively smaller samples from the shorter of the two texts and looks for a match in the longer one on the basis of an XSLT fn:contains(). When a match is found, the results are separated by XSLT tan:substring-before() and tan:substring-after(). The tan:diff() function is efficient and fast, even on pairs of very long strings (100K to 1M characters), in part because of its staggered-sample approach, in part because of its optimization strategy for long strings.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-04-12:00. Literary Texts

Encode semantic relationships in literary texts.

Stakeholder(s)

Fotini Koidaki :
Aristotle University of Thessaloniki

Katerina Tiktopoulou :
Aristotle University of Thessaloniki

Wednesday 12:00 - 12:30 EDT (+ Q&A 12:30 - 12:45) ~ Encoding semantic relationships in literary texts: A methodological proposal for linking networked entities into semantic relations ~ Encoding meaningful semantic relationships in literary texts is almost as difficult as defining and identifying them. Defining the types and the components of semantic relationships that can be extracted from literary texts is challenging because literature is full of implicit and oblique messages and references. Relations may not have a clear or standard linguistic form and they often overlap. We discuss issues involved in modeling and encoding the mapping of relationships in literary and humanities texts, illustrated by the case of the ECARLE project annotation campaign. We propose using minimalistic and flexible annotation techniques to generate human annotated training data for a Relation Extraction machine learning system. Using the TEI tagset, without customization, we are able to encode the mapping of relations formed by named entities in a simple yet flexible way, open to reuse, interchange, conversion and visualization.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-04-13:00. Editing

Enable users to quickly access the most often used structures and operations to edit XML documents

Stakeholder(s)

Alex Jitianu :
Oxygen

Wednesday 13:00 - 13:30 EDT (+ Q&A 13:30 - 13:45) ~ Sponsor Presentation: Oxygen | Using XSLT and XQuery Update to Define Actions for Editing XML ~ When editing XML documents, we often need to perform more complex operations than inserting/deleting text or XML elements. To maximize efficiency and productivity, we also need to provide users with a set of actions that allow them to quickly access the most often used structures and operations. These actions can be encoded as custom operations specific to an XML vocabulary or to a project and usually the language for defining these actions is the application language, Java, Javascript, etc. Since we are talking about processing XML documents, the best languages for processing XML are XSLT and XQuery! Let's explore: The benefits of using XSLT and XQuery Update to define custom actions. How we can enable operation selection using XPath to dispatch to the correct operation based on the document context. How we can employ XSpec for testing custom actions.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-04-14:00. PubMed

Modernize XML conversion at PubMed Central.

Stakeholder(s)

Martin Latterner :

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health

Dax Bamberger :

Kelly Peters & Jeff Beck, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health

Kelly Peters :

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health

Jeff Beck :

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health

National Library of Medicine

U.S. National Institutes of Health

Wednesday 14:00 - 14:30 EDT (+ Q&A 14:30 - 14:45) ~ Modernizing XML conversion at PubMed Central ~ PubMed Central® (PMC) is a free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine. PMC ingests about 70,000 XML articles a month, using XSLT to perform "schema conversion" to transform from 250 variant schemas to JATS XML (Journal Archiving and Interchange Tag Set, ANSI/NISO Z39.96). In addition to converting from many vocabularies to JATS, PMC needs to perform data normalization, standardizing core metadata to meet the PMC tagging guidelines, which are enforced with a style-checker. The main conversion work is currently performed by over 400 XSLT 2.0 stylesheets (c.200,000 lines of code), one full transformation per source DTD. PMC is working on a proof-of-concept that uses the XSLT 3 fn:transform() function to break this single conversion operation into multiple, discrete transformations that can handle article collections, schema conversion, collection-specific processing, and the data normalization necessary to produce PMC-compliant JATS.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-04-15:00. Citations

Provide a way for documents to declare their own internal structure along with a way to resolve citations conforming to that structure

Stakeholder(s)

Hugh Cayless :
Duke University Libraries

Jonathan Robie :
Clear Bible, Inc

Thibault Clérice :
École nationale des chartes

Wednesday 15:00 - 15:30 EDT (+ Q&A 15:30 - 15:45) ~ Introducing Citation Structures (LB) ~ Text Encoding Initiative documents are notoriously heterogeneous in structure, since the Guidelines are intended to permit the encoding on any type of text, from tax receipts written on papyrus to Shakespeare plays or novels. Citation Structures are a new feature in the TEI Guidelines that provide a way for documents to declare their own internal structure along with a way to resolve citations conforming to that structure. This feature will allow systems like the Distributed Text Services (DTS) API which process heterogeneous TEI documents to handle tasks like automated table of contents generation, the extraction of structural metadata, and the resolution of citations without prior knowledge of document structure.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-04-16:00. Scriptural Markup

Make software deal well both with USFM for scriptural texts and various flavors of XML for scriptural and other materials.

Stakeholder(s)

Jonathan Robie :
SIL International

Bible Translation Community

Wednesday 16:00 - 16:30 EDT (+ Q&A 16:30 - 16:45) ~ Scriptural markup in the Bible translation community ~ Many XML formats have been proposed for scriptural markup, but the format most often used by the thousands of Bible translators actively working today is a non-XML markup language called USFM (universal standard format markers), a backslash-delimited markup language which developed bottom-up as users invented new tags and persuaded programmers to support them. USFM is specialized for the problems of scriptural markup and allows relatively lightweight solutions to the problems that arise in marking up scripture. That specialization, however, makes it less well suited to the markup of lexica, handbooks, commentaries, critical apparatus, and other resources translators use in their work. Supporting such materials in interactive tools for Bible translators requires finding a way to make software deal well both with USFM for scriptural texts and various flavors of XML for scriptural and other materials. Many of the hardest problems have to do with the complexity of the documents, not with the form of markup used to encode them. There are lessons here for anyone who needs to work with heterogeneous data or live outside of a walled garden.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-05-10:00. XProc & XSLT

Convert legislation from MS Word to XML and handle parallel documents in multiple languages.

Stakeholder(s)

Geert Bormans :
C-Moria BV

Thursday 10:00 - 10:30 EDT (+ Q&A 10:30 - 10:45) ~ XProc 3 and XSLT 3: Adventures of an early adopter ~ Although converting a large corpus of legislation from MS Word to XML may not be a new concept, nor is handling parallel documents in multiple languages, doing so with new tools while maintaining a production environment is nonetheless a considerable challenge. To make certain there would be no surprises at the end, the information model was not developed starting with the easiest or most common cases, but with the 200 most complex cases, which had been showstoppers for previous attempts at conversion. The work, based in XProc 3.0 and XSLT 3.0, has been successful, and the initial effort has paid off.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-05-11:00. Manuscripts

Make manuscripts project findable, accessible, interoperable, and reusable

Stakeholder(s)

Pietro Maria Liuzzo :
Universität Hamburg

Thursday 11:00 - 11:30 EDT (+ Q&A 11:30 - 11:45) ~ Serving IIIF and DTS APIs from TEI data with XQuery with support from a SPARQL Endpoint (LB) ~ When presenting manuscript materials online, an IIIF (International Image Interoperability Framework) interface makes it possible for the user to zoom in and out dynamically on a manuscript page, providing far better access to the page than any set of pre-determined zoom levels, as well as a better user experience. Distributed Text Services (DTS) provide similar flexibility for navigation of text, enabling consistent user interfaces for distributed text collections. Both IIIF and DTS use linked-data APIs with JSON as a data format. We describe how we can use the RESTxq interface in eXist-db to support these APIs and make the manuscripts of the Beta mašāḥəft project findable, accessible, interoperable, and reusable (FAIR).

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-05-12:00. XML & CSS

Print recipes.

Stakeholder(s)

Peter Flynn

Thursday 12:00 - 12:30 EDT (+ Q&A 12:30 - 12:45) ~ Printing recipes: Continuing adventures in XML and CSS for recipe data ~ Most XSLT programmers would likely find the task of transforming recipe markup into HTML and CSS suitable for printing to be a relatively straightforward task. But what if you didn't want to do it that way? Browsers will apply CSS style to XML. Would it be practical to render robust recipe markup directly from XML in the browser with only CSS? Would it even be possible? The answer is "yes." The exercise is an interesting one and the details may very well teach you something new about CSS.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-05-13:00. Balisage Bard

Exercise literary creativity with original poems, short stories, jokes, songs, and other masterpieces.

Stakeholder(s)

Lynne Price :
Gamemaster

Thursday 13:00 - 14:00 EDT ~ Once again, Balisage Bard gives you the opportunity to exercise your literary creativity with original poems, short stories, jokes, songs, and other masterpieces. Subject matter must be related to Balisage (markup, papers presented this or previous years, virtual conferences, and so forth). Read your effort or play it on video during the game session. Translations of works in languages other than English are not required but will be appreciated. There is a two-minute time limit per presentation. Sign up by entering your name in the Bard chat room. Presentation sequence at that gamemaster's discretion. One submission per person/team unless there is time for more at the end. And listen closely. Vote for your favorite three works after the last presentation. Who will be the 2021 Balisage Poet Laureate?

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-05-14:00. XForms

Handle invariants that express structure.

Stakeholder(s)

Steven Pemberton :
CWI, Amsterdam

Thursday 14:00 - 4:30 EDT (+ Q&A 14:30 - 14:45) ~ Structural constraints in XForms ~ XForms! XForms defines relationships called “invariants.” When one member of an invariant relationship changes, the other member(s) are updated automatically (similar to the way spreadsheets work). Currently XForms only allows invariants to be expressed between simple values (values that can be calculated with a simple expression), while structural changes can only be detected using XForms events. Could the invariant mechanisms of XForms be extended to handle invariants that express structure, without resorting to events? Yes, by considering structural invariants to be just a higher-level form of invariant, where the work of the structural invariant is to rebuild networks of lower-level simple invariants. Structural invariants can then be merged into the general XForms invariant recalculation, by treating the rebuild phase of the XForms update mechanism as a higher-level version of the current recalculate.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-05-15:00. STAR

Develop an alternate serialization format in XML.

Stakeholder(s)

Michael R. Gryk :

UCONN Health & University of Illinois Urbana-Champaign

Protein Data Bank

Thursday 15:00 - 15:30 EDT (+ Q&A 15:30 - 15:45) ~ Deconstructing the STAR file format ~ Markup comes in many flavors. The STAR (self-defining text archival and retrieval) file format, first proposed in 1991, predates both XML and JSON, but shares many of their features: it's a machine-independent textual format designed for simplicity of reading and writing and flexibility in the face of change. STAR is used primarily in scientific data exchange; it's the basis for the Crystallographic Information File (CIF) and the default format used by the Protein Data Bank. STAR encompasses both a model of information and a syntax for serializing it. If we can decouple the model from the syntax, we can understand STAR better, and develop an alternate serialization format in XML.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-05-16:00. Web Semantics

Examine the history and immediate future of semantic markup on the Web.

Stakeholder(s)

Simon St.Laurent :
LinkedIn Learning

Thursday 16:00 - 16:30 EDT (+ Q&A 16:30 - 16:45) ~ Semantics and the Web: An Awkward History ~ The vast bulk of the markup that gets sent out into the world over networks keeps getting simpler and simpler, using markup with fewer features than was common in the 1980s or 1990s. Support from non-markup technologies provides the meaning. Are semantics receding from markup? Perhaps. Let's examine our history and immediate future.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-06-10:00. TEI Text

Present TEI text in a way that resembles the printed edition

Stakeholder(s)

Tony Graham :
Antenna House, Inc.

Friday 10:00 - 10:30 EDT (+ Q&A 10:30 - 10:45) ~ Call me PasticheMael: Recreating the Moby-Dick first edition (LB) ~ Descriptive markup is all about capturing structure, not formatting, isn't it? But what if you have a good TEI text of the first edition of a novel and want to present it in a way that resembles the printed first edition? And what if that first edition had all sorts of 19th century typographic quirks? Then you have to get creative with XSLT and FO (and know how to bend your formatter to your will).

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-06-11:00. XML Strings

Process XML strings.

Stakeholder(s)

Michael Kay :
Saxonica

Friday 11:00 - 11:30 EDT (+ Q&A 11:30 - 11:45) ~ ZenoString: A data structure for processing XML strings ~ XML documents (usually) contain lots of strings of Unicode characters, sometimes very, very long strings. Representing and processing these efficiently in Java poses several challenges. Variable-width encodings improve space efficiency at the cost of direct addressability. Operations, both string construction and the implementation of XPath functions, may require strings to be copied. Copying very large strings requires large blocks of contiguous free memory, putting pressure on the garbage collector. Conversely, a large collection of very short strings has more per-object overhead and poorer locality of reference, a critical factor in the performance of modern hardware architectures. Taking inspiration from several sources, a novel data structure called ZenoString supports direct addressing in many common cases, strings of unlimited length, and good locality of reference, while maintaining high performance.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-06-12:00. CSS & XSLT

Embed CSS style rules in XSLT templates.

Stakeholder(s)

Liam Quin :
Delightful Computing

Friday 12:00 - 12:30 EDT (+ Q&A 12:30 - 12:45) ~ CSS Within: An application of the principle of locality of reference to CSS and XSLT ~ “CSS Within” is a method of embedding CSS style rules into XSLT templates, so that the CSS rules that govern look-and-feel of elements in the result are as near as possible to where those elements are generated. The proximity of the rules to the corresponding element generation code provides greater clarity and an increase in programmer efficiency especially when doing maintenance. CSS Within also reduces impediments both to stylesheet refactoring and to CSS changes, because the effects of making changes are more readily apparent. To make all of this work, some extension elements are proposed, and a separate pass on the stylesheet to create CSS files.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-06-13:00. Sponsor Presentation

Discuss topics of interest

Friday 13:00 - 13:45 EDT ~ Sponsor Presentation ~ A Balisage Sponsor will discuss topics of interest.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-06-14:00. SVG Text

Determine the dimensions of text objects in SVG

Stakeholder(s)

David J. Birnbaum :
University of Pittsburgh

Charlie Taylor :
University of Pittsburgh

Friday 14:00 - 14:30 EDT (+ Q&A 14:30 - 14:45) ~ How long is my SVG <text> element? (LB) ~ SVG expects the creator of a graphic to give it the dimensions of the objects it is to create and place. That's fine for a lot of graphical objects, but not if what you want to do involves placing text: SVG doesn't know its dimensions until after the fact. You can get around this problem, using XSLT, if you're willing first to extract font metrics to an XML file. Or you can use JavaScript in the browser to determine the text dimensions as SVG renders them, then use the results in the final placement of objects.

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

8-06-15:00. Complexity

Discuss limitations

Stakeholder(s)

C. M. Sperberg-McQueen :
Black Mesa Technologies

Friday 15:00 - 15:30 EDT ~ Catastrophic complexity ~ Eventually, things reach their limit...

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

Administrative Information

Start Date: 2021-08-02

End Date: 2021-08-06

Publication Date: 2021-07-31

Source: <https://www.balisage.net/2021/Program.html>

Submitter:

Given Name: Owen

Surname: Ambur

Email: Owen.Ambur@verizon.net

Phone:

_e52b43ce-f22f-11eb-994d-49871283ea00