

Balisage 2019 Program

Contents

Mission.....	5
Values	5
7-30-9:15. Explicit Markup.....	6
Forms & Media.....	6
Hierarchy.....	6
Visibility & Knowledge	7
7-30-9:45. TEI Annotation.....	8
7-30-11:00. Workflows & Systems.....	9
7-30-11:45. Scholarly Content	10
Best Practices	10
7-30-14:00. Interviews	11
7-30-14:45. JSON Schema.....	12
Schema Validation	12
7-30-16:00. XProc	13
7-31-9:00. Text & Markup Processing	14
7-31-9:45. Character Data.....	15
7-31-11:00. XML Everywhere.....	16
GUIs.....	16
WebSocket	16
State Chart XML	16
7-31-11:45. XForms Multitasking	18
Control	18
Monitoring	18
Tasks	18
7-31-14:00. Interviews.....	20
7-31-14:45. Document Dysfunction	21
7-31-16:00. Invisible XML.....	22
7-31-16:45. Aparecium.....	23
Non-XML Data	23
8-1-9:00. MVC Paradigm	24
8-1-9:45. EXPath Packaging.....	25
8-1-11:00. Security Content.....	26
8-1-11:45. Accessibility	27
8-1-13:15. Literary Creativity	28
8-1-14:00. Interviews.....	29
8-1-14:45. Vocabularies.....	30
8-1-16:00. Parts of Speech	31
8-1-16:45. Encodings.....	32
8-2-9:00. Security Controls.....	33
Schemas & Tooling	33
Documentation	33
Parallel Schemas	33
Conversion Tools	34
8-2-9:45. Loose-Leaf Publishing	35
8-2-11:00. XML DBs & CMSs	36
Digital Editions	36
8-2-11:45. Rules & Requirements.....	38
Administrative Information.....	38

DEMONSTRATION ONLY



Balisage 2019: The Markup Conference (BS2019)

Stakeholder(s):

Balisage 2019 Sponsors

Role: Sponsorship

antennahouse :

Silver Sponsor

Role: Sponsorship

Oxygen XML Editor :

Sponsor

Role: Sponsorship

Data Conversion Laboratory :

Sponsor

Role: Sponsorship

Text Encoding Initiative

Role: Co-Sponsorship

Washington Area SGML/XML Users Group

Role: Co-Sponsorship

The XML Guild :

Co-Sponsor

Role: Co-Sponsorship

Sister/Related Conferences

Role: Conferencing

XML Prague Markup

Role: Conferencing

UK XML Summer School

Role: Conferencing

Mulberry

Role: Balisage Conference Producer

Balisage People :

People involved with Balisage — The people making Balisage include markup theoreticians and practitioners, data modelers, designers, architects, and both aficionados and deep thinkers. We work as software developers, system architects, academics, integrators, librarians, data miners, lexicographers, archivists, document managers, standards developers, programmers, and publishers.

Role: Balisage Conference Making

Balisage Conference Committee

Role: Balisage Conference Direction

B. Tommie Usdin :

Mulberry Technologies — B. Tommie Usdin is President of Mulberry Technologies, Inc., a consultancy specializing in XML and SGML. Ms. Usdin has been working with SGML since 1985 and has been a supporter of XML since 1996. She chairs the Balisage conference. Ms. Usdin has developed DTDs, Schemas, and XML/SGML application frameworks for applications in government and industry. Projects include reference materials in medicine, science, engineering, and law; semiconductor documentation; historical and archival materials. Distribution formats have included print books, magazines, and journals, and both web- and media-based electronic publications. She is co-chair of the NISO Z39-96, JATS: Journal Article Tag Suite Working Group and a member of the NISO STS Standing Committee. You can read more about her at <http://www.mulberrytech.com/people/usdin/index.html>

Role: Balisage Conference Committee Chair

Deborah A. Lapeyre :

Mulberry Technologies — Deborah Aleyne Lapeyre is a Senior Consultant for Mulberry Technologies, Inc., a consulting firm specializing in helping their clients toward better publishing through XML, XSLT, and Schematron solutions. She works with Tommie Usdin as architects and Secretariat for JATS (ANSI NISO Z39.96-2019 Journal Article Tag Suite) and BITS (Book Interchange Tag Suite). She teaches hands-on XML, XSLT, DTD and schema construction, and Schematron courses as well as numerous technical and business-level introductions to XML and JATS. Debbie has been working with XML and XSLT since their inception and with SGML since 1984 (before SGML was finalized as an ISO standard). In a previous life, she wrote code for systems that put ink on paper and used, taught, and documented a proprietary generic markup system named "SAMANTHA". Hobbies, besides Balisage, include pumpkin carving parties.

Role: Balisage Conference Committee Co-Chair

James David Mason :

James David Mason, originally trained as a mediaevalist and linguist, became a writer, systems developer, and manufacturing engineer at U.S. Department of Energy facilities in Oak Ridge since the late 1970s. In 1981, he joined the ISO's work on standards for document management and interchange. He chaired ISO/IEC JTC1/SC34, which is responsible for SGML, DSSSL, Topic Maps, and related standards, for more than 20 years. Dr. Mason has been a frequent writer and speaker on

— continued next page

Stakeholders (continued)

standards and their applications. For his work on SGML, Dr. Mason has received the Gutenberg Award from Printing Industries of America and the Tekkie Award from GCA. He recently retired from working on information systems to support manufacturing and documentation at DOE's Y-12 National Security Complex (Y-12) in Oak Ridge, Tennessee.

Role: Balisage Conference Committee Co-Chair

C. M. Sperberg-McQueen :
Black Mesa Technologies

Role: Balisage Conference Committee Co-Chair

Norman Walsh :

MarkLogic Corporation — Norm Walsh is a Principal Engineer at MarkLogic Corporation where he helps to develop APIs and tools for advanced content applications. He has also been an active participant in international standards efforts at both the W3C and OASIS. At the W3C, Norm was chair of the XML Processing Model Working Group, co-chair of the XML Core Working Group, and an editor in the XQuery and XSLT Working Groups. He served for several years as an elected member of the Technical Architecture Group. At OASIS, he was chair of the DocBook Technical Committee for many years and is the author of *DocBook: The Definitive Guide*. Norm has spent more than twenty years developing commercial and open source software.

Role: Balisage Conference Committee Co-Chair

Balisage Conference Advisory Board

Role: Advice

Syd Bauman :

Northeastern University — Syd Bauman was introduced to the world of markup languages through the world of computer typesetting. He became interested in SGML just prior to its publication in 1986, but did not start seriously engaging with it until late 1990 when he started working with TEI P1.1. He has since become immersed in markup, XML, and TEI. Syd is a Senior XML Programmer/Analyst at the Women Writers Project, part of Northeastern University's Digital Scholarship Group, where he does not actually do much programming. When he does it is usually XSLT designed to read TEI. But regardless of language, Syd's programs are always copylefted. Syd often teaches XML, TEI (in particular, TEI customization), and XSLT. He served as the North American Editor of the TEI Technical Council since 2013.

Role: Advice

Jeff Beck :

National Library of Medicine

Role: Advice

David J Birnbaum :

University of Pittsburgh — David J. Birnbaum is Professor and Chair of the Department of Slavic Languages and Literatures at the University of Pittsburgh. He has been involved in the study of electronic text technology since the mid-1980s, has delivered presentations at a variety of electronic text technology conferences, and has served on the board of the Association for Computers and the Humanities, the editorial board of *Markup*

languages: theory and practice, and the Text Encoding Initiative Technical Council. Much of his electronic text work intersects with his research in medieval Slavic manuscript studies, but he also often writes about issues in the philosophy of markup.

Role: Advice

Jon Bosak

Role: Advice

Robin Cover :
OASIS

Role: Advice

Steve DeRose :
Independent Consultant

Role: Advice

Bob DuCharme :
CCRi

Role: Advice

Patrick Durusau
Role: Advice

Eric Freese :
TopQuadrant

Role: Advice

Eduardo Gutentag
Role: Advice

G. Ken Holman :
Crane Softwrights

Role: Advice

Sam Hunting
Role: Advice

Michael Kay :
Saxonica

Role: Advice

David A. Lee :
Nexstra, Inc

Role: Advice

Chris Lilley :
World Wide Web Consortium

Role: Advice

Yves Marcoux :
Université de Montréal

Role: Advice

— continued next page

Stakeholders (continued)

Sean McGrath :
Propylon

Role: Advice

Mary McRae :
IQ Solutions

Role: Advice

Steven R. Newcomb :
Coolheads Consulting

Role: Advice

Wendell Piez :
Piez Consulting Services — A long-time contributor to Balisage and its predecessor conferences, since 2018 Wendell Piez has been serving as IT Specialist in the Information Technology Laboratory, National Institute of Standards and Technology (Gaithersburg Maryland). There he is putting his XML and XSLT skills to daily use.

Role: Advice

Ari Nordström :
Karnov Group — Ari Nordström is a Senior XML Specialist (fancy speak for "markup geek") at Karnov Group, a Scandinavian legal publisher. He is based in Göteborg, Sweden, but has been known to provide angled brackets across a number of borders over the years. Ari is the proud owner and head projectionist of Western Sweden's last functioning 35/70mm cinema, situated in his garage, which should explain why he

once wrote a paper on automating commercial cinemas using XML.

Role: Advice

Allen H. Renear :
University of Illinois at Urbana-Champaign

Role: Advice

Bruce Rosenblum :
Inera

Role: Advice

Jeni Tennison :
Jeni Tennison Consulting

Role: Advice

Henry S. Thompson :
University of Edinburgh

Role: Advice

Fabio Vitali :
University of Bologna

Role: Advice

Lauren Wood
Role: Advice

Ann Wrightson
Role: Advice

Mission

To document and share the agenda for the Balisage 2019 conference.

Values

Theory: There is nothing so practical as a good theory

Practicality

7-30-9:15. Explicit Markup

Separate meaning from format.

Stakeholder(s)

B. Tommie Usdin :
Mulberry Technologies

Role: Presenter

Brian Reid
Role: Past Presenter

In 1998, at a Balisage predecessor conference, Brian Reid told us we couldn't have the world we wanted. XML wouldn't deliver. He used twenty-year-old slides, slides that he had originally presented at a conference in 1981 to make his point.

Tuesday 9:15 am - 9:45 am -- Explicit markup: a fool's errand or the next big thing? ... I still want the world that Brian Reid told us we could not have; I still want Brian Reid to have been wrong. I still believe that separating meaning from format will enable our documents to be displayed in many forms and media, that a markup format that makes hierarchy explicit makes complex documents tractable, that when content creators author in systems that make declarative markup visible and use the author's knowledge to add value to their content, we will be able to make documents sing! And I have the twenty-year-old slides to prove it.

Forms & Media

Display documents in many forms and media.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

Hierarchy

Make hierarchy explicit to make complex documents tractable.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

-

Visibility & Knowledge

Make declarative markup visible and use the author's knowledge to add value to content.

Stakeholder(s):

Content Creators

Role: **Content Creation**

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

7-30-9:45. TEI Annotation

Implement TEI standoff annotation in the browser.

Stakeholder(s)

Hugh Cayless :

Duke Collaboratory for Classics Computing (DC3)

Role: Presenter

Tuesday 9:45 am - 10:30 am -- Standoff markup allows you to add information to a text without modifying the source. Often this can be achieved by linking between different documents. Various mechanisms exist for handling the connections involved. But some cases such as named entity recognition appear to require inline markup. Could we do this with standoff markup too? The answer is yes, using the TEI Critical Apparatus model, but it isn't completely straightforward.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

7-30-11:00. Workflows & Systems

Use XML to build rich models, extensive workflows, and robust systems.

Stakeholder(s)

Ari Nordström

Role: **Presenter**

Tuesday 11:00 am - 11:45 am -- Eating your own dog food -- Declarative solutions generally - and XML specifically - invite experimentation, iterative development, and play. In this way they encourage the self-described "non-programmer" to build rich models, extensive workflows, and robust systems. But can you build the whole application this way? And if the application is critical to getting paid, do you have the courage to do so? We Swedes are a courageous lot.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

7-30-11:45. Scholarly Content

Optimize the reusability of scholarly content.

Stakeholder(s)

Jeff Beck :
 NCBI/NLM/NIH, JATS4R Steering Committee

Role: **Presenter**

Tuesday 11:45 am - 12:30 pm -- Rules for the Rulemakers: JATS4R's Self Guidance on Attributes (LB) -- Maximal flexibility of rules, or ease of reuse -- choose one. The tighter the rules, the more consistent documents will be and the easier it will be to reuse them, but only if the rules are reasonable enough to be adopted. (If all the data creators ignore the rules, reuse doesn't get easier.)

Best Practices

Develop best-practice recommendations for tagging content in JATS XML.

The group has devoted particular attention to the flexibility/reuse tradeoff for rules on attribute use and controlled values, and we eventually decided that we needed some rules for ourselves, on how to write rules for attributes in our recommendations. In the process of developing our guidance document for writing rules for attribute values in our recommendations, we learned (or at least articulated) some things along the way.

Stakeholder(s):

JATS4R :
 JATS4R (JATS for Reuse) is a NISO working group devoted to optimizing the reusability of scholarly

content by developing best-practice recommendations for tagging content in JATS XML.

Role: **Best Practice Recommendations**

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

7-30-14:00. Interviews

Get to know about the work and lives of some of the people in our markup community.

Tuesday 2:00 pm - 2:45 pm -- Interviews to be determined -- Balisage is a gathering of remarkable people! We have developed standards, tools, languages, and applications. We have written books, blogs, tweets, and code. We have changed organizational cultures - and in a few cases we have changed the world. In these interviews, we will get to know a bit about the work and lives of some of the remarkable people in our markup community.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

7-30-14:45. JSON Schema

Apply Brzowski derivatives to JSON Schema.

Stakeholder(s)

Mary Holstege :

MarkLogic Corporation — Mary Holstege has been developing software in Silicon Valley for decades, in and around markup technologies and information extraction. She currently works at MarkLogic Corporation, where she mainly works on search. She holds a Ph.D. from Stanford

University in Computer Science, for a thesis on document representation.

Role: Presenter

Tuesday 2:45 pm - 3:30 pm -- In 1964, Janusz Brzowski defined a new technique for computing whether a string of symbols is in the language defined by an extended regular expression. Brzowski derivatives have been used for content model validation in several XML schema processors; they can also be applied to the task of model validation for JSON Schema.

Schema Validation

Shed light on matching problems outside the original problem scope of JSON Schema validation.

As it turns out, applying them to JSON Schema requires several extensions to cover "type-tagged" expressions, which sheds light on certain interesting matching problems outside the original problem scope of JSON Schema validation.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

7-30-16:00. XProc

Enable XML centric workflows.

Stakeholder(s)

Achim Berndzen :
<xml-project />

Role: **Presenter**

Geert Bormans

Role: **Presenter**

Gerrit Imsieke :
le-tex

Role: **Presenter**

Norm Walsh

Role: **Presenter**

Tuesday 4:00 pm - 5:30 pm -- XProc 3.0 (LB) -- XProc is an XML pipeline language designed for XML centric workflows. XProc 3 is currently under development. The editorial team believes that the core language specification is in "last call". XProc 3.0 is designed to improve the usability of XProc. Features include: handling XML, text, binary, and JSON documents, text value and attribute value templates, typed variables and options using XDM 3.1, and a lot of shortcuts. XProc's language design is still about encapsulated data processing steps with defined inputs, outputs, and options. What sets it apart from other scripting languages, Make, and Ant is this: It is a truly functional language with immutable inputs and state. This allows composition of arbitrarily complex steps without risking unexpected side effects and without jeopardizing manageability. In contrast to other functional languages, it offers multiple return "values" (on the named output ports) that don't have to be consumed at once, or at all. Apart from becoming less verbose, XProc 3.0's major strength is that JSON, text, HTML and binary data are now first-class citizens, making it suitable for data processing in the Web age. In addition to describing the new XProc 3.0 we will show code (including both XProc 1.0 and 3.0) and demonstrate XProc tools.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

7-31-9:00. Text & Markup Processing

Outline a plan to make text and markup processing easier.

Stakeholder(s)

Sam Wilmott :

Sam Wilmott designed his first programming language in the winter of 1967-1968 and was using early non-standardized markup languages in the late '60s. Since then he has led the development of typesetting/text-formatting systems for the Canadian Government Printing Office (in the '70s) and for a major real-estate company (in the '80s), implemented one of the first SGML parsers (which was also the first pull-model markup parser), and is the originator of the OmniMark programming language (in the early '90s), with its strong support of SGML, XML, and

text transformation. After leaving OmniMark, Sam worked in the XSLT world: he contributed to the implementation of an XSLT compiler and worked as an XSLT programmer and analyst (in the early 2000s). Currently he is largely retired, happily married, does voluntary work locally and walks a little dog every day, but in spite of his advancing age, he is nonetheless working on new programming language ideas for markup language and text processing.

Role: Presenter

Wednesday 9:00 am - 9:45 am -- Text and markup processing languages, past, present, and future -- Programming language design is in continual flux, with significant new languages coming along every few years. In the field of text and markup programming languages, things seem stable at the moment, with XSLT in a dominant position and a few other languages filling in the gaps. But text and markup processing is no more exempt from change than any other field. What should the next language for this application domain look like? Can we make text and markup processing easier than it is now? What direction should we take? For the last ten years or so, I have been working on this problem. I have a plan.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

7-31-9:45. Character Data

Streamline character data for tokenization.

Stakeholder(s)

Ashley M. Clark :

Ashley M. Clark is XML Applications Developer for the Northeastern University Women Writers Project and the Digital Scholarship Group.

Role: Presenter

Full-Text Search Tool Developers :

Role: Full-Text Search Tool Development

Textual Analysis Tool Developers

Role: Textual Analysis Tool Development

Wednesday 9:45 am - 10:30 am -- "With one voice": streamlining character data for tokenization -- Some full-text search and textual analysis tools operate exclusively on sequences of tokens. Deriving input for these tools from XML documents can be challenging and depends heavily on the encoding practices and assumptions which produced the XML. Does metadata information, for example, carry the same weight as the text? If a document includes annotations about nuances of the transcription, including those annotations may aid researchers attempting to find relevant documents, but may hinder a process that is performing textual analysis of the work authored. Rather than attempting to make all tools powerful enough to deal with these issues, a modular approach to tokenization has been developed.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

7-31-11:00. XML Everywhere

Demonstrate that XML can be used up and down the application stack.

Stakeholder(s)

Zahra Al-Awadai :
Technical University of Munich (TUM)

Role: **Presenter**

Anne Brüggemann-Klein :
Technical University of Munich (TUM)

Role: **Presenter**

Christina Grubmüller :
Technical University of Munich (TUM)

Role: **Presenter**

Philipp Ulrich :
Technical University of Munich (TUM)

Role: **Presenter**

Wednesday 11:00 am - 11:45 am -- Graphical user interfaces in the X stack -- "XML Everywhere" isn't just a slogan: it actually works, up and down the XML application stack. Recent developments, such as the inclusion of custom elements in HTML5, allow the declarative approach of XML to come into the browser/server interaction.

GUIs

Apply XForms as the basis for graphical user interfaces.

XForms, supported by SVG and CSS, can serve as the basis for a graphical user interface.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

WebSocket

Use a custom WebSocket element to support client-to-client and server-push communication.

A custom WebSocket element can support client-to-client and server-push communication of XML data.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

State Chart XML

Apply State Chart XML (SCXML) to extend to models of operations in applications.

Applications of State Chart XML (SCXML) mean that the "XML Everywhere" approach can be extended all the way to models of operations in an application. Interactive games offer living proof of the stack.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

DEMONSTRATION ONLY

7-31-11:45. XForms Multitasking

Multitask in XForms.

Stakeholder(s)

John M. Boyer :

IBM Canada — John Boyer is an IBM Distinguished Engineer and Master Inventor. He joined IBM in 2005 via its acquisition of PureEdge Solutions. While at PureEdge, John undertook graduate studies in theoretical computer science, earning his Ph.D. in 2001 from the University of Victoria. John has over two dozen refereed journal and conference publications, over 30 granted patents, and numerous other

publications of professional articles and industry standards. John has served as Chair of the XForms working group, Editor of XForms, and Editor or Co-author of XML Signatures, XML Canonicalization, and related W3C Recommendations.

Role: Presenter

Wednesday 11:45 am - 12:30 pm Multitasking algorithms in XForms -- XForms aren't just about simple data collection: they can process data as well. Even operations that have traditionally been considered "procedural" rather than "declarative" can be specified -- and executed -- from within an XForms instance. XForms actions provide form authors with the ability to solve basic data manipulation use cases such as changing data values and copying and deleting elements. XForms offers multitasking techniques to control execution, to monitor execution progress, and to set task execution priorities dynamically. There's a lot more there than you might have expected!

Control

Control execution.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

Monitoring

Monitor execution progress.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

Tasks

Set task execution priorities dynamically.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

DEMONSTRATION ONLY

7-31-14:00. Interviews

Get to know about the work and lives of some of the people in our markup community.

Wednesday 2:00 pm - 2:45 pm -- to be determined -- Balisage is a gathering of remarkable people! We have developed standards, tools, languages, and applications. We have written books, blogs, tweets, and code. We have changed organizational cultures -- and in a few cases we have changed the world. In these interviews, we will get to know a bit about the work and lives of some of the remarkable people in our markup community.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

7-31-14:45. Document Dysfunction

Fix document dysfunction.

Stakeholder(s)

Jean Paoli :

Jean Paoli is the Founder of Docugami Inc., a startup that uses AI to transform the unique document business processes of individual companies, making frontline users more efficient while giving COOs better compliance and insights – inspired by his deep belief that openness and interoperability raises all boats. He was formerly President of Microsoft Open Technologies, Inc., and one of the co-creators of the XML 1.0 standard with the World Wide Web Consortium (W3C). Throughout his career, Jean has worked in startups: before Microsoft, with Inria, the renowned French research Labs (Gipsi S.A. and Grif S.A.); and within Microsoft creating four new startups: XML, InfoPath, opening the Office formats and MS OpenTech

(Microsoft's open source subsidiary). The startups he built created breakthrough platform technologies used today by millions. He is the recipient of multiple industry awards for his work on XML, semi-structured data, the convergence of documents and data and openness at large. In addition to core technical design, Jean takes deep care at building healthy ecosystems at worldwide scale. He is credited as one of the key leaders responsible for shifting in a fundamental way, under the guidance of the CEO, Microsoft's strategy to embrace and love open source.

Role: Presenter

Wednesday 2:45 pm - 3:30 pm -- We Created Document Dysfunction. It Is Time to Fix It. (LB) -- Some of us building software need to take a hard look in the mirror. For years, we have promised that technology would solve the world's information management problems, but 85% of business information is still "dark data," with potentially useful insights lost in a rising tide of disconnected documents, emails, Slack conversations, voice-to-text messages, etc. We need an effective approach to documents and want to start a public conversation about these issues. We believe that effective solutions should be based on: Declarative Markup; AI sympathetic to "Small Data"; focus on company-specific documents; applying AI to documents as a whole; and solutions that do not disrupt existing workflows or require massive investment. The future isn't about AI making human beings obsolete; the future is about AI making human beings and companies more productive, effective, and creative.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

7-31-16:00. Invisible XML

Discuss whether it is necessary to see the markup.

Stakeholder(s)

James David Mason

Role: **Presenter**

Wednesday 4:00 pm - 4:45 pm Do we really want to see markup? -- Markup fanatics have long cried, "We need to see the markup!" Yet since the earliest stages of developing the SGML standard, there has been an urge even among standards developers to avoid having to write tags everywhere. The recent urge to create "Invisible XML" is but the latest symptom of a smoldering disease, from which I, too, suffer.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

7-31-16:45. Aparecium

Provide an XQuery/XSLT library to make the use of "invisible XML" convenient.

Stakeholder(s)

C. M. Sperberg-McQueen :
Black Mesa Technologies LLC

Role: **Presenter**

XSLT Users
Role: **XSLT Usage**

XQuery Users
Role: **XQuery Usage**

Wednesday 4:45 pm - 5:30 pm -- Aparecium: an XQuery/XSLT library for invisible XML -- This paper introduces Aparecium, a library intended to make the use of "invisible XML" convenient for users of XSLT and XQuery. Invisible XML, a method for treating non-XML documents as if they were XML, holds great promise for immediately and easily bringing our array of XML technologies to bear on the non-XML data that we encounter (CSS, wiki markup, domain-specific notations, JSON, LaTeX, etc.). Aparecium uses an Earley parser to ensure that any context-free grammar can be used.

Non-XML Data

Bring XML technologies to bear on non-XML data.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

8-1-9:00. MVC Paradigm

Test the Model-View-Controller (MVC) paradigm in XForms.

Stakeholder(s)

John J. Chelsom :

Seven Informatics — John Chelsom is CEO of Seven Informatics Ltd. He trained as an electrical engineer before gaining a PhD in artificial intelligence in medicine. He has been a Visiting Professor in Health Informatics at City University, London and the University of Victoria, Canada. As Managing Director of CSW Group from 1993 to 2008, John was responsible for implementation of XML workflow and production systems for many major organisations, including the British Medical Journal, Jaguar Cars and the Royal Pharmaceutical Society. The Case Notes product developed by CSW was based on XML and other open

standards. In 2003 the UK government chose Case Notes as the primary clinical system in the national architecture for a shared electronic health record covering the 55 million citizens in England. In 2000, John founded the XML Summer School and continues as a board member and lecturer in this annual event. Since 2010 he has been the lead architect of the open source cityEHR product - an XRX (Xforms, REST, XQuery) health records system currently used in a number of hospitals in England.

Role: Presenter

Thursday 9:00 am - 9:45 am -- XForms Space Invaders -- The Model-View-Controller (MVC) paradigm is a design pattern for creating applications in which: the View (web page) interacts with the user; the Model controls manipulation of the data; and the Controller orchestrates the work of the view and the model. Implementing the classic arcade game Space Invaders in an XForms workbench proved to be a successful testbed for this approach. Key functionalities required for Space Invaders are an application "heartbeat" to control the speed/progression of the invaders; animated graphics for the invaders, the Mystery Ship, and laser fire; and the user-controlled laser cannon. The workbench was implemented using Orbeon Forms, an open source framework which supports XForms 1.1 with a number of custom extensions, including Javascript actions, Attribute Value Templates on XHTML elements, and listeners for "keypress" events. Most of the extensions required are included in the draft XForms 2.0 specification (albeit with slightly modified syntax).

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

8-1-9:45. EXPath Packaging

Manage code modules for languages based on XPath.

Stakeholder(s)

Adam Retter :
Evolved Binary

Role: **Presenter**

Joe Wickentowski :

U.S. Department of State, Office of the Historian — Joseph Wickentowski is a historian who specializes in the use of open standards to improve the accessibility and utility of scholarly editions. Since completing his Ph.D. in History at Harvard University in 2007, he has spearheaded a project to convert a major diplomatic documentary edition to TEI, leveraging the XML family of technologies to enable editors,

researchers, and the public to access texts online in multiple open formats. Wickentowski has led workshops on the XQuery language and the eXist-db open source native XML database at TEI@Oxford Summer School in 2010-11 and Digital Humanities 2017, serves as a community liaison for the eXist-db community, and is co-author of a forthcoming book on XQuery for digital humanists in the Coding for Humanists series from Texas A&M University Press.

Role: **Presenter**

Thursday 9:45 am - 10:30 am -- Improving upon the EXPath Packaging System (LB) -- How can we best manage code modules for languages that are based on XPath? Users have had experience with the EXPath packaging system for nearly a decade and can now see both its strengths and weaknesses, particularly the handling of dependencies. The authors propose a replacement packaging system, based on experience with Maven, which leverages mature versioning and dependency management technologies.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

8-1-11:00. Security Content

Package XML security content into bundles for easy deployment.

Stakeholder(s)

Joshua Lubell :

National Institute of Standards and Technology — Joshua Lubell is a computer scientist in the NIST Engineering Laboratory's Systems Integration Division. His interests include model-based engineering, cybersecurity, cyber-physical systems, long-term preservation of digital data, information modeling, and XML and other markup technologies. He received the United States Department of

Commerce Silver Medal for his leadership in developing ISO 10303-203, a standard for representation and exchange of computer-aided designs. He is also a Balisage hyper-local, residing in the heart of Rockville, Maryland.

Role: Presenter

Thursday 11:00 am - 11:45 am -- SCAP composer: a DITA Open Toolkit plug-in for packaging security content -- The Security Content Automation Protocol (SCAP) schema for source data stream collections standardizes the requirements for packaging XML security content into bundles for easy deployment. SCAP bundles must be self-contained (each bundle contains all necessary information without external references) and reversible (XML components must be unmodified so they can be rebundled). These requirements (along with very long identifiers) make authoring the content and bundling very difficult. SCAP Composer is an authoring product which uses a DITA specialized element type for source data stream collections that makes the authoring process easier. SCAP Composer takes an incremental approach to aiding SCAP content authors: it helps only with creating source data stream collections; it does not offer any help with creating the XML resources encapsulated in a data stream collection. SCAP Composer is implemented using the DITA Open Toolkit and can be used with any DITA authoring software that includes the Toolkit, or with a standalone Toolkit.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

8-1-11:45. Accessibility

Discuss accessibility and XML.

Stakeholder(s)

Chandi Perera :

Chandi is the CEO of Typefi Systems, who specialise in automating content production. He joined Typefi in 2006, and has over two decades of publishing and media technology experience. He has acted as a technology consultant to corporations and government agencies around the world, and is a frequent conference speaker in the areas of Accessi-

bility, content management, publishing, media, structured content and digital rights management. Chandi is a board member of a number of industry bodies and has degrees in Engineering and Computer Science.

Role: Presenter

Thursday 11:45 am - 12:30 pm -- Accessibility and XML (LB)

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

8-1-13:15. Literary Creativity

Exercise your literary creativity with poems, short stories, jokes, and songs.

Stakeholder(s)

Lynne Price

Role: **Gamemaster**

Thursday 1:15 pm - 2:00 pm (during lunch - location: Sinequa) Balisage Bard -- Once again, Balisage Bard gives you the opportunity to exercise your literary creativity with poems, short stories, jokes, and songs. Subject matter must be related to Balisage (markup, venue, papers, and so forth). Read your effort 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 for each presentation. As many submissions as time permits will be taken; authors will be called in the order they sign up (there will be a sign-up sheet at conference registration). If time permits, additional volunteers will be accepted during the game.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

8-1-14:00. Interviews

Get to know about the work and lives of some of the people in our markup community.

Thursday 2:00 pm - 2:45 pm -- to be determined -- Balisage is a gathering of remarkable people! We have developed standards, tools, languages, and applications. We have written books, blogs, tweets, and code. We have changed organizational cultures -- and in a few cases we have changed the world. In these interviews, we will get to know a bit about the work and lives of some of the remarkable people in our markup community.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

8-1-14:45. Vocabularies

Describe changes to vocabularies, both successful and unsuccessful.

Stakeholder(s)

Liam Quin :

Delightful Computing — Liam Quin runs an information design company, Delightful Computing, and previously was XML Activity Lead at the World Wide Web Consortium; before that he was involved in the creation of XML itself and in SGML, most notably at SoftQuad Inc. in Toronto. His

backgrounds are in digital typography and computer science.

Role: Presenter

Thursday 2:45 pm - 3:30 pm -- Extending vocabularies: the rack and the weeds -- Markup languages such as XML, JSON, and SGML divide documents into two parts: markup and content. While in theory markup could be created ad hoc for every document, this would mean that markup had no meaning (and thus no value) to anyone but the creator of the document. In order to realize the value of marked up documents for interchange and longevity, we create, write documentation for, and share markup vocabularies. Vocabularies are created in specific contexts and for specific purposes. Like all human constructs, they are flawed and need to be repaired and changed over time. As people bump up against the limitations of their markup vocabularies, they often want to extend those vocabularies. Understanding these processes requires sensitivity of the human needs involved and the social contexts in which people interact with and around the vocabularies. This paper characterizes some of these contexts and their properties, and in the light of this characterization describes changes to vocabularies both successful and unsuccessful.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

8-1-16:00. Parts of Speech

Select the most likely POS and meaning for a given word token.

Stakeholder(s)

Bethan Tovey :
Prifysgol Abertawe (Swansea University)

Role: **Presenter**

Thursday 4:00 pm - 4:45 pm -- You're not the POS of me: the centrality of markup for part-of-speech tagging (LB) -- Part-of-speech tagging is a markup problem: it takes a text and returns it as a series of tokens, each one marked up with information about its grammatical function. This is a standard early-stage process for most kinds of natural language processing, including speech recognition and machine translation. POS tagging for monolingual texts is challenging, but bilingual colloquial texts are even harder. They contain words from both languages, some governed by the grammar of one language, some by the other, and (often) some by both. POS-tagging processes which are effective for either language are unlikely to be sufficient, even in combination, for dealing with bilingual text. In the case of Welsh-English bilingual data, there is the additional challenge of "Wenglish". Wenglish texts combine features of both languages, for example by using Welsh orthography for an English word or English morphological suffixes on a Welsh root. In building DERWen -- a POS Tagger for Wenglish texts -- I have created a mixed Constraint Grammar which takes advantage of the marked-up data produced by the programme in order to select the most likely POS and meaning for a given word token. As the function of markup in the pipeline from raw text to POS-tagged output show, markup is central to computational linguistics.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

8-1-16:45. Encodings

Explore the cascade of "encodings".

Stakeholder(s)

Allen H. Renear :

University of Illinois at Urbana-Champaign — Allen Renear is the dean of the School of Information Sciences at the University of Illinois at Urbana-Champaign. His research interests include information organization and access, particularly the development of formal ontologies for cultural

and scientific objects and the application of those ontologies in information system design, scholarly publishing, and data curation in the sciences and humanities.

Role: Presenter

Thursday 4:45 pm - 5:30 pm -- Encoding -- In their model of digital objects, David Dubin and others postulate three entity types (propositions, symbols, and documents) with three relationships: "expresses", "encodes", and "inscribes". We can "express" an assertion with a sentence. We can also "inscribe" symbols in physical media. I'd like to investigate the cascade of "encodings" that we find in every digital computing system, and the articulation of those encodings that is bound up in everything we do. Encoding can be recursive, but do we really understand it? What is happening when we encode a sentence as a character string? A character as an integer? An integer as an octet? Is encoding a well-understood linguistic or mathematical relationship? Is encoding just a mapping (function)? Is it the same as the relationship between a name and its referent? Is it the same as the relationship between a sentence and the proposition it expresses? I don't think so. So let's explore some possibilities.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

8-2-9:00. Security Controls

Develop technical standards for documentation related to systems security.

Stakeholder(s)

Wendell Piez :

National Institute of Standards and Technologies / Information Technology Laboratory

Role: Presenter

Friday 9:00 am - 9:45 am -- The Open Security Controls Assessment Language (OSCAL): schema and metaschema -- The Information Technology Lab at NIST is developing technical standards for documentation related to systems security. The Open Security Controls Assessment Language (OSCAL) defines lightweight schemas, along with related infrastructure, for tagging system security information to support routine tasks like cross-checking, validating against arbitrary constraints, and producing punchlists. OSCAL is not conceived as "another big XML application" but as a metaschema. This approach allows us to simplify the design and maintenance of schemas and related tooling; support generation of documentation; produce multiple parallel schemas for XML, JSON, and YAML; and construct conversion tools more easily. Documents and tools leverage basic HTML, or even Markdown, for simplicity even though it limits the complexity of what can be directly imported. Conversion is simplified by the metaschema approach, even when multiple schemas apply to a single data collection. We hope that these simplifications will lead not only to more documents but also to more useful documents.

Schemas & Tooling

Simplify the design and maintenance of schemas and related tooling.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

Documentation

Support generation of documentation.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

Parallel Schemas

Produce multiple parallel schemas for XML, JSON, and YAML.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

Conversion Tools

Construct conversion tools more easily.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

DEMONSTRATION ONLY

8-2-9:45. Loose-Leaf Publishing

Typeset and print only pages that have changed.

Stakeholder(s)

Eliot Kimber :

Contrext, LLC — Eliot Kimber has been working with and thinking about hyperdocuments for a very long time, first in his role as a technical writer and writing tool implementor at IBM and then as an SGML and XML consultant. Eliot was a co-editor, with Dr. Charles Goldfarb and Steven Newcomb, of HyTime 2nd Edition (ISO/IEC 10744-1996), a

founding member of the W3C XML Working Group, and a founding and current member of the OASIS DITA Technical Committee. Eliot lives and works in Austin, Texas.

Role: Presenter

Friday 9:45 am - 10:30 am -- Loose-leaf publishing using Antenna House and CSS -- Loose-leaf publishing is the ability to typeset and print only the pages in a document that have changed since its last publication. This presents many interesting challenges. We developed a loose-leaf publication system using Antenna House Formatter, CSS for pagination, and XSLT for post processing the area tree into "change packages" which include only the changed pages. Both the CSS markup and the publication workflow warrant a closer look.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

8-2-11:00. XML DBs & CMSs

Integrate XML databases and content management systems.

Stakeholder(s)

David J. Birnbaum :
University of Pittsburgh

Role: **Presenter**

Hugh Cayless :

Duke University — Hugh Cayless is Senior Digital Humanities Developer at Duke University, where he provides architecture, design, and programming support for the Duke Collaboratory for Classics Computing (DC3). He has served as an elected member of the TEI Technical Council since 2012 (as its Chair from 2015-2017), and he is a founding member of the EpiDoc Collaborative. Hugh earned a PhD in Classics and an MSIS, both from UNC Chapel Hill. His research interests focus on digital critical editions and Linked Open Data.

Role: **Presenter**

Leif-Jöran Olsson :

University of Gothenburg (Sweden) — Leif-Jöran Olsson has been employed since 2005 as a systems developer at Språkbanken, the Swedish Language bank, University of Gothenburg, where he develops research infrastructure for language technology, both nationally and within CLARIN ERIC. His project management experience involves both long-term partner projects (e.g., the Swedish Literary Bank, the Selma Lagerlöf Archive, the Swedish Drama web) and short-term domain-specific toolboxes (including training and use case analysis). He has extensive experience with teaching in language technology and programming. Leif-Jöran obtained his MA in Language Technology from Uppsala University in 2004, and he is one of the core developers of the open-source eXist-db native XML database.

Role: **Presenter**

Joseph Wicentowski :

Office of the Historian, US Department of State — Joseph Wicentowski is a historian who specializes in the use of open standards to improve the accessibility and utility of scholarly editions. Since completing his Ph.D. in History at Harvard University in 2007, he has spearheaded a project to convert a major diplomatic documentary edition to TEI, leveraging the XML family of technologies to enable editors, researchers, and the public to access texts online in multiple open formats. Wicentowski has led workshops on the XQuery language and the eXist-db open source native XML database at TEI@Oxford Summer School in 2010-11 and Digital Humanities 2017, serves as a community liaison for the eXist-db community, and is co-author of a forthcoming book on XQuery for digital humanists in the Coding for Humanists series from Texas A&M University Press.

Role: **Presenter**

Emmanuelle Morlock :

French National Center for Scientific Research (CNRS); History and Sources of the Ancient Worlds (HiSoMA) Research Center, Lyon (France) — Emmanuelle Morlock has been working as an engineer in Digital Humanities at the French National Center for Scientific Research (CNRS) since 2008. Educated in French literature and Library and Information Science, she has specialized in the application of the encoding standard TEI EpiDoc, a version of the guidelines of the Text Encoding Initiative (TEI) that is dedicated to ancient documents. Emmanuelle manages several digital scholarly editions projects within the History and Sources of the Ancient Worlds (HiSoMA) Research Center, and she is also involved in the French-speaking digital humanities association Humanistica as a member of the steering committee and co-director of the digital open access journal *Humanités numériques*.

Role: **Presenter**

Friday 11:00 am - 11:45 am -- Reese's Peanut Butter Cups and eXist-db: integration of XML databases and content management systems in digital editions -- We have identified four models for integrating digital edition content into eXist-db: TEI Publisher; the eXist-db app framework using HTML templating; the eXist-db app framework without HTML templating; and Apache and PHP mediating between the user and eXist-db, so that eXist-db provides only XML database services.

Digital Editions

Examine and compare four ways of conceptualizing and implementing the infrastructure for a digital edition.

We examine and compare these ways of conceptualizing and implementing the infrastructure for a digital edition. Each of them has advantages and disadvantages, primarily from the perspective of sustainability. Our considerations apply to edition frameworks generally and are therefore not specific to eXist-db.

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

DEMONSTRATION ONLY

8-2-11:45. Rules & Requirements

Address the need for rules, schemas, and conformance.

Stakeholder(s)

C. M. Sperberg-McQueen :
Black Mesa Technologies LLC — C. M. Sperberg-McQueen is the founder and principal of Black Mesa Technologies, a consultancy specializing in helping memory institutions improve the long term preservation of and access to the information for which they are responsible. He served as

editor in chief of the TEI Guidelines from 1988 to 2000, and has also served as co-editor of the World Wide Web Consortium's XML 1.0 and XML Schema 1.1 specifications.

Role: Presenter

Friday 11:45 am - 12:30 pm -- Thinking, wishing, saying -- Can we have rules for our documents we cannot write down in a schema language? If a conformance requirement is not mechanically checkable, is it a conformance requirement? If a rule is not testable, is it a rule?

Performance Indicators

Description	Type	Start Date	End Date
[To be described]	Target		
	Actual		

Administrative Information

Start Date: 2019-07-30

End Date: 2019-08-02

Publication Date: 2019-06-18

Source: <https://www.balisage.net/2019/Program.html>

Submitter:

Given Name: Owen

Surname: Ambur

Email: Owen.Ambur@verizon.net

Phone: