NATIONAL ENDOWMENT FOR THE
Humanities

OFFICE OF DIGITAL HUMANITIES

Narrative Section of a Successful Application

The attached document contains the grant narrative and selected portions of a previously funded grant application. It is not intended to serve as a model, but to give you a sense of how a successful application may be crafted. Every successful application is different, and each applicant is urged to prepare a proposal that reflects its unique project and aspirations. Prospective applicants should consult the Office of Digital Humanities application guidelines at http://www.neh.gov/grants/odh/institutes-advanced-topics-in-the-digital-humanities for instructions. Applicants are also strongly encouraged to consult with the NEH Office of Digital Humanities staff well before a grant deadline.

Note: The attachment only contains the grant narrative and selected portions, not the entire funded application. In addition, certain portions may have been redacted to protect the privacy interests of an individual and/or to protect confidential commercial and financial information and/or to protect copyrighted materials.

Project Title:	XQuery Summer Institute: Advancing XML-Based Scholarship from Representation to Discovery
Institution:	Vanderbilt University
Project Directors:	Clifford Anderson
Grant Program:	Institutes for Advanced Topics in the Digital Humanities

Summary

A great deal of work in the digital humanities depends on the effective use of markup languages. Scholars in the humanities have readily mastered the techniques required to annotate texts in the eXtensible Markup Language (XML). However, the full scholarly potential of machine-readable XML documents eludes many digital humanists. They are able to encode in XML, but lack the skills required to query their texts computationally. This leaves many feeling "stuck"—with texts all dressed up in markup but nowhere to go.

The XQuery Summer Institute at Vanderbilt University will be aimed at archivists, librarians, professors, and students who have experience marking up texts in XML, but do not yet know how to work computationally with those documents. Our institute aspires to recruit twelve members of the digital humanities community and help them to get "unstuck" and working productively with their XML-encoded texts.

The interdisciplinary team of leaders number among the best qualified in their fields. Clifford Anderson and two other members of the Vanderbilt University community will conduct the instruction in consultation with Jonathan Robie, Lead Editor of the XQuery and XPath Recommendations at the WC3, and other expert consultants.

The leaders will offer the training in XQuery required for intermediate digital humanists to move to an advanced level of XML expertise. The institute will teach digital humanists how to ask questions across a corpus of digital texts using XQuery. XQuery provides an easy way for humanists to pick up programming because it does not require a lot of "boiler plate" code up front. Powerful queries can be written in one or two lines of code. Using an open source XML database called eXist, participants will learn how to ingest their digital texts into a database and write simple XQuery programs to analyze them. The institute will purposely keep the programming toolset simple while maximizing its expressive power. Participants will go beyond using XML for representation to querying XML for discovery.

The institute will take place on the campus of Vanderbilt University from June 9 to June 20, 2014. Vanderbilt University is a top private research university located in Nashville, Tennessee. The institute will be hosted by the University Library, which has a long track record of research and development of emerging technologies.

Participants will leave with sufficient proficiency in XQuery to work computationally with their XML documents, querying them for specific items of information, searching them for new discoveries, combining them with other data sources, and rendering them publicly through interactive websites. We also anticipate indirect effects from the institute, including publicity for XQuery and the development of a more cohesive community of XQuery users in the digital humanities.

Table of Contents

XQuery Summer Institute:
Advancing XML-Based Scholarship from Representation to Discovery

Significance

Introduction
One of the milestones in digital humanities scholarship has been the adoption of eXtensible Markup Language (XML) as a basic standard for representing information in digital form. Courtesy of a variety of gentle introductory tools, a beginning scholar can quickly learn to encode text following the guidelines of the Text Encoding Initiative (TEI) or the metadata schemas of the library community (EAD, MODS, MADS, METS, etc.).

The story of XML literacy in the digital humanities is a success story, but only a partial one. The full scholarly potential of machine-readable XML documents eludes many digital humanists. They are able to encode in XML but lack tools to program the machine reading of their own documents. A significant number of scholars in the humanities have learned the rudiments of TEI, for example, but very few have experience with XQuery or XML databases. This leaves many feeling "stuck"—with texts all dressed up in markup but nowhere to go.

The XQuery Summer Institute at Vanderbilt University will be aimed at archivists, librarians, professors, and students who have some experience marking up texts in XML, but do not yet know how to work computationally with those documents. Our institute aspires to recruit twelve members of the digital humanities community and help them to get unstuck and working productively with their XML-encoded texts.

The leaders of the institute will offer the training needed for intermediate digital humanists to move to an advanced level of XML expertise. Participants will go beyond using XML for representation to querying XML for discovery.

Why XQuery?
Learning how to mark up digital texts and learning how to program are significantly different intellectual undertakings. Scholars may learn how to encode scholarly texts using the Text Encoding Initiative (TEI) Guidelines but further skills are needed to investigate them computationally, specifically training in XQuery (http://www.w3.org/TR/xquery/).

XQuery is a simple yet powerful language developed by the World Wide Web Consortium (http://www.w3.org/) for querying and extracting information from multiple XML documents. XQuery has become the principal query language for a new breed of XML databases. With XQuery, scholars can learn a single language to ingest their texts into an XML database, ask questions of them, connect them with other sources of information, and publish them on the web. In short, XQuery is the discovery tool for XML.

XQuery is already in use in digital humanities scholarship including:

- The Theological Commons (http://commons.ptsem.edu), the largest open access library in the field of religious studies and theology, is built entirely in XQuery.
- The Rheinische Friedrich-Wilhelms-Universität Bonn and the Cologne Center for eHumanities have jointly developed a digital edition in XQuery of the Egyptian Book of the Dead (http://totenbuch.awk.nrw.de//).
- Rotunda, a digital imprint of the University of Virginia Press, has produced commercial editions using XQuery, including Emily Dickinson's Correspondences (http://rotunda.upress.virginia.edu/edc/) and the Dolley Madison Digital Edition (http://rotunda.upress.virginia.edu/dmde/).

These projects illustrate the flexibility and scalability of XQuery as a programming language for the digital humanities.

Training in XQuery alleviates two challenges scholars frequently encounter when seeking to perform computational analyses of XML datasets.

- Many scholars rely on the programming language Extensible Stylesheet Language Transformations (XSLT) for analysis of marked up texts. While XSLT excels at many uses, it is not optimized for querying across large XML datasets. XQuery provides a robust discovery-oriented alternative to XSLT.
- Computational analyses of digital texts require an understanding of ancillary XML technologies beyond a particular schema. Scholarly expertise in the TEI or EAD is often domain specific and does not necessarily translate into general expertise in XML. XQuery is a programming language applicable to all forms of XML.

Moreover, training in XQuery offers many scholarly benefits.

- XQuery provides an easy way for humanists to pick up programming. XQuery does not require learning a lot of "boiler plate" code up front. Powerful queries can be written in one or two lines of code.
- Understanding XQuery enhances encoding expertise. Scholars who understand how their data will be analyzed can make better informed decisions about how to encode.
- XQuery enables scholars to create and query their own XML corpora directly rather than relying on indexing or analysis performed by third parties.

XQuery Summer Institute at Vanderbilt University

To further scholarly fluency in XQuery, we propose to host a two week summer institute on XQuery for twelve participants at Vanderbilt University from June 9 to 20, 2014. The institute will teach digital humanists how to ask questions across sets of digital texts using XQuery. Using the open source XML database eXist (http://exist-db.org/), participants will learn how to ingest digital texts into a database and write simple XQuery programs to analyze them. As its sample data set, the institute will use thirteen Shakespeare plays in TEI recently released by the Folger Library (http://www.folgerdigitaltexts.org/) under a Creative Commons (CC-BY-NC 3.0) license. By using a common corpus of familiar texts, we will have a

shared basis for exploration and discovery. We will also invite participants to bring examples of their own digital texts for use during afternoon workshops.

Over the course of two weeks, participants will also learn more advanced techniques, such as how to connect marked up documents of various kinds with other sources of data to create new scholarly "mash-ups." The institute will purposely keep the programming toolset simple while maximizing its expressive power. Among other skills, participants will learn:

- To query digital texts for highly specific and contextual results by taking full advantage of the XML markup;
- To encode entities such as people, places, and events automatically in their digital texts using an Application Programming Interface (API) of a third party service;
- To develop an interactive interface to their digital texts with full-text search capability;
- To deploy an eXist database on a cloud service provider for public access.

At the conclusion of the institute, participants will have learned how to store, query, collate, and publish scholarly texts using a single programming language: XQuery.

The proposed institute will fill an important, but neglected, niche in the ecosystem of the digital humanities. Currently, there are excellent options for learning scholarly markup languages such as the NEH-funded Institutes on Scholarly Text Encoding offered by the Brown University Women Writers Project. However, options for learning how to query scholarly texts are scarcer. In part, this paucity of options is due to the fact that XQuery is a relatively new recommendation from the W3C. In addition, few digital humanists had developed the requisite skill in digital markup to make it feasible to offer advanced courses in XQuery explicitly for them. However, a populous and growing community of scholars in the humanities now understands the basics of scholarly markup and is ready to take that knowledge to the next level. An institute on XQuery for digital humanists would fulfill the spirit of the NEH's Institutes for Advanced Topics in the Digital Humanities by providing an opportunity for these scholars to expand their ability to work computationally with digital texts. The institute would also foster the formation of a community of XQuery enthusiasts in the digital humanities by bringing together scholars and practitioners who ordinarily attend non-overlapping conferences.

Institutional Profile

Vanderbilt University, a top private research university in Nashville, Tennessee, has an extensive track record with digital projects and digital scholarship. The Jean and Alexander Heard Library at Vanderbilt University was among the first to recognize the historical value of public news, preserving the broadcast news from the major networks since August 5, 1968 and making them available digitally to contemporary historians and cultural critics. (The NEH has generously awarded Vanderbilt University a number of significant grants during the past two decades to preserve,

catalog and digitize the Television News Archive.) Today, the Heard Library continues to stay at the forefront of the digital revolution in scholarship. In October 2012, Vanderbilt University and the Council on Library and Information Resources (CLIR) established the Committee on Coherence at Scale for Higher Education to foster coordination and collaboration among institutions of higher education in the field of digital scholarship (see http://news.vanderbilt.edu/2012/10/council-library-vanderbilt-scale-higher-education/).

The strength of Vanderbilt's digital leadership team and its award-winning facilities make it an ideal location for a two week institute on advanced topics in the digital humanities. The institute will be supported by the Library Digital Services group in the Vanderbilt University Library. The Library recently undertook a $6 million renovation of its Central Library. The renovation, which received "gold" LEED certification, inaugurated a state-of-the-art digital exhibitions infrastructure as well as several digital classroom environments. The institute will be held in the 4th Floor Electronic Classroom in the Central Library, which has ample space for participants and their laptops as well as two digital projectors and screens. Laptops will be provided for any participant who does not bring a personal laptop to the institute.

The staff of Library Digital Services has experience in all aspects of digital project management, including expertise in XML, XQuery, and XML databases. We have an established track record of teaching non-programmers how to program and are sensitive to the challenges scholars in the humanities face when exploring new technologies. The leaders are also exponents of the distinctive vantage point of the digital humanities, seeing technology not as an end-in-itself but as a means for humanistic exploration and discovery. (See "Staff and Consultants" for details about the qualifications of the leaders of the proposed institute.)

Participants will have the option of lodging in the dorms at Vanderbilt University. The daily cost of a single room (with linens) is $43.60. A partial meal plan (breakfast and dinner) is available for $17.08 per day. Lunches will be included as part of the institute.

The Vanderbilt University campus is readily accessible by public transportation. A bus connects the airport with downtown Nashville and Vanderbilt University. The Nashville International Airport is served by major airlines, including Delta, US Airways, and Southwest. And, of course, Nashville itself offers unparalleled opportunities for dining and recreation during the evenings and weekends.

Curriculum and Work Plan

The goal of the two week institute will be for every participant to leave with a sufficient understanding of XQuery and XML databases for them to continue working productively with those tools at their home institutions. While designed for scholars who have already gained some experience with XML-based markup languages, the institute will not take any background knowledge for granted.

Broadly speaking, the first week will cover the basics of XML, TEI, XHTML and related standards before proceeding to an introduction to XPath and XQuery. The second week will introduce participants to the practical uses of XQuery in XML databases, leading up to a workshop on how to develop full-blown digital projects using nothing other than XQuery and the eXist database. The morning sessions of the institute will consist of formal instruction while the afternoons will feature workshops and practical exercises.

The first day will review XML essentials. During the morning session, participants will review the fundamentals of XML. Particular attention will be given to features which trip up intermediate users such as namespaces, character entities, and the use of CDATA sections in XML. In the afternoon workshop, participants will practice encoding a number of plain text documents in XML.

The second day will review the basics of the TEI P5 specification using the Folger Library's digital texts of Shakespeare as reference. In the afternoon, participants will mark up texts in TEI and discuss how they made encoding decisions.

The third day will introduce the XML Path Language (XPath). In the morning session, participants will use XPath expressions to navigate XML document hierarchies. The full set of XPath axes will be covered along with their abbreviated syntax. The morning session will also cover how to use node tests and predicates. During the afternoon, participants will work through practical exercises using XPath to navigate XML documents.

The fundamentals of the XQuery language will be introduced on the fourth day of the institute. After introducing the concept of functional programming languages, participants will learn how to write FLWOR expressions, which form the building blocks of XQuery. During the afternoon, participants will write simple XQuery expressions to query the Folger Library's editions of Shakespeare in TEI.

The Friday of the first week will conclude with an introduction to functions and operators in XPath and XQuery. The morning session will explore the standard library of functions, highlighting the functions that work on textual data. During the afternoon, the class will write simple XQuery programs that combine FLWOR expressions with a range of built-in functions to render the Shakespeare TEI documents as XHTML.

The goal of the second week is to teach participants practical uses of XQuery for the digital humanities. We will reserve time during the second week for participants to discuss their projects and solicit feedback from fellow participants and institute consultants. Where feasible, the afternoon workshops will deal with challenging problems raised by participants.

On Monday, participants will conduct full-text searches with XQuery in the eXist database. The morning session will introduce eXist and its extension functions.

During the afternoon session, participants will use FLWOR expressions to conduct full-text searches of their documents and filter the result set.

The concept of user defined functions will be explained the following day. Participants will write custom functions and organize those functions into modules in order to build large systems and to share code with others. The morning will also treat the type system and error handling. The class will rewrite the expressions from the previous day as custom functions during the afternoon session.

On Wednesday, the class will learn how to index their documents efficiently in eXist and how to connect them with other sources of information on the web. The morning session will review the fundamentals of indexing in eXist and how to "tune" indexes to different kinds of XML documents. The afternoon session will teach participants how to create mash-ups with their documents by connecting to online application programming interfaces (APIs). As a practical exercise, the class will connect with an online entity extraction API to identify people, places, and events in their XML documents. Participants will reflect on the benefits and drawbacks of using entity extraction services for digital humanities projects.

On Thursday, participants will learn to build a web interface for the Folger Library's edition of Shakespeare in TEI. The morning session will draw together what the participants have learned to this point to create a rich application interface that accepts search terms and displays formatted results for users. During the afternoon session, participants will have the opportunity to add features to their web interface and to customize it according to their interests.

The final day of class will cover the deployment of eXist databases to the "cloud." The morning session will show how eXist databases may be moved from participants' desktops to cloud-based servers. During the afternoon, participants will discuss the advantages of using XQuery and the eXist database to deploy TEI-based projects publicly and we will handle any final questions about deployment.

There will be three assigned texts for the institute: Joe Fawcett, Liam R. E. Quin, and Danny Ayers, *Beginning XML,* 5th edition (Indianapolis: Wiley, 2012); Ron Hitchens, *Getting Started With XQuery: Query XML Like You Mean It* (Dallas: The Pragmatic Bookshelf, 2008); and Priscilla Walmsley, *XQuery* (Sebastopol: O'Reilly Media, 2007). Students will be expected to have familiarized themselves with the basics of XML markup. Exposure to TEI will be helpful, but not presumed. The fundamentals of XML and TEI will be reviewed during the first week of the institute.

The institute will inculcate good programming habits in our participants through practical exercises. In particular, we will practice pair programming (see http://en.wikipedia.org/wiki/Pair_programming) during afternoon workshops. Students will work through problems in pairs, alternating turns at the keyboard. This practice, which arises from the extreme programming community, lessens

anxiety about learning programming and alleviates frustration when participants inevitably receive cryptic error messages instead of the expected results.

Selection of Participants

Our call for applications will be issued in October 2013. We will publicize the announcement through the Vanderbilt University website and also through our partners' websites and distribution lists, including the Association of Southeastern Research Libraries (ASERL), the Council on Library and Information Resources (CLIR), and the Southeastern Universities Research Association (SURA). We will also post announcements through the TEI-L listserv, HASTAC, and Digital Humanities Now, among other venues.

Candidates will submit applications by February 2014 using a simple web-based form. The selection of candidates will take place according to the following criteria.

- The selection of candidates will privilege those who have some prior experience with XML-based markup languages. For practical purposes, this will likely mean some experience with DocBook, EAD, METS, or TEI, though any XML-based markup language will suffice.
- The selection of candidates will favor applications where two or more candidates have applied from the same institution. In our experience, learning to program works best as a social activity. We hope that candidates will continue to practice pair programming after returning from the institute. Pair programming can be a good means to pass along their newly-developed skills to others at their home institutions.
- The selection of candidates will seek a balanced distribution of scholars from different areas of the digital humanities, including archivists, librarians, and professors. We will make an effort to include at least two graduate students and at least two #alt-ac professionals at the institute.

The selection committee will be composed of Clifford Anderson, David Michelson, and Dale Poulter. Candidates will be notified in March 2014 about their acceptance to the institute.

Impact and Evaluation

The key to learning any programming language is practicing over time. The institute will not succeed if participants absorb a great deal of XQuery during their two week residency at Vanderbilt but do not implement any practical projects after leaving. An emphasis of the course will be to get participants up and running with an eXist database loaded with documents of relevance to their research so that they can continue refining and developing their XQuery skills. We recognize that we must take account of several potential barriers to achieve this result.

- Some scholars may not receive permission from their I.T. groups to host an eXist database on their institutional networks. Our institute will work around this limitation by showing how to install the eXist database in the "cloud" at

low cost. (Of course, it is possible to install eXist on a private computer as well.)

- Some scholars will not have collaborators after leaving the institute. We will encourage them either to list their projects on DH Commons or to assist other scholars on DH Commons with querying their XML data (see (http://dhcommons.org/projects)
- Some scholars will get swamped with other academic work and leave off working with XQuery. We will provide moral and technical support by staying in touch with participants via a listserv during the six months following the institute, providing practical tips about working with XQuery and offering suggestions for new scholarly applications.

The institute will publish the entire curriculum of the institute, including all code samples and documentation, to GitHub (https://github.com/). GitHub provides free hosting for open source projects with excellent features for collaborative, shared coding, including documentary web pages, wikis, and version control. In particular, XQuery code from the institute will be shared as Gists (https://gist.github.com/), making it easy to pass them around, adapt them, and drop them into personal projects. Since many digital humanists are already using GitHub to host their projects, our institute will have good visibility among members of the digital humanities community. We will maintain our GitHub site in an open manner—inviting nonparticipants to contribute comments and code alongside participants. We anticipate that our GitHub site will become a central hub for the XQuery community in the digital humanities.

We will evaluate the success of the institute in two ways. First, we will ask participants to provide feedback through a survey mechanism at the conclusion of the institute. We will hope to identify aspects of our curriculum, teaching style, and facilities that may be improved. Six months after the conclusion of the institute, we will send an additional survey to participants to discern its impact on participants' scholarly practices. We will ask, among other things, whether participants continue to work with XQuery on a regular basis and, if so, what have they found most useful to their scholarship. The successful outcome of the institute will be measured primarily by this second survey.

We expect that the institute will have significant direct and indirect impact on digital humanists who work with XML documents. We anticipate that participants in the institute will gain sufficient proficiency in XQuery to work computationally with their XML documents, querying them for specific items of information, searching the text nodes, combining their XML documents with other data sources, and rendering them publicly through interactive websites. We also anticipate indirect effects from the institute, including publicity for XQuery, XML databases, and XML-based markup languages within the digital humanities community. Indeed, an important goal of the institute is to create a stronger, more cohesive community of XQuery users in the digital humanities. We hope that such publicity will raise the profile of open source

projects like eXist among digital humanists and encourage aspiring humanities programmers to consider learning XQuery as their primary programming language.

Staff and Consultants

Clifford Anderson is the Director for Scholarly Communications at Vanderbilt University. He will be the director of the institute and a primary instructor. Anderson has a decade of experience developing and deploying XML-based digital projects. He was among the first in the academic world to develop XQuery-based applications using native XML databases. He has developed many digital projects using XQuery, including the *Theological Commons*, a full-text database with more than 75,000 books and articles (http://commons.ptsem.edu). Anderson has also led a training series in XQuery for librarians, helping non-programmers to become productive coders. He has a M.Div. from Harvard Divinity School, a Th.M. and Ph.D. from Princeton Theological Seminary, and a M.S.L.I.S. from the Pratt Institute.

Kevin Clarke is Institutional Repository Programmer at the University of California Los Angeles. Previously, he was the Digital Library Applications Programmer for the Grateful Dead Archive at the University of California Santa Cruz. He has taught introductory classes on XQuery for, among other institutions, the Library of Congress. He has a M.S.I.S. from the University of North Carolina at Chapel Hill. Clarke will serve as a visiting lecturer during the final three days of the institute, assisting with the development of a search application for the digital texts of Shakespeare.

David Michelson is Assistant Professor of Early Christianity and Classical Studies at Vanderbilt University. An historian, Michelson is currently the director of *The Syriac Reference Portal* (www.syriaca.org) a digital reference project for Middle Eastern history. *The Syriac Reference Portal* employs XML and the semantic web to publish linked open historical data. This project is currently funded by the NEH (Division of Preservation and Access). Michelson has taught research methods in the digital humanities at the graduate and undergraduate level and has been a previous participant in the NEH/ODH-funded Linked Ancient World Data Institute. Michelson has a M.A. from Trinity Evangelical Divinity School and an M.A. and Ph.D. from Princeton University. Michelson will co-lead instruction during the first two days and assist with the afternoon sessions throughout the institute.

Dale Poulter is the Coordinator of Search and Core Services at Vanderbilt University. He received his M.L.I.S. from University of North Texas and B.S. Ch. E from Texas Tech. Dale has been working with XML for more than five years. His main focus has been manipulating data from various XML sources to be used in web applications. Poulter will assist throughout the institute and co-lead instruction during the final day.

Jonathan Robie is a Senior Consultant at EMC Corporation, working primarily on query architecture and REST architecture. He is the Lead Editor of the XQuery and

XPath specifications at the World Wide Web Consortium (W3C) and a co-editor of JSONiq, a query language for JSON based on XQuery. He is also the list owner of B-Greek and biblicalhumanities.org. Jonathan has a M.S. in Computer Science from Michigan State University. Robie will serve as a visiting lecturer, providing an introduction to XPath and XQuery during the third, fourth, and fifth days of the institute.

Winona Salesky is an independent digital library consultant with 10 years' experience building digital collections with XML technologies, including XQuery, XSLT, and native XML databases. She was previously the Digital Initiatives Librarian at the University of Vermont where she developed and deployed The Center for Digital Initiatives (http://cdi.uvm.edu), an entirely XML based digital library project run on eXist. Salesky has a M.L.I.S. from Rutgers University. Salesky will serve as a visiting instructor, co-leading instruction on the eXist database from Monday to Thursday during the second week of the institute.

Vanderbilt University
Application to the National Endowment for Humanities
Institutes for Advanced Topics in the Digital Humanities

Detailed Curriculum:
Schedule and Readings

Monday, June 9

- Morning Session I (9:00 a.m. to 10:15 a.m.) – Clifford B. Anderson
  Introduction to the course and participants; discussion of the institute's goals and anticipated outcomes

- Morning Session II (10:30 a.m. to 11:45 a.m.) – David A. Michelson
  Review of XML essentials, including namespaces, character entities, and CDATA sections

  Readings: Joe Fawcett, Liam Quin, and Danny Ayers, *Beginning XML,* Chapter 1: What is XML? and Chapter 3: XML Namespaces

- Afternoon Session I (1:00 p.m. to 2:15 p.m.) – Clifford B. Anderson
  Review of Document Type Definitions (DTDs) and XML Schemas; a brief look at Relax NG and Schematron

  Readings*:* Joe Fawcett, Liam Quin, and Danny Ayers, *Beginning XML*, Chapter 4: Document Type Definitions; Chapter 5: XML Schemas; Chapter 6: Relax NG and Schematron

- Afternoon Session II (2:30 p.m. to 3:45 p.m.) – Institute Faculty
  Practical exercises encoding sample documents in XML

- Open session (4:00 p.m. to 5:00 p.m.)
  Commentary, discussion, questions

Tuesday, June 10

- Morning Session I (9:00 a.m. to 10:15 a.m.) – David A. Michelson
  The challenges of scholarly editing in XML

  Reading: Hans Walter Gabler, "Theorizing the Digital Scholarly Edition" in *Literature Compass* 7:2 (2010): 43-56.

- Morning Session II (10:30 a.m. to 11:45 a.m.) – David A. Michelson
  Introduction to the TEI P5 specification

  Reading: *P5: Guidelines for Electronic Text Encoding and Interchange*, Chapter 3: Elements Available in All TEI Documents

- Afternoon Session I (1:00 p.m. to 2:15 p.m.) – Institute Faculty
  Exploration of the Folger Library's digital edition of *Julius Caesar* (TEI P5)

- Afternoon Session II (2:30 p.m. to 3:45 p.m.) – Institute Faculty
  Questions that we would like to ask across the corpus of the Folger Library's digital edition of Shakespeare

- Open session (4:00 p.m. to 5:00 p.m.)
  Commentary, discussion, questions

Wednesday, June 11

- Morning Session I (9:00 a.m. to 10:15 a.m.) – Jonathan Robie
  Introduction to XPath; XPath axes (unabbreviated and abbreviated syntax)

  Readings: Joe Fawcett, Liam Quin, and Danny Ayers, *Beginning XML*, Chapter 7: Extracting Data from XML

- Morning Session II (10:30 a.m. to 11:45 a.m.) – Jonathan Robie
  XPath node tests and predicates

  Reading: Priscilla Walmsley, *XQuery*, Chapter 4: Navigating Input Documents Using Paths

- Afternoon Session I (1:00 p.m. to 2:15 p.m.) – Jonathan Robie
  XPath operators and functions

- Afternoon Session II (2:30 p.m. to 3:45 p.m.) – Institute Faculty
  Exploring *Julius Ceasar* using XPath

- Open session (4:00 p.m. to 5:00 p.m.)
  Commentary, discussion, questions

Thursday, June 12

- Morning Session I (9:00 a.m. to 10:15 a.m.) – Jonathan Robie
  Introduction to XQuery

Readings: Ron Hitchens, G*etting Started with XML*, Chapter 4: XQuery, Start your Engine and Chapter 5: Taking the Plunge; Priscilla Walmsley, *XQuery*, Chapter 1: Introduction to XQuery

- Morning Session II (10:30 a.m. to 11:45 a.m.) – Jonathan Robie
  Writing simple XQuery FLWOR expressions -

  Readings: Ron Hitchens, G*etting Started with XML*, Chapter 7: The Power of the FLWOR; Priscilla Walmsley, *XQuery*, Chapter 6: Selecting and Joining Using FLWORs

- Afternoon Session I (1:00 p.m. to 2:15 p.m.) – Jonathan Robie
  The XQuery type system; atomic types and sequence types; checking and casting types

  Reading: Ron Hitchens, G*etting Started with XML*, Chapter 5: Just Typical; Priscilla Walmsley, *XQuery*, Chapter 2: XQuery Foundations

- Afternoon Session II (2:30 p.m. to 3:45 p.m.) – Institute Faculty
  Querying Shakespeare's corpus using XQuery FLWOR expressions

- Open session (4:00 p.m. to 5:00 p.m.)
  Commentary, discussion, questions

Friday, June 13

- Morning Session I (9:00 a.m. to 10:15 a.m.) – Jonathan Robie
  Using built-in functions in XQuery: sequence functions

  Reading: Priscilla Walmsley, *XQuery*, Chapter 11: A Closer Look at Types

- Morning Session II (10:30 a.m. to 11:45 a.m.) – Jonathan Robie
  Using built-in functions in XQuery: string functions

  Reading: Priscilla Walmsley, *XQuery*, Chapter 17: Working with Strings

- Afternoon Session I (1:00 p.m. to 2:15 p.m.) – Jonathan Robie
  Other XQuery control structures: if…then…else; switch expressions; group by

  Reading: Priscilla Walmsley, *XQuery*, Chapter 3: XQuery Building Blocks

- Afternoon Session II (2:30 p.m. to 3:45 p.m.) – Institute Faculty
  Using built-in functions with FLWOR expressions to render Shakespeare as XHTML

- Open session (4:00 p.m. to 5:00 p.m.)
  Commentary, discussion, questions

Monday, June 16

- Morning Session I (9:00 a.m. to 10:15 a.m.) – Winona Salesky
  Exploring eXist- An Open Source XML Database

  Reading: Joe Fawcett, Liam Quin, and Danny Ayers, *Beginning XML*, Chapter 10: XML and Databases

- Morning Session II (10:30 a.m. to 11:45 a.m.) – Winona Salesky
  Combining eXist's extension functions with FLWOR expressions

- Afternoon Session I (1:00 p.m. to 2:15 p.m.) – Winona Salesky
  Experimenting with eXist's full-text search extension functions; a brief look at the XQuery and XPath Full Text 1.0 recommendation

  Reading: "XQuery Function Documentation" (http://exist-db.org/exist/apps/fundocs/browse.html?extensions=true)

- Afternoon Session II (2:30 p.m. to 3:45 p.m.) – Institute Faculty
  Conducting full-text searches across Shakespeare's corpus using eXist's extension functions

- Open session (4:00 p.m. to 5:00 p.m.
  Participant presentation

Tuesday, June 17

- Morning Session I (9:00 a.m. to 10:15 a.m.) – Winona Salesky
  Writing user defined functions; function signatures; checking types

  Reading: Ron Hitchens, G*etting Started with XML*, Chapter 9: Functionality; Priscilla Walmsley, *XQuery*, Chapter 8: Functions

- Morning Session II (10:30 a.m. to 11:45 a.m.) – Clifford Anderson
  Handling errors in XQuery; try…catch; strategies for diagnosing and preventing errors

  Reading: "Try-Catch Expression" (http://atomic.exist-db.org/HowTo/XQuery3/Try-CatchExpression)

- Afternoon Session I (1:00 p.m. to 2:15 p.m.) – Winona Salesky
  Developing modules of user defined functions in eXist

  Reading: Ron Hitchens, G*etting Started with XML*, Chapter 8: Modulation; Priscilla Walmsley, *XQuery*, Chapter 12: Queries, Prologs, and Modules

- Afternoon Session II (2:30 p.m. to 3:45 p.m.) – Institute Faculty
  Rewriting our search expressions as custom functions and modules

- Open session (4:00 p.m. to 5:00 p.m.)
  Participant presentation

Wednesday, June 18

- Morning Session I (9:00 a.m. to 10:15 a.m.) – Kevin Clarke
  Reviewing HTTP; a brief introduction to REST

  Reading: Cornelia Davis, "What if the web were not RESTful?" In *Proceedings of the Third International Workshop on RESTful Design (WS-REST '12)*, Rosa Alarcon, Cesare Pautasso, and Erik Wilde (Eds.). New York: ACM, 3-10

- Morning Session II (10:30 a.m. to 11:45 a.m.) – Winona Salesky
  Introduction to eXist's httpclient extension functions

- Afternoon Session I (1:00 p.m. to 2:15 p.m.) – Institute Faculty
  Connecting with online data sources using HTTP

  Reading: Charles Severance, "Discovering JavaScript Object Notation," *Computer*, 45:4 (2012): 6-8

- Afternoon Session II (2:30 p.m. to 3:45 p.m.) – Institute Faculty
  Writing a module to extract entities such as people, places, and events from the Shakespeare corpus

- Open session (4:00 p.m. to 5:00 p.m.)

  Participant presentation

Thursday, June 19

- Morning Session I (9:00 a.m. to 10:15 a.m.) – Kevin Clarke

Architectural patterns for building larger applications; introduction to the Model-View-Controller pattern

Reading: Cerstin Mahlow, et. al., "A Framework for Retrieval and Annotation in Digital Humanities using XQuery Full Text and Update in BaseX" in *Proceedings of the 2012 ACM symposium on Document engineering (DocEng '12)* (New York: ACM, 2012): 195-204.

- Morning Session II (10:30 a.m. to 11:45 a.m.) – Winona Salesky
  Understanding and configuring eXist's indexes

  Reading: "Configuring Database Indexes" (http://exist-db.org/exist/apps/doc/indexing.xml)

- Afternoon Session I (1:00 p.m. to 2:15 p.m.) – Institute Faculty
  Building a simple web search engine for the Shakespeare corpus – Part One

- Afternoon Session II (2:30 p.m. to 3:45 p.m.) – Institute Faculty
  Building a simple web search engine for the Shakespeare corpus – Part Two

- Open session (4:00 p.m. to 5:00 p.m.)
  Participant presentation

Friday, June 20

- Morning Session I (9:00 a.m. to 10:15 a.m.) – Dale Poulter
  Deploying our Shakespeare search engine to the "cloud" using Amazon Web Services – Part One

- Morning Session II (10:30 a.m. to 11:45 a.m.) – Dale Poulter
  Deploying our Shakespeare search engine to the "cloud" using Amazon Web Services – Part Two

- Afternoon Session I (1:00 p.m. to 2:15 p.m.) – Kevin Clarke
  Review of XQuery fundamentals and questions

- Afternoon Session II (2:30 p.m. to 3:45 p.m.) – Kevin Clarke
  Review of eXist fundamentals and questions

- Open session (4:00 p.m. to 5:00 p.m.) – Clifford B. Anderson
  Institute wrap up; survey; recommendations for developing a stronger community of XQuery experts in the digital humanities