

## ***HOW TO DO THINGS WITH MILLIONS OF WORDS***

### **A Multi-Disciplinary Digital Humanities Workshop for Experts, Novices, the Sceptical and the Curious**

*co-presented by*

Green College (UBC), The Early Modern Conversions Project (McGill University) and the Cultural Evolution of Religion Consortium (UBC)

**Wednesday 2 November – Friday 4 November, 2016**

Green College UBC, 6201 Cecil Green Park Road, Vancouver

\*\*\*

*How To Do Things With Millions of Words* brings an international group of researchers together at Green College on the Vancouver campus of the University of British Columbia for an intensive two-day meditation on the theoretical and technical challenges of macro-scale textual analysis. On the theoretical side, participants will work toward a better understanding of the limits and optimal uses of technologies for measuring very large corpora: How can researchers most effectively integrate computer-assisted techniques with more conventional forms of scholarship? What sort of questions and conclusions work best with this style of research? In what ways does macro-scale textual analysis count as a form of reading? Key areas of discussion on the technical side will include topic modeling, data mining, data visualization and other techniques for interrogating, interpreting or otherwise engaging with texts at the scale of millions—or even billions—of words.

The workshop is open to members of the UBC and wider academic community without charge. *Pre-registration is required*, except for the pre-workshop lecture at Green College on Wednesday 2 November and the noon-hour public lecture at the Peter A. Allard School of Law on Friday 4 November. To register, contact [gc.programs@ubc.ca](mailto:gc.programs@ubc.ca) no later than Monday 24 October. Places are limited. Priority will be given to graduate students and UBC faculty and staff.

\*\*\*

## Wednesday 2 November

Pre-workshop lecture and piano recital at Green College

5:00-6:15 pm Public lecture (Coach House) [45 minutes + < 30 mins Q&A]

### **THE DIGITAL HUMANITIES: SOME CONTEXTS, CONCEPTS AND INITIATIVES**

Ray Siemens

6:15-7:00 *Reception* (Piano Lounge)

6:15-8:00 *Dinner* (Great Hall)

8:00-9:00 Piano recital by Réa Beaumont, “Key Notes for a Conversation Piece”  
(Piano Lounge)

\*\*\*

## Thursday 3 November

Day 1 of workshop: all sessions in Coach House, Green College UBC

*Minibus departs Pinnacle Hotel Harbourfront for Green College*

### **9:00 - 11:00 am      Session 1**

Welcome and opening remarks

[Three talks: each 15 minutes + 15 minutes Q&A]

Chair: Mark Vessey

### **NEAR HORIZONS IN DIGITAL HUMANITIES RESEARCH**

Christopher Donaldson *Implementing GIS and Corpus Analysis to Investigate  
Historical Travel Writing and Topographical Literature about the English  
Lake District*

Devin Vartija and Melvin Wevers *Using Text Mining to Trace the  
Evolution of Enlightenment Encyclopaedias*

Michael Ulliyot *Unnatural Language and Natural Thinking: Shakespeare and  
His Contemporaries*

11:00 - 11:30 *Break*

**11:30 am - 12:30 pm Session 2**

[Team presentation: 30 minutes + 30 minutes Q&A]

Chair: Bruce Rusk

**THE DISTANT READING OF RELIGIOUS TEXTS: A “BIG DATA”  
APPROACH TO MIND-BODY CONCEPTS IN EARLY CHINA**

Ryan Nichols, Kristoffer Laigaard Nielbo and Edward Slingerland

12:30 - 2:30 pm      *Lunch*

**2:30 - 4:30                      Session 3**

[Two team presentations: each 30 minutes + 30 minutes Q&A]

Chair: Paul Yachnin

**NEW VISIONS OF EARLY MODERN ENGLAND: NAVIGATING THE  
CITY AND THE TEXTS**

Janelle Jenstad, Tye Landels-Gruenwald and Joseph Takeda  
*How to Do Things With 6000 Toponyms: MoEML Mines DEEP*  
Kosta Gligorijevic, Catherine Nygren and Stephen Wittek  
*Distant Reading Early Modernity (DREaM)*

4:30 - 5:00              *Break*

**5:00 - 6:00                      Session 4**

[Team presentation: 30 minutes + 30 minutes Q&A]

Chair: Jessica Wang

**DESIGNING TOOLS FOR MACRO-SCALE DATA ANALYSIS IN THE  
HISTORY OF SCIENCE**

Elyse Graham and Robert Crease

6:15 - 6:45              *Reception*

6:15 – 7:30              *Dinner served at the College*

*Minibus departs Green College for Pinnacle Hotel Harbourfront*

\*\*\*

## Friday November 4

Day 2 of workshop: all sessions at Green College except noon-hour lecture

*Minibus departs Pinnacle Hotel Harbourfront for Green College*

NOTE: Sessions 5A and 5B will run in parallel

### **9:00 - 11:00 am      Session 5A** (Coach House)

[Three talks: each 15 minutes + 15 minutes Q&A]

Chair: Patricia Badir

#### **DISTANT READING PROJECTS: AUTHORS, GENRES, READERS**

Amrita Dhar *Distant Reading Blind Language?*

Matthew Lavin *Continuity and Discontinuity in Gothic and Horror*

Allison Hegel *Reading Genre in Online Book Reviews*

### **9:00 - 11:00 am      Session 5B** (Gibson Room)

[Three talks: each 15 minutes + 15 minutes Q&A]

Chair: Matthew Milner

#### **ANALYZING COMPLEX LITERARY CORPORA: RHETORIC, STYLE AND OTHER SPECIAL EFFECTS**

Eleni Bozia *Evaluating Attic, Imperial Greek and Roman Oratory:*

*Towards a Schema of Rhetorical Constructions*

Anupam Basu *Spenser, Linguistic Change and Historical Stylometrics*

Elise Thorsen *Refactoring the Bag of Words into Meter, Rhythm, Rhyme  
in Russian Verse Corpora*

11:00-11:30      *Break*

### **12:00 - 1:00 pm      Session 6** (The Lew Forum, Peter A. Allard School of Law)

[Public lecture: 40 mins + 10 mins Q&A]

Chair: Ted Slingerland

Introduction: Stéfan Sinclair

**HUNDREDS, THOUSANDS OR MILLIONS OF WORDS: FITTING OUR METHODS TO THE QUESTION**

Ted Underwood

1:15 - 2:45                      *Lunch*

**2:45 - 4:00    Session 7** (Coach House)

[Two talks: each 15 minutes + 15 minutes Q&A]

**TOPIC MODELING WITH GOVERNMENT CORPORA**

Chair: Larissa Ringham

Fredrik Norén *Finding Co-Occurring Topics with MALLET in Swedish Governmental Reports*

Fabian Brinkmann *Topic Modeling Approaches in Research on Turkish Foreign Policy in Sub-Saharan Africa*

4:00 - 4:30                      *Break*

**5:00 - 6:30    Session 8** (Coach House)

[Team Presentation: 30 minutes + 30 minutes Q&A]

Chair: Stephen Wittek

**INTERVENTIONS IN ACADEMIC COMMUNITY**

Iter Community, the Renaissance Knowledge Network, Social Knowledge Creation and Open Social Scholarship

Alyssa Arbuckle, Lindsey Seatter, Randa El Khatib, Tracey El Hajj

Closing Remarks: Ted Slingerland, Paul Yachnin

*Minibus departs Green College for Pinnacle Hotel Harbourfront*

## SPONSORS

### **Early Modern Conversions**

[www.earlymodernconversions.com](http://www.earlymodernconversions.com)

The five-year, \$2.3 million project, “Early Modern Conversions,” is based at McGill University’s Institute for the Public Life of Arts and Ideas (IPLAI), and is funded by the Social Sciences and Humanities Research Council of Canada (SSHRC), with additional support from artistic and academic partner institutions in Australia, the United Kingdom, and the United States. With the goal of developing a better understanding of the diverse range of conversional activity that spread across Europe and the New World from approximately 1400-1700, the project has brought together an international, interdisciplinary team of scholars in Art History, Architecture, Literature, Music, Theatre, Religion, History of Cognition, and Digital Humanities.

### **Green College**

[www.greencollege.ubc.ca](http://www.greencollege.ubc.ca)

Green College (founded 1993) is a graduate residential facility at the University of British Columbia with a mandate to promote advanced interdisciplinary inquiry and engagement between academics and society at large. The College is committed to the cultivation of intellectual and creative connections at the edge of the main disciplinary space of the University. To that end, and in partnership with other UBC units and locally based non-profit organizations, it provides extracurricular academic, artistic and cultural programs that are open to all-comers at no charge.

### **The Cultural Evolution of Religion Research Consortium**

[www.hecc.ubc.ca/cerc/project-summary/](http://www.hecc.ubc.ca/cerc/project-summary/)

The Cultural Evolution of Religion Research Consortium (CERC), centred at UBC, brings together the expertise of over fifty scientists, social scientists and humanities scholars from universities across North America, Europe and East Asia—along with postdocs and graduate students—into a research network that aims to answer the question of what religion is, how it is linked to morality, and why it plays such a ubiquitous role in human existence. Funded by a \$3 million Partnership Grant from the Canadian Social Sciences and Humanities Research Council (SSHRC), with approximately \$3 million in matching funds from partner institutions and other contributors, CERC is currently in the fifth year of its six years of funding. For details, see <http://www.hecc.ubc.ca/cerc/project-summary/>

## BIOGRAPHIES AND ABSTRACTS FOR INDIVIDUAL PRESENTERS

**Basu, Anupram** <[prime.lens@gmail.com](mailto:prime.lens@gmail.com)>

Anupam Basu is an Assistant Professor of English at Washington University in Saint Louis. He works at the intersection of literature and computational analysis, drawing on emerging techniques to make vast digital archives of early modern print more tractable for scholars. The ngram and collocation browser at <http://earlyprint.wustl.edu> makes some of the data behind his research accessible to a wider audience. His current research book project on crime and social change in Tudor and Stuart literature explores the popular representation of criminality, poverty and vagrancy in the period.

Presentation: *Spenser, Linguistic Change, and Historical Stylometrics*

This paper will investigate one of the basic truisms about Edmund Spenser's language – that it is distinguished by more or less systematic archaizing. A long editorial and critical tradition has held that the orthography of Spenser's texts was distinctive and that Spenser extended the normal latitudes of mid-sixteenth-century practice to achieve particular effects. Situating Spenser's texts within the general evolution of orthographic change over the first two centuries of English print will let us explore the ways in which early readers might have read these texts and how they might have interpreted Spenser's spelling and lexical choices as indicators of his poetic archaism.

**Bozia, Eleni** <[bozia@ufl.edu](mailto:bozia@ufl.edu)>

Eleni Bozia is an Assistant Professor in the Department of Classics and the Digital Worlds Institute at the University of Florida. She is currently pursuing a Dr. Phil. at the Universität Leipzig. Dr. Bozia also serves as the Associate Director of the *Digital Epigraphy and Archaeology Project*. She is the author of the book *Lucian and his Roman Voices: Cultural Exchanges and Conflicts in the Late Roman Empire*.

Dr. Bozia is the recipient of collaborative grants from the National Endowment for the Humanities and the French Ministry of Education and several national and international Awards, including the Young Researcher Fellowship from La Fondation Hardt and the Mary A. Sollman Scholarship of the American Academy in Rome.

Presentation: *Evaluating Attic and Imperial Greek Oratory: Towards a Schema of Rhetorical Constructions*

This paper attempts to quantify compositional complexity in Greek oratory via the syntactical annotation of Classical Attic orators and orators of the Imperial Era against the backdrop of Dionysius of Halicarnassus. For the stylistic analysis of rhetorical structures, passages from speeches of Lysias, Isocrates, Demosthenes, Dio Chrysostom, Lucian, and Aelius Aristides were annotated syntactically in the form of Treebank annotation, using the Arethusa annotation framework through Perseids.<sup>1</sup> Additionally, I developed a unified node-based metric formulation for implementing syntactical construction metrics. These metrics were then applied to the annotated texts, which were comparatively examined using Principal Component Analysis. The results furnish patterns of syntactical constructions that may apprise us of a schema of syntactical constructions.

**Brinkman, Fabian** <Fabian.Brinkmann@ruhr-uni-bochum.de>

Fabian Brinkmann is a doctoral student at the Ruhr-University Bochum, Germany. Based at the interdisciplinary Centre for Mediterranean Studies and the chair for the history of the Ottoman Empire and Turkey he currently works on a dissertational project titled “A Computer-Assisted Analysis of Turkish Foreign Policy in Sub-Saharan Africa: Diachronic Topics and Networks (2002-2016)” funded by the German Academic Scholarship Foundation (Studienstiftung des Deutschen Volkes). Before starting on his doctoral studies he got degrees in History, Protestant Theology and Gender Studies from the Ruhr-University Bochum and the University of Basel, Switzerland.

Presentation: *Topic Modeling Approaches in Research on Turkish Foreign Policy in Sub-Saharan Africa*

Many researchers have dealt with the different geopolitical spaces of Turkish foreign policy under the AKP. However, Turkish policy in Sub-Saharan Africa still remains a blind spot and the networks of actors and different political discourses have not yet been highlighted by scholars. This paper will show how Structural Topic Modeling can be used to uncover discursive structures and (discursive) networks of actors. It will describe the attempt to untangle the different discourses across intertwined actors via the mass data approach of Topic Modeling across different covariates in a diachronic and synchronic way, using a sample of texts from this discursive field.

**Dhar, Amrita** <amritad@umich.edu>

Amrita Dhar was educated at Jadavpur and Cambridge universities before coming to the University of Michigan for her doctoral studies. Her thesis “Writing Sight and Blindness in Early Modern England” examines the perceptions, representations and negotiations of sight and blindness in England between 1564 and 1674. Her passion for the archives engages her in palaeography just as her passion for information science and new methods of scholarship engages her increasingly in the digital humanities. Outside her early modern interests, she is also an active traveller of mountains and coordinates the University of Michigan’s Mountaineering Culture Studies Group.

Presentation: *Distant Reading Blind Language?*

Can we use distant reading methods to register fundamental changes in language generated by differently-sighted, or blind, writers? In my dissertation, I use close reading to excavate and analyse John Milton’s blind language: language generated by a man whose intensely sighted training perforce adapts to an irremediable visual affliction and becomes profoundly generative. Although epic in scope and size, Milton’s blind poetry nevertheless constitutes a limited corpus to study blind and brilliantly aural mnemonics, syntax, and cohesion. This makes me wonder about other—larger—corpora of differently-abled or differently-sighted work. Can distant reading help us understand how writerly language changes with changes in cognitive abilities? Using a preliminary case study of Milton’s blind language, my presentation will explore possible ways to



interrogate other early modern texts conceived of and executed in blindness, all the while asking how language bears the imprint of a differently-abled creative consciousness.

**Donaldson, Christopher** <C.Donaldson@bham.ac.uk>

Christopher Donaldson is Lecturer in Regional History in the Department of History at Lancaster University, and a specialist in the cultural history of the English Lake District. Chris is an investigator on the Leverhulme Trust research project “Geospatial Innovation in the Digital Humanities: A Deep Mapping of the English Lake District” (2015-2018); he is also co-editor of the Digital Forum of the *Journal of Victorian Culture* and co-editor of *Literary Mapping in the Digital Age* (Routledge, 2016).

Presentation: *Implementing GIS and Corpus Analysis to Investigate Historical Travel Writing and Topographical Literature about the English Lake District*

This paper models the implementation of an interdisciplinary approach to the examination of historical corpora. It presents a case study that combines corpus analysis, automated geo-parsing, and GIS technology in order to investigate the geographies underlying some of the key aesthetic terminology historically used in writing about the English Lake District. The focus of this investigation is a 1,500,000+ word corpus of writing about the Lake District that comprises works published between the years 1622 and 1900. The paper exemplifies how a hybrid geographical and corpus-based methodology can be used to study historical relationships between landscape writing and physical geography.

**Hegel, Allison** <ahegel@ucla.edu>

Allison Hegel is a PhD candidate in the Department of English at UCLA. Her dissertation, “Social Reading in the Digital Age,” uses digital humanities methods to investigate contemporary literature and how we talk about that literature online. She is also an editor for The Programming Historian, a website that publishes peer-reviewed tutorials to help humanists learn to use digital tools.

Presentation: *Reading Genre in Online Book Reviews*

How well do established literary theories of genre line up with how contemporary readers discuss and experience genre? For example, science fiction is often characterized as a genre of defamiliarization—so do readers often mention being surprised, or science fiction books being “weird”? Or do more outdated ways of classifying genre, such as the presence of tropes like aliens and spaceships, better express the way readers experience genre? In this presentation, I will explore these questions through a textual analysis of over 100,000 online book reviews ranging from the *Los Angeles Review of Books* to the social reading website Goodreads.com.

**Lavin, Matthew** <lavin@pitt.edu>

Matthew J. Lavin earned his PhD in English from the University of Iowa in 2012. His dissertation, “Collaborative Momentum: The Author and the Middle Man in U.S. Literature and Culture, 1890-1940” examines the supportive and adversarial relationships among authors, agents, editors, publishers and other literary go-betweens that helped construct modern authorial credibility. From 2012 to 2013, Lavin served as a Council on Library and Information Resources (CLIR) Postdoctoral Fellow at the University of Nebraska – Lincoln’s Center for Digital Research in the Humanities. From 2013 to 2015, he was Associate Program Coordinator for the Andrew W. Mellon Foundation Initiative “Crossing Boundaries: Re-Envisioning the Humanities for the 21st Century” at St. Lawrence University. He is currently a Clinical Assistant Professor of English and Director of the Digital Media Lab at the University of Pittsburgh. His most recent scholarship focuses on the intersection of digital humanities, book history and U.S. literature.

Presentation: *Continuity and Discontinuity in Gothic and Horror*

In his recent essay “The Life Cycles of Genres,” Ted Underwood suggests that quantitative methods are perhaps uniquely positioned to “allow scholars to coordinate textual and social approaches to genre.” The present paper contributes to digital humanities discussions of genre (and distant reading generally) by fleshing out what happened with the Gothic/horror genre between 1830 and 1945. It builds upon Underwood’s findings by expanding his dataset, experimenting with unsupervised learning to look for topical subcategories, and using multi-class/multi-label machine learning techniques to attempt to separate horror’s genre and date signals.

**Norén, Fredrik** <fredrik.noren@umu.se>

In 2014, Fredrik Norén started as a doctoral candidate in media and communication studies—geared towards media history and digital humanities—at Umeå University, where he also works at the digital humanities hub, HUMlab. His PhD project concerns the formation of governmental information in Sweden during the 1960s and 1970s and its relationship to the evolving media and information landscape at the time. His research uses different methodologies to capture this development, with digital text methods, particularly topic modeling, a cornerstone. From September to November 2016 Norén is a visiting student researcher at CESTA at Stanford.

Presentation: *Finding Co-Occurring Topics with LDA/Mallet*

Topic modeling with LDA and Mallet has long been practised and has proved to be a strong and fairly easy method to find clusters of words that tend to co-occur in large collections of text. Searching for co-occurring topics is less common, but nevertheless important in order to gain better knowledge and understanding of the latent themes within a corpus. In this talk I will present a way to identify co-occurring topics by using the weight measure from Mallet’s output data. The text corpus I am using is a collection of 7000 Swedish Governmental Official Reports.

**Siemens, Ray** <siemens@uvic.ca>

Ray Siemens is Distinguished Professor in the Faculty of Humanities at the University of Victoria, in English and Computer Science, and formerly holder of a Canada Research Chair in Humanities Computing (2004-2015). He is founding editor of the electronic scholarly journal *Early Modern Literary Studies* (at UBC, 1994), and his publications include Blackwell's *Companion to Digital Humanities* (2004, 2015 with Schreibman and Unsworth), Blackwell's *Companion to Digital Literary Studies* (2007, with Schreibman), *A Social Edition of the Devonshire MS* (2012, 2015; MRTS/Iter, Wikibooks), *Literary Studies in the Digital Age* (2014; MLA, with Price), and *The Lyrics of the Henry VIII MS* (2016; RETS). He directs the Implementing New Knowledge Environments project, the Digital Humanities Summer Institute, and the Electronic Textual Cultures Lab, recently serving also as Vice President / Director of the Canadian Federation of the Humanities and Social Sciences for Research Dissemination, Chair of the MLA Committee on Scholarly Editions, and Chair of the international Alliance of Digital Humanities Organisations. <http://web.uvic.ca/~siemens/>

Presentation: *The Digital Humanities: Some Contexts, Concepts and Initiatives*

This introductory talk lays out some key contexts and concepts in the digital humanities, focusing on themes such as social knowledge creation, consensus-driven pedagogical communities and collaborative research partnership. Examples are drawn from the speaker's experience as director of several digital humanities interventions into research, scholarship, and teaching, including the Implementing New Knowledge Environments Major Collaborative Research Initiative funded by the Social Sciences and Humanities Research Council of Canada (<http://inke.ca>), the Digital Humanities Summer Institute (<http://www.dhsi.org>), and a project in "social" editing ([http://en.wikibooks.org/wiki/The\\_Devonshire\\_Manuscript](http://en.wikibooks.org/wiki/The_Devonshire_Manuscript)). Ultimately, the talk urges a closer examination of what lies at the intersection of computing and the humanities as a necessary step in the humanities' digital self-determination.

**Thorsen, Elise** <enthorsen@gmail.com>

Elise Thorsen's research in digital humanities deals with computational methods for enabling corpus-scale analysis of unprocessed Russian verse texts, including at the structural level. Published work on this subject can be found at "Meter, Rhythm, and Rhyme" (<http://poetry.obdurodon.org>, co-authored with Dr. David J. Birnbaum). She holds a PhD in Russian Literature from the University of Pittsburgh, where her dissertation, "Territory and Empire in Early Soviet Poetry," examined the global spatial implications of revolution and the legacy of imperial colonialism in interwar Soviet civic poetry. Elise Thorsen currently combines language and area-studies expertise with metadata-driven analysis in open-source advanced analytics projects.

Presentation: *Refactoring the Bag of Words into Meter, Rhythm, Rhyme in Russian Verse Corpora*

While of smaller volume than prose corpora, poetic corpora are larger than they may appear to be by virtue of the information they encode not only at and above, but also below the level of the word. Poetry privileges properties rooted in phonetics (meter, rhyme and other regularities in sound texture) that are of only occasional importance in prose, providing literary information that should not be lost in computationally-assisted analysis. In this study of Russian verse, the operative questions orient toward how to do things (conceptually and in implementation) with information that is present in words without being directly accessible at the word level.

**Ullyot, Michael** <michael.ullyot@gmail.com>

Michael Ullyot is an Associate Professor of English at the University of Calgary, specializing in early modern literature and the digital humanities. He has published articles on anecdotes, abridgements and Edmund Spenser. His current projects include a monograph on the rhetoric of exemplarity, and a computer program that detects rhetorical figures of repetition and variation in literary texts.

Presentation: *Unnatural Language and Natural Thinking: Shakespeare and His Contemporaries*

Critics of computational text-analysis tend to perceive its focus on language patterns as a flattening of qualitative texts into quantifiable patterns. They're right. But a text's linguistic operating-system deserves close scrutiny when it reveals features of the text that a human reader can't perceive, or when it flags evidence beyond our capacity to gather. The Augmented Criticism Lab has developed algorithms to detect features of repetition and variation in the works of Shakespeare and his contemporaries (starting with drama, namely the Folger's Digital Anthology). We've begun with features like rhetorical figures that repeat lemmas (heed, heedful, heeding) or morphemes (heeding, wringing, vexing). We use natural-language processing to gather evidence of these unnatural formulations, to ask whether they signal natural habits of thought. The interpretive payoff is our ability to make more definitive arguments not just about these figures, but also about underlying cognitive habits. This paper describes our process and our corpus, and presents a range of our results with this initial corpus before we expand to the billion words in the EEBO-TCP corpus (1473-1700). For more information about the Augmented Criticism Lab, visit [acriticismlab.org](http://acriticismlab.org).

**Underwood, Ted** <tunder@illinois.edu>

Professor of English and Information Sciences at the University of Illinois, Ted Underwood was trained as a scholar of Romantic-era literature, and has written two books on that topic: *The Work of the Sun: Literature, Science, and Political Economy 1760-1860* (Palgrave, 2005) and *Why Literary Periods Mattered* (Stanford, 2013). More recently, he has been using quantitative methods to cast new light on long literary-historical timelines; that research will be gathered in a third book to be called *The Horizon of Literary History* (tentatively, late 2017).

Presentation: *Hundreds, Thousands or Millions of Words: Fitting Our Methods to the Question*

Many humanists rarely used numbers before discovering algorithms (like topic modeling) that are suited to exploring enormous libraries. As a result, our conversations about method tend to be organized by a strong opposition between “close” and “distant” approaches, aimed at radically different scales. But interesting things can be done at every scale of analysis — from a single passage, to dozens or hundreds of texts, to millions of volumes. In this talk, I’ll briefly survey that range of scales before dwelling on some approaches suited to the middle of the spectrum. Here scholars rarely have open-ended exploratory goals. Instead we begin with a loosely formulated theme, traced in a few examples, and need a way to scale up our question. In a situation like this, it can be valuable to keep human readers in the loop: we are, after all, very good at reading with a specific end in view. But algorithmic models also have unique strengths, catching certain kinds of patterns we would miss. I’ll describe ways of pairing human expertise with algorithmic flexibility to address questions about genre, gender, and the representation of time in fiction.

**Vartija, Devin** <D.J.Vartija@uu.nl> and **Wevers, Melvin** <M.J.H.F.Wevers@uu.nl> (co-presenters)

Devin Vartija is a PhD candidate in intellectual history at Utrecht University. His research focuses on the evolution of equality as a political concept in Enlightenment encyclopaedias and how this idea conflicted with a nascent, Eurocentric racial classificatory system. He holds a Bachelor of Arts and Science from McMaster University and a Master’s in History from Utrecht University.

Melvin Wevers is a PhD candidate at Utrecht University. His project is titled “Consuming America: The United States as a Reference Culture in Dutch Public Discourse on Consumer Goods.” He holds a Master’s degree in Psychology and American Studies, as well as a Research Master’s degree in Cultural Analysis. In 2016, Wevers was a fellow at the research program “Culture Analytics” organized by UCLA’s Institute of Applied Mathematics. His interests include digital humanities, cultural history, conceptual history, data science, and computational linguistics.

Presentation: *Using Text Mining to Trace the Evolution of Enlightenment Encyclopaedias*

In our paper, we demonstrate how to utilize computational techniques to trace Enlightenment religious and political thought in two influential eighteenth-century encyclopaedias: Diderot and d’Alembert’s *Encyclopédie* (1751–72) and De Felice’s *Encyclopédie d’Yverdon* (1770–80). We will both demonstrate and reflect on the usability within historical scholarship of a number of techniques, such as topic modelling, collocate extraction, and tf-idf comparison. The eventual goal of our set up is to trace the evolution of some key philosophical, political and religious concepts between these two encyclopaedias in order to sketch the thin coherence that characterized Enlightenment

thought, beyond the divergent Enlightenments that historians have been sketching over the past three to four decades.

BIOGRAPHIES AND ABSTRACTS GROUP PRESENTERS  
*in alphabetical order by name of first listed presenter*

**Arbuckle, Alyssa** <alyssaa@uvic.ca>, **El Hajj, Tracy** <traceyelhajj@gmail.com>, **El Khatib, Randa** <elkhatib.randa@gmail.com> and **Seatter, Lindsey** <lseatter@uvic.ca>

Presentation: *Interventions in Academic Community: Iter Community, the Renaissance Knowledge Network, Social Knowledge Creation and Open Social Scholarship*

This panel will discuss interventions and contextually-related environmental scans initiated by members of the Electronic Textual Cultures Lab at the University of Victoria (Canada). Alyssa Arbuckle leads the presentations of the panel, focusing on work toward defining and scoping the notion of *social knowledge creation*, a phrase intended to reflect “the emergent body of evidence and scholarly products that indicate a qualitative change in how knowledge is produced in a networked age ... shift[ing] from single-author, long-form prose research (especially in the humanities) to distributed models of intellectual production that intersect with social media connectivity, digitally facilitated coauthorship, and well-defined and traceable patterns of intellectual contribution” (Scholarly Communication Institute, [goo.gl/nyjN2J](http://goo.gl/nyjN2J)). An extension of this is the concept of *open social scholarship*; Tracey El Hajj will present on work toward understanding its intellectual situation at a time when some of its core tenets, including open access, are reflected in emerging scholarly practice and beginning to be codified in academic policy nationally and internationally (see Canadian Federation of Humanities and Social Sciences, [goo.gl/QY5A7k](http://goo.gl/QY5A7k)). Randa El Khatib will look at Iter Community (see [goo.gl/dDZRwS](http://goo.gl/dDZRwS)), an early-stage, iteratively-prototyped platform for social knowledge creation, situated as a critical intervention in scholarly production and publication in an identifiable academic community, and developed as part of a larger research partnership, which has also conceptualized the Renaissance Knowledge Network (ReKN, <http://rekn.itercommunity.org>). ReKN builds on a decade-long foundation of scholarly pursuit, which has also led to Early English Books Online and its Text Creation Partnership as well as the concept of the “social scholarly edition” ([goo.gl/8iBHCg](http://goo.gl/8iBHCg)). Lindsey Seatter will discuss REKn’s history, its current state as an amalgamative resource for early modern studies, and some thoughts toward its vision for the future.

Alyssa Arbuckle is the Assistant Director, Research Partnerships & Development, in the Electronic Textual Cultures Lab (ETCL) at the University of Victoria, where she works with the Implementing New Knowledge Environments (INKE) group and assists with the coordination of the Digital Humanities Summer Institute (DHSI). She is also an interdisciplinary PhD candidate at the University of Victoria, studying open social scholarship and its implementation (planned completion 2019). She holds a BA Honours in English from the University of British Columbia and an MA in English from the University of Victoria. Website: [alyssaarbuckle.com](http://alyssaarbuckle.com)

Tracey El Hajj is a PhD student in the English Department at the University of Victoria. She completed her Master's at the American University of Beirut where she developed a social networking tool that serves for literary purposes. Her current research focuses on programming for the humanities and 3D design. She is the technical developer at the Electronic Textual Cultures Lab at the University of Victoria, where she monitors the online presence of the lab's projects. Tracey is part of the IC, INKE, and Personas teams.

Randa El Khatib is an English Literature doctoral student at the University of Victoria, where she works on digital mapping tool development for humanities research and its implementation in Early Modern British literature. She is also a Research Assistant at the Electronic Textual Cultures Lab and is on the organizing team of the Digital Humanities Summer Institute. Randa is currently the local coordinator of the University of Victoria's campus-based open scholarship fellowships. She completed her BA and MA in English at the American University of Beirut.

Lindsey Seatter is a doctoral candidate in the Department of English at the University of Victoria. Studying the British Romantic period, she focuses on female writers and in the past has conducted detailed studies on works by Anna Barbauld, Mary Shelley, and Jane Austen. Lindsey completed her Master's degree at Simon Fraser University, where she developed an open-access, critical, digital anthology of short Romantic works by female writers with the aim of assisting undergraduate students in literary studies. Lindsey's dissertation is focused on exploring patterns across Austen's print and manuscript novels, reader engagement with narrative style, and online communities of practice. She works as a Research Assistant in the Electronic Textual Cultures Lab at the University of Victoria and is the Colloquium Assistant for the Digital Humanities Summer Institute Colloquium.

**Crease, Robert** <rcrease@notes.cc.sunysb.edu> and **Graham, Elyse** <jean.graham@stonybrook.edu>

Robert P. Crease is a professor in the Department of Philosophy at Stony Brook University, New York, and former chairman of the department. He has written, translated or edited over a dozen books on history and philosophy of science.

Elyse Graham is Assistant Professor of Digital Humanities at Stony Brook University and a research affiliate at MIT. Her first book, "The Republic of Games," is under advance contract with McGill-Queen's University Press.

Presentation: *Designing Tools for Macro-Scale Data Analysis in the History of Science*

Our project involves developing a new kind of digital resource to capture the history of research at scientific facilities in the era of the "New Big Science." The phrase "New Big Science" refers to the post-Cold War era at US national laboratories, when large-scale materials science accelerators rather than high-energy physics accelerators became marquee projects at most major basic research laboratories. The extent, scope and

diversity of research at such facilities makes keeping track of it difficult to compile using traditional historical methods and linear narratives; there are too many overlapping and bifurcating threads. In this talk, we will discuss existing methods of data collection and curation for a specific case project, the National Synchrotron Light Source Digital Archive. We will focus on strategies for curating open data sets and fostering communities of users to interpret and engage with the data. We are especially interested in the functional potential, in the context of this kind of tool development, of the humanistic concepts of narrative, metaphor, and performance.

**Jenstad, Janelle** <jenstad@uvic.ca>, **Landels-Gruenewald, Tye** <tyelandels@gmail.com>, and **Takeda, Joseph** <joey.takeda@gmail.com>

Janelle Jenstad is Associate Professor of English at the University of Victoria and Associate Coordinating Editor of the Internet Shakespeare Editions ([internetshakespeare.uvic.ca](http://internetshakespeare.uvic.ca)). She founded and directs The Map of Early Modern London ([mapoflondon.uvic.ca](http://mapoflondon.uvic.ca)). With Jennifer Roberts-Smith and Mark Kaethler, she co-edited *Shakespeare's Language in Digital Media* (Routledge). She has co-edited John Stow's *A Survey of London* (1598 text) for MoEML, and is currently editing *The Merchant of Venice* for the ISE. Her articles have appeared in *Elizabethan Theatre*, *Early Modern Literary Studies*, *Shakespeare Bulletin*, *Renaissance and Reformation*, and *The Journal of Medieval and Early Modern Studies*. She contributed chapters to *Approaches to Teaching Othello* (MLA); *Teaching Early Modern Literature from the Archives* (MLA); *Institutional Culture in Early Modern England* (Brill); *Shakespeare, Language, and the Stage* (Arden); *Performing Maternity in Early Modern England* (Ashgate); *New Directions in the Geohumanities* (Routledge); *Early Modern Studies and the Digital Turn* (Iter); *Making Humanities Matter* (Minnesota); and *Placing Names: Enriching and Integrating Gazetteers* (Indiana). For more details, see [janellejenstad.com](http://janellejenstad.com).

Tye Landels-Gruenewald is an MA candidate in English at Queen's University and consultant for the Map of Early Modern London. His current SSHRC-funded thesis project considers how the popularity of Shakespeare during the romantic period contributed to the emergence of secularism. In 2015, he published a blog series on "Georeferencing the Early Modern Booktrade" using NLP and other data mining techniques.

Joseph Takeda is a SSHRC-funded MA student in English (Science and Technology Research Stream) at the University of British Columbia and Junior Programmer and consultant for *The Map of Early Modern London*. He has written on the digital humanities and pedagogy (with Janelle Jenstad) and is currently working on papers discussing workflows in the digital humanities, and encoding and processing early modern calendars in TEI (with the MoEML team). For his MA, he is currently researching diasporic and Indigenous literature in Canada in relation to the entangled histories of colonialism, uranium mining and the atomic bomb.

Presentation: *How to Do Things With 6000 Toponyms: MoEML Mines DEEP*



Well-structured data, deeply tagged in highly granular ways, vastly improves the results of data mining and facilitates linking. MoEML's gazetteer, which aggregates 6000+ toponyms, was built by mining MoEML's own TEI-XML documents. The Database of Early English Playbooks (DEEP) has similarly well-formed data that is openly available, granular and comprehensive. Of particular interest to MoEML is their imprint line transcriptions and interpretive location metadata (from Pantzer's location identifiers in STC vol. 3). In this workshop, we explain the creation of our gazetteer and the process of parsing and processing DEEP's data against our toponymic data to argue for the role of editors and data curators in the process of data mining and linking. We show the power of mobilizing the different expertise in two teams to mine, link and enhance two discrete datasets.

**Slingerland, Edward** <edward.slingerland@gmail.com>, **Nichols, Ryan** <ryantatenichols@gmail.com> and **Nielbo, Kristoffer Laigaard** <nielbo23@gmail.com>

Edward Slingerland is Professor of Asian Studies at the University of British Columbia. His research specialties and teaching interests include Warring States (5th-3rd c. BCE), Chinese thought, religious studies, cognitive linguistics, ethics, and the relationship between the humanities and the natural sciences. His publications include several academic monographs and edited volumes, a trade book, and approximately thirty refereed articles in top journals in a wide variety of fields. He is currently PI on a large, multiyear grant on "The Evolution of Religion and Morality" and Director of the Cultural Evolution of Religion Research Consortium (CERC) and Database of Religious History (DRH).

Ryan Nichols is a Professor of Philosophy at Cal State Fullerton University in Orange County California and an affiliate of UBC's Centre for Human Evolution, Cognition & Culture. He specializes in cultural transmission, religion and Chinese thought. His current work falls within an absurd 20-year project trying to understand what made China Chinese. Text analytics contribute to that work.

Kristoffer Laigaard Nielbo is a humanist scholar with specialization in computational and quantitative methods for analysis, interpretation and storage of cultural data. He has participated in a range of collaborative and interdisciplinary research projects involving researchers from the humanities, social sciences, health science and natural sciences. His research covers two broad areas: automated text analysis and modeling of cultural behavior). Both areas explore the cultural information space in new and innovative ways by combining cultural data and humanities theories with statistics, computer algorithms and visualization.

Presentation: *The Distant Reading of Religious Texts: A "Big Data" Approach to Mind-Body Concepts in Early China*

This talk will focus on the debate about mind-body concepts in early China in order to demonstrate the usefulness of large-scale, automated textual analysis techniques for scholars of religion. In a series of studies we show how three different machine-based techniques—word collocation, hierarchical clustering and topic modeling analysis—provide convergent evidence that the authors of early Chinese texts viewed the mind-body relationship as unique or problematic. We conclude with reflections on the advantages of adding “distant reading” techniques to the methodological arsenal of scholars of religion, as a supplement and aid to traditional, close reading.

**Wittek, Stephen** <stephen.wittek@mcgill.ca>, **Gligorijevic, Kosta** <kosta.gligorijevic@mail.mcgill.ca> and **Nygren, Catherine** <catherine.nygren@mail.mcgill.ca>

Stephen Wittek is a scholar of early modern English drama and Manager of the Early Modern Conversions project at McGill University. His first book is entitled *The Media Players: Shakespeare, Middleton, Jonson, and the Idea of News* (University of Michigan Press, 2015). With the goal of establishing a more thorough understanding of the central, formative function of Shakespeare’s theatre in the news culture of early modern England, the project combines historical research with recent developments in public sphere theory, and argues that the unique discursive space created by commercial theatre helped to foster the conceptual framework that made a culture of news possible. In 2014, Dr. Wittek’s research was the subject of an hour-long episode of the CBC Radio One program *Ideas* (see <http://www.cbc.ca/ideas/episodes/2014/06/05/ideas-from-the-trenches-the-theatre-of-news/>).

Kosta Gligorijevic is a PhD student in philosophy at McGill University. His main research interest is ancient political philosophy, and especially the intersection between Aristotle’s politics, metaphysics and biology. His work in the digital humanities has included the transcription, modernization and digital analysis of Early Modern and Renaissance texts on gender, including the works of John Knox, Thomas Elyot and William Bercher. He is especially interested in using large-scale text analysis tools to study the reception and transmission of Classical political ideas in said authors.

Catherine Nygren is a PhD student in English at McGill University. Her current research examines paratexts of travel literature in the early modern and eighteenth-century. Other research includes topographies of literature and culture at the Grub Street Project and working with letters and other documents at the Burney Centre. She can be found on Twitter @broomgrass

Presentation: *Distant Reading Early Modernity (DREAM)*

This presentation will discuss the creation and practical application of DREAM, an online platform that facilitates distant reading for early modern texts. Studies that make use of distant reading techniques have tended to focus on texts from the 19<sup>th</sup> century or later, not for lack of scholarly interest, but because early texts present special challenges

that make computer analysis difficult. To help overcome these hurdles, DREAM enables users to create customized subsets from a corpus that includes all of the transcribed material available from Early English Books Online (EEBO). Unlike EEBO, DREAM facilitates batch downloading of custom subsets rather than obliging users to download texts on a one-by-one basis. In other words, it functions at the level of ‘sets of texts’ rather than ‘individual texts.’ Examples of subsets one might potentially generate include ‘all texts by Ben Jonson,’ ‘all texts published in 1623,’ or ‘all texts printed by John Wolfe.’ As it now stands, the corpus of transcribed EEBO texts comprises more than 1.2 billion words and includes some 44,000 titles (EEBO Phases I and II), or approximately one third of the 125,000 surviving texts from the period. To make distant reading of this corpus possible, DREAM provides a version of all texts with normalized spellings, and also gives users the option to transfer custom subsets directly to Voyant Tools, a digital environment for textual visualization and analysis. Please feel free to try the public beta version of DREAM, which comprises all texts in the public domain:  
<http://dream.voyant-tools.org/dream/?corpus=dream>

## BIOGRAPHIES FOR NON-PRESENTING PARTICIPANTS

**Badir, Patricia** <patribad@mail.ubc.ca>

Patricia Badir (University of British Columbia) has published on community identity and public space in Medieval and Reformation dramatic entertainments and on religious iconography and post-Medieval devotional writing. She is the author of *The Maudlin Impression: English Literary Images of Mary Magdalene* (Notre Dame UP, 2009). She is currently working on playmaking and the perils of mimesis on Shakespeare’s stage. She also studies modernism and Shakespeare in Canadian theatre and has recently published on this topic in *Shakespeare Quarterly*.

**Beaumont, Réa** <info@reabeaumont.com>

Dr. Réa Beaumont is described as “a tremendous pianist” (CBC Radio 2) with international performances and acclaimed recordings that have been broadcast in 13 countries, most recently on BBC Radio 3. In collaboration with leading composers, such as Order of Canada recipient R. Murray Schafer, she has premiered over 50 works, including those commissioned for her by the Canadian Broadcasting Corporation. Beaumont is a specialist in the music of Canada and has lectured at academic institutions internationally. Following the release of her book *Composer Barbara Pentland*, Beaumont was named “a world authority in the field” (CBC Radio 2). Her additional publications appear in *Clavier Companion*, *American Music Teacher*, *The WholeNote* and *Crescendo*. Beaumont is the recipient of awards from The Canada Council for the Arts, the Canadian Federation of University Women, Ontario Arts Council, British Columbia Arts Council and The Banff Centre for the Arts. She earned a Doctor of Musical Arts in Piano and Conducting as a Fellowship recipient, Master of Music, Bachelor of Music in Music Education (Hons), Artist Diploma and an ARCT Performance Diploma. Her coaches

included Anton Kuerti, Robert Silverman, Robert G. Rogers, Menahem Pressler, and Marc Durand. Beaumont was a Faculty member at the University of British Columbia and is now affiliated with the Royal Conservatory of Music in Toronto.

[www.reabeaumont.com](http://www.reabeaumont.com)

**Milner, Matthew** <m@matthewmilner.name>

Matthew Milner holds a PhD in History from the University of Warwick, and a Licentiate in Mediaeval Studies from the Pontifical Institute of Mediaeval Studies in Toronto. His historical research focuses on the nature of sense perception and religious life in fifteenth- and sixteenth-century England. Among his publications are *The Senses and the English Reformation*, the first full-length study of the history of the senses in the Reformation (Ashgate, 2011), 'The Senses in Religion', in *A Cultural History of the Senses: The Renaissance* (Bloomsbury, 2014), and 'The Physics of Holy Oats' (JMEMS, 2013). From 2014-16 he was the Assistant Director of the McGill Centre for Digital Humanities, following a postdoctoral fellowship with the Making Publics Project. As a digital historian, his interests centre on collaborative research environments, linked open data, chronometry, and historical social network modeling. His new historical event documentation tool, Nano-History ([www.nanohistory.org](http://www.nanohistory.org)) launched in the fall of 2015. It allows scholars to model and document large scale historical interactions and relationships. As part of the Conversions digital humanities team, he's used his open data experience to create new metadata for the EEBO corpus for the DREaM project.

**Rusk, Bruce** <bruce.rusk@ubc.ca>

Bruce Rusk is an Associate Professor in the Department of Asian Studies at UBC. His research focuses on the cultural and intellectual history of early modern China, ca. 1400-1800, especially the history of authentication and forgery, issues of authorship and authority, Confucian thought, and the digital humanities.

**Sinclair, Stéfan** <stefan.sinclair@mcgill.ca>

Stéfan Sinclair is an Associate Professor of Digital Humanities at McGill University. His primary area of research is in the design, development, usage and theorization of tools for the digital humanities, especially for text analysis and visualization. He has led or contributed significantly to projects such as Voyant Tools, the Text Analysis Portal for Research (TAPoR), the MONK Project, the Simulated Environment for Theatre, the Mandala Browser, and BonPatron. He has numerous publications related to research and teaching in the Digital Humanities, including *Hermeneutica, Computer-Assisted Interpretation in the Humanities* with Geoffrey Rockwell (MIT 2016) and *Visual Interface Design for Digital Cultural Heritage* with Stan Ruecker and Milena Radzikowska (Ashgate 2011).

**Vessey, Mark** <mark.vessey@ubc.ca>

Mark Vessey learnt close reading and old-fashioned literary historicism as a student of English at Cambridge in the late 1970s and has spent much of his time since then trying to make critical and literary-historical sense of very large authorial corpora (Jerome, Augustine, Erasmus). He is currently writing an extremely short history of the Latin literature of the later Roman Empire and preparing a student edition of Erasmus' neglected essay on literary interpretation and representation (with special reference to the Bible), the *Ratio seu compendium verae theologiae* (1518/19), itself an early essay in combining close and distant readings of texts. He is professor of English at UBC, sometime holder of a Canada Research Chair in Literature / Christianity and Culture, Principal of Green College at UBC and a team-member of the Early Modern Conversions project.

**Wang, Jessica** <jessica.wang@ubc.ca>

Jessica Wang is Associate Professor of U.S. History at the University of British Columbia and the author of *American Science in an Age of Anxiety: Scientists, Anticommunism, and the Cold War*, among other writings. She is currently completing a book manuscript on the social and medical history of rabies in nineteenth-century New York City, as well as beginning a new project on inter-imperial collusion and American empire in the early twentieth century.

**Yachnin, Paul** <paul.yachnin@mcgill.ca>

Paul Yachnin is Tomlinson Professor of Shakespeare Studies at McGill University and Director of the Early Modern Conversions project. He teaches and publishes on the social creativity of the arts, principally on Shakespeare and other writers of early modern Europe. His books include *Stage-Wrights, The Culture of Playgoing in Early Modern England* (with Tony Dawson), and *Making Publics in Early Modern Europe* (with Bronwen Wilson), as well as editions of *The Tempest* and *Richard II*. In addition to serving as Director of the Shakespeare and Performance Research Team and President of the Shakespeare Association of America (2009-2010), he founded the Making Publics (MaPs) Project (2005-10), which brought together scholars across the disciplines to rethink the history of early modern Europe by understanding how works of art and intellect created "publics," new forms of association based on the shared interests, tastes, and desires of individuals.