Joint CSDH/SCHN & ACH Digital Humanities Conference 2015

Ottawa, Canada | June 1-3, 2015

📄 **Hide Presentations**   📊 **Table View**

Session Overview

Date: Monday, 01/Jun/2015

8:30am - 10:00am **Lamoureux 122**	**1a: ACCUTE and CDSH/SCHN joint session** "Digital literary studies vs Digital humanities" with Dean Irvine, Jennifer Drouin, Marcello Vitali-Rosati and Michael E. Sinatra

8:30am - 10:00am
Lamoureux 215

1b: Named Entity Viz

Visualizing Philosopher and Topic Frequency Data Gathered from Named Entity Recognition Tools

Kevin Schenk, John Simpson, Geoffrey Rockwell
E-mail: kschenk@ualberta.ca

How can we study the evolution of ideas in philosophy? Jean-Baptiste Michel and colleagues (2010) have used n-grams to do what they call "culturnomics", but such approaches do not lend themselves to exploration. In this paper we discuss using a combination of Topic Modeling and Named Entity Recognition to study the flow of ideas within twentieth century philosophy as revealed within a full text corpus of 10 major philosophy journals. The tool we are developing lets users explore the history of ideas in philosophy by exploring networks comprised of both philosophers and philosophical topics.

Creating visualizations of large data sets to study changes to culture and ideas over time is not new. Most systems, like the Google Books NGram Viewer (https://books.google.com/ngrams), display keywords over time. By using the visual metaphor of a river's flow to vary the width and length of topics ThemeRiver visualizes variations in theme over time in a large corpus, showing that time is "an important type of relationship among documents that is important for some analytical tasks" (Havre 2002, p. 9). Our visualization uses a similar approach to show one of the variables over time, but overlays the others over it in a line graph to differentiate the two.

While important, the visual components of the visualization are inextricably connected with the underlying data and making sure that only relevant information is available to be visualized is a major contribution towards ensuring that the visualization is not ultimately a hairball waiting to be untangled (Regattieri et al. 2014). To select the necessary data from this multi-volume corpus we build on topic modelling work previously reported on at CSDH-SCHN (Brown et al. 2014) and augment it with named entity recognition (Barbosa et al. 2012; Yuval et al. 2012) that has been trained in conjunction with the ontology available from InPhO (https://inpho.cogs.indiana.edu/). In conjunction these methods create a pair of datasets that can be used individually or in tandem to reveal the flow of influential ideas across the corpus.

Currently being prototyped, the exploratory tool will allow the selection of topics from a dataset generated list sorted by frequency. Selecting an alternate date on the x-axis will rearrange the topics in the list by frequency. When selected, the topic frequency will be shown on the graph as an area chart. The frequencies of philosopher citations in texts are plotted as dots on the same graph. Selecting one of these dots draws a line connecting the frequencies of that philosopher for each date. Elements of the interface, such as the data shown by dots and area chart, can be switched to match the preference of the user.

Finding the frequencies and comparing them visually shows the rise and fall of ideas. Introducing Named Entity Recognition allows us to see what thinkers correlated with the topics of keywords, and how they change over the course of time.

Extracting and Visualizing Named Entities using Interactive Streamgraphs – A Case Study on First World War Data

Alaa Abi Haidar, Bin Yang, Jean-Gabriel Ganascia
E-mail: alaa.abi-haidar@lip6.fr

Huge amounts of printed manuscripts from old French news journals have been recently digitized and published by the National French Library, Bibliotheque Nationale de France (BnF). However, the massive amounts of produced textual data are highly unstructured and hard to index, search, or visualize, needless to mention the digitization errors resulting from ill-preserved or damaged manuscripts and imperfect Optical Character Recognition (OCR) techniques.

Named Entity Recognition (NER) is a task of information extraction that aims to identify in-text references to concepts such as people, locations and organizations, mainly in unstructured natural-language text. NER is very useful for text indexing, text summarization, question answering and several other tasks that enhance the experience between humans and literature. Furthermore, advanced NER and disambiguation techniques are capable of dealing with noise resulting from digitization errors. Information Visualization (InfoVis) is yet another indispensable task in Digital Humanities. Massive amounts of data is being produced at exponential rates that users cannot process. Information extraction and visualization both offer solutions to filter out useless data and organise or illustrate useful information to provide meaningful insights about the data.

Here, we use unsupervised named entity recognition and streamgraphs in order to extract useful concepts and visualize them from massive amounts of unstructured textual stream data, namely, French newspapers (e.g. Le Figaro, La presse, L'humanité) from the first world war period. Such a visualization allows us to identify main characters, events and locations involved in or relevant to the First World War, according to the French press. Furthermore, our visualization technique can help visually identify correlations between major people, places, organizations and events on multiple aligned streamgraphs. Our streamgraph visualization offers a much more intuitive overview of what happened, when and where during the first world war. Events can be easily correlated and their labels displayed by simply hovering over a stream with the mouse. Our method can be applied to unstructured data streams of any language, domain or time period as our method of unsupervised named entity recognition and disambiguation has already been tested on French and English benchmarks. An interactive demo of our streamgraphs from the Figaro titles between 1914 and 1918 is accessible at http://alahay.org/labs/ACASA/ .

What We Talk About When We Talk About Books: Topic Modelling Reader Responses on Social Reading Sites

Harvey Noel Quamen, Anouk Lang
E-mail: hquamen@ualberta.ca

Social reading sites like Goodreads and LibraryThing bring together readers from around the world to voice their opinions about books. Hosting huge communities of readers— LibraryThing claims 1.8 million users while Goodreads boasts 30 million members—these sites also provide abundant opportunities to observe the discourses of social reading in action.

This paper attempts to understand how these communities of readers evaluate and discuss books. Recognizing that a rating from one to five cannot adequately summarize a reader's overall evaluation, we chose instead to use topic modelling on prose reviews to see whether we could learn something about the discourses used by these readers. Scholars including Finn, Gutjahr and Nakamura have pointed to the scholarly value of working with reviews from social reading sites, but to our knowledge no work has yet been done on this material with topic modelling.

We built three corpora by collecting reviews from LibraryThing for three novels: Robin Sloan's *Mr. Penumbra's 24-Hour Bookstore* (2012; 264 reviews), Gillian Flynn's *Gone Girl* (2012; 757 reviews), and *The Girl with the Dragon Tattoo* by Stieg Larsson (English translation from Swedish 2008; 1,236 reviews). What, we asked, could a method such as topic modelling tell us about the ways people talk about books, especially in the informal, non-institutional environment of a social reading site to which they presumably come for pleasure rather than for mandated educational reasons?

To find out, we topic modelled each corpus with Mallet's LDA algorithm, using a range of settings from two to twelve topics in order to determine whether the reviews clustered into cohesive, semantically meaningful groups. We visualized the results on an HTML webpage that displays an interactive, force-directed graph.

What we have found has been surprising. The topic modelling algorithm seems to have some difficulty splitting the reviews into distinct clusters. In fact, the most definitive clustering occurs with as few as two or three topics, where reviews that summarize the plot (a mere 5% of reviews) are distinguished from those that do not (the remaining 95%). Repeated topic modelling with a greater number of topics (up to twelve) does little to split that 95% into smaller topic clusters. The greatest distribution of reviews across clusters seems to occur consistently in all three corpora at about eight to ten topics.

What do these results mean? In contrast to the diversity and variety of discursive tropes that a human reader might observe, the topic modelling algorithm suggests that the majority of these reviews have a high degree of topical homogeneity. However, the "sweet spot" witnessed consistently between eight and ten topics may indeed point to a repeated set of popular discursive modes used in such reviews. Other tools might help us understand more fully the dynamics we see. Nonetheless, despite the variance of keywords from novel to novel, the consistent behaviours of these topic models suggest the presence of meaningful patterns that begin to reveal what we talk about when we talk about books.

8:30am - 10:00am **Lamoureux 217**	**1c: Architecture and Mapping**

Visualizing Early Modern Lima—The Experiences of Lima as a Theater of Conversion

Jason Richard Young, **Francisco Laucirica**, Stéfan Sinclair, José Jouve-Martin
E-mail: jason.r.young@mail.mcgill.ca, francisco.laucirica@mail.mcgill.ca

Cities are spaces for public and private civic performances, and understanding the production of that space requires critically engaging with how actors narrate space and how interactions between these narrations enact, reiterate, challenge, and re-narrate space. This paper deepens humanists' understanding of space as a simultaneously constructed and experienced set of representations by combining traditional historical, literary, and cartographic scholarship with Geographic Information Systems (GIS) tools to analyze the social and urban development of Lima from a multiplicity of experiential viewpoints over a ~200 year period. Beginning with the 16th century, this paper represents the first study to offer a historically and geographically imbedded biopolitical mapping of Lima as a sacralized, gendered, racialized, and securitized experience of space.

The Representation Is Not the Thing: Digital Methods in Art and Architectural History

Alison Langmead
E-mail: adl40@pitt.edu

Art and architectural historians have long taken advantage of the latest in imaging technologies to perform their research and teaching. From the large-format glass lantern slides of the early twentieth century, to the small, plastic 35mm slides of mid-century, to the most recent digital imaging technologies, scholars of the visual arts have always required access to technologically-advanced representations of the objects they study. When an art or architectural historian projects an image behind them during a lecture, they do not do so to illustrate a point, they do so because they are representing for their audience the actual primary-source object of inquiry. Because of this established relationship, art and architectural historians have also been made continually aware of the ways in which these representations are markedly different from the objects themselves. The projected images stand in for works of art. They are not to be mistaken for the works of art.

This long-standing distinction is critical to an understanding of the place that digital methods have so far assumed in the fields of art and architectural history. Many projects in other fields that take digitized texts as their object are able to use the inherent text-processing capabilities of the computer to fascinating ends. The digitized image, however, simply cannot be mined in the precise same ways. While some scholars have begun to use the logical properties of the digital image object, such as encoded color information or even content-based image recognition, to investigate potential historical patterns, these approaches are novel, and still somewhat tangential, to the field. This is not only because they use innovative digital methods, but also because it has been somewhat unthinkable before the age of digital computing to sort representations of works of art by their color information in order to tease out historical relationships. And, indeed, digital approaches that do not focus on processing the digitized representations of primary-source art objects may have a more instinctive applicability to the study of art and architectural history today.

This paper discusses some of the lessons learned in the Visual Media Workshop at the University of Pittsburgh as it has transitioned from a traditional 35mm-slide library to a digital humanities lab. In particular, the project *Itinera* (http://itinera.pitt.edu), which takes as its focus culturally-motivated travel in the eighteenth and nineteenth centuries, has served as a testing ground for some of the roles that digital methods can take in the study of the visual arts above and beyond working with the digitized image. After running this project for almost two years, we have found it most fruitful to speak of *Itinera* as a "classroom." The project has proven to be successful as a way for all scholarly inquirers—whether faculty or students—to make use of the data processing powers of digital computing in order to learn from each other and work collaboratively to better contextualize, model, and imagine the complicated and reciprocal historical relationships between humans and works of art and architecture.

Moving the Mess to the Map: Studying locations and themes with messy data

Cameron Chown, **Randall Keefer**, **Marc Plamondon**
E-mail: crh.chown@gmail.com, rwkeefer473@community.nipissingu.ca, marcp@nipissingu.ca

We propose to discuss the importance and difficulties of dealing with "messy data" as part of undergraduate research projects in Digital Humanities.

Geothemography, a supervised undergraduate research project undertaken independently of a course, is a development of an earlier group project that was undertaken as part of an advanced undergraduate Digital Humanities course. In the earlier project, we examined ten related novels, half encoded using TEI XML and half with only minimal encoding, and developed digital tools to extract information from both sets. Two of the aims of the project are: to identify the degree to which pre-defined themes are present in the novels, and to identify and map automatically the various locations specified in the texts.

Geothemography takes these two concerns and joins them together. The current project uses over 150 British books (fiction and non-fiction) published between 1790 and 1900, automatically identifies U. K. geographic locations

specified in the books, and then attempts to match pre-defined themes to the locations using keyword proximity matching. None of the books have been encoded: they are plain text files. The simple keyword matching algorithms produce many distracting results, but our challenge has been to find ways to automatically reduce the amount of distracting, potentially false contributions to the results. We will present a few of our techniques for dealing with the messiness of the data.

Our tool and its results allow for an in-depth examination of the thematic significance of locations in Great Britain. The results are mapped using Google Maps to help visualize the significance of these theme-location pairs. For instance, using a general theme of nature, we find clusters of locations, around London, Bournemouth, etc., pointing to possible anxieties about the industrial revolution. We map places in Great Britain that localize British responses to the French Revolution. We also show changes in these associations on a decade-by-decade basis. We feel we reveal locales of thematic discussions by uniting automatically the thematic and geographic information.

Many text mapping projects, such as *Dislocating Ulysses* , *Hestia* , and *A Literary Atlas of Europe* , base their results on manually encoded geographical data and often focus on one text or small collection of texts: our project requires no manual encoding of texts and aims to incorporate as many texts as possible. We have experimented with adopting the "fuzzy shapes" that Piatti, et al., adopt ("Mapping Literature: Towards a Geography of Fiction") for the imprecision inherent in the results as a contrast to the precision usually implied by maps. In spite of the imprecision, we are confident that we (in response to an important distinction is raised by Moretti in *Graphs, Maps, Trees*) are not schematizing geometrical, but geographical relationships.

We offer our project as a model for undergraduate projects. The challenges of the messy data require students to think critically about the techniques involved, exploring problems and potential solutions. The scalability of the project, the lack of text preparation, and the simple text analysis techniques make this an accessible model for undergraduate Digital Humanities research.

8:30am - 10:00am **Lamoureux 219**	**1d: Innovative Integrations**

Developing Text-to-Speech for Religious Scripture: Shiri Guru Granth Sahib

Gurpreet Singh, Sandeep Kaur
E-mail: singhg@uleth.ca

This abstract outlines the development of schwa deletion algorithm for an efficient Text-to-Speech system for the Sikh religious scripture, Shri Guru Granth Sahib (SGGS), written in Gurmukhi Panjabi (Pa).

Punjabi is the only living Indo-European language which is fully tonal. It is also one of the few languages which is officially written in two different scripts, Gurmukhi and Shahmukhi.

These are both abugida (phonosyllabaray) scripts, in which the unit of notation is the syllable (other examples include Hebrew and Arabic). In such systems, the syllable is constructed around the consonant and the vowel is either implied or secondary.

Because Modern Punjabi is a relatively shallow language orthographically, it is quite easy to devise a Letter-to-Sound system: all you need to do is replace the letter in the script with the corresponding sound from the International Phonetic Association (G. Singh).

This changes, however, when it comes to the SGGS. This is a voluminous text of 1430 pages with 511,874 words, 1,720,345 characters, and 28,534 lines and contains hymns of 36 composers written in twenty-two languages in Gurmukhi script (Lal). This text presents several problems that are far less significant for Text-to-Sound systems designed for Modern Punjabi:

1) The use of vowels is different in that some of the vowel sounds at the end of words are used for grammatical purposes and are not part of the syllable

2) The texts are in multiple languages

3) There are words that use non-native phonemes.

Unless these problems are handled properly, there is no way to achieve an accurate system capable of converting the SGGS text to sound.

This paper reports on the state-of-the-art in terms of schwa deletion for Indian languages, in general, and for Gurmukhi Punjabi, one of the two Punjabi scripts, in particular. It discusses how we can improve the accuracy of the output in terms of perfect matching of IPA to the letters of each word in the DGGS, taking care of schwa deletion. Some attention is also paid to computational efficiency, though this is a lesser concern.

Works Cited

Lal, Harbans. "BOUNDLESS SCRIPTURE OF GURU GRANTH SAHIB." Web. 28 Nov. 2014. Omniglot. "Punjabi." 1998 - 2014.

Padda, S., R. Kaur, and Nidhi. "Punjabi Phonetic: Punjabi Text to IPA Conversion." International Journal of Emerging Technology and Advanced Engineering. 2012. Print.

Singh, Gurpreet. "Letter-to-Sound Rules for Gurmukhi Panjabi (Pa): First Step Towards Text-to-Speech for Gurmukhi." Language Resources and Evaluation Conference: LRE-Rel: Language Resources and Evaluation for Religious Texts. Istanbul, TURKEY, May 2012. Print.

Singh, Parminder. and G. S. Lehal. "A rule based schwa deletion algorithm for Punjabi TTS system." ICISIL. 2011. 98–103. Print.

Disability: The Last Frontier for DH

Kendra Rawluk, Carissa Alexander
E-mail: kendra.rawluk@uleth.ca, carissa.alexander@uleth.ca

The Digital Humanities (DH) continues to grow. Recent endeavors have sought to broaden participation to under-represented groups. The introduction of globalization and feminism, as discussed in the works of Fiormonte and Losh respectively, within the field aim to include groups that have been previously marginalized by language and gender. While this has resulted in substantial progress, DH has failed to recognize another marginalized group that is often easily overlooked: the disabled, in particular the developmentally disabled.

"Disability" includes a range of accommodations. We aim in this paper to focus specifically on the developmentally disabled. Disabilities, such as those that impede an individual's physical condition (eg: deaf) have been more closely examined through work by researchers such as Williams and his concept of universal design for individuals with disability. The developmentally disabled (eg: Autism) on the other hand seem to have been overlooked within DH.

When considering the disabled, it is important to recognize the spectrum that exists within the group. Disability cannot be narrowly categorized, but rather contains a wide variety of conditions that inhibit the individual's ability to adequately access information. While the individual's ability to access information may be restricted, their ability to comprehend may not. A disability does not equate to an individual's level of intelligence and in order for the expansion of the field, must not be considered as such. Disregard for a particular group within academia is unfortunate in consequence, as many of these individuals may have groundbreaking ideas that they cannot communicate as easily as an individual without a disability.

Inclusion is not a new concept within the scholarly world; integration is a necessity that stems from possibility. Organizations, such as GO::DH have been established for the incorporation of marginalized groups within the DH field. In our paper, we will draw on the comparisons between globalization, feminism and developmental disability to demonstrate that inclusion is a continuous process that provides new perspectives and innovative ideas within academia. We aim to demonstrate the possibilities that can arise from creating an inclusive and accessible environment for the developmentally disabled. If the field attempts to integrate this marginalized group, as it has

done in the past in terms of globalization and feminism, then it will be making effort towards increasing communication more broadly. Broader communication will allow for the integration of new ideas and the elaboration of existing. It will open up the DH field to perspectives previously untapped and can result in the betterment of the field as a whole.

Works Cited

Fiormonte, Domenico. "Towards a Cultural Critique of the Digital Humanities."

Controversies around the Digital Humanities, Vol. 37, No. 141. Historical Social Research, 2012. 59-76. Web. Nov 14, 2014.

< http://www.academia.edu/1932310/Towards_a_Cultural_Critique_of_Digit al_Humanities>

Global Outlook :: Digital Humanities. Web. Nov 14, 2014.

<http://www.globaloutlookdh.org/>

Losh, Elizabeth. "Feminist Making and Breaking in the Digital Humanities" Web..Nov. 28, 2014

<http://www.academia.edu/5066265/Feminist_Making_and_Breaking_in_the_Digital_Humanities>

Williams, George. "Disability, Universal Design, and the Digital Humanities." Debates in Digital Humanities. 202-12. Web. Nov 14, 2014. <http://books.google.ca/bookshl=en&lr&id=_6mo2tApzQQC&oi=fnd&pg=PA202& dq=disabilities+and+digital+humanities&ots=9XuhhvK8UF&sig=4bLl8LcYVH9YGRICMgOEfQU24z4#v=onepage& q=disabilities%20and%20digital%20humanities&f=false>

"Q i-jtb the Raven": Taking Dirty OCR Seriously

Ryan Cordell
E-mail: r.cordell@neu.edu

Discussions of large-scale digital archives inevitably return to the question of optical character recognition (OCR)—and, more specifically, to its failings. Whether the archive under discussion is Early English Books Online (EBBO) or the Chronicling America newspapers, in presentation after presentation scholars groan at screenshots of OCRed text that mangles the words of its proof documents, replacing the clearly-printed characters of paper books with jumbles of symbols—in ugly computer fonts, no less—that in many cases cannot be read as semantic content. For editors and creators of digital content, such discussions are, perhaps, evergreen, as scholars continue to suss out best practices for doing what Jerome McGann has long argued we must: re-curating and re-editing the "the whole of our cultural inheritance…in digital forms" (New Republic of Letters 1). We are torn between competing desires to digitize as much as possible and to ensure that we digitize well. Perhaps due to the digital humanities' roots in critical editing, however, this dilemma leans always toward the latter. The intricate detail-work of TEI encoding is cast as the preferable option—as it no doubt would be, were time and money no object—while the unedited OCR of large-scale projects is labeled "dirty OCR," a phrase that invokes the jumbledness of such texts to the human reader but implies too an uncleanliness in such archives.

This paper will argue that such conversations signal a foreshortening of the bibliographic imagination, in which the digital archive can be only a transparent window into the "actual" objects of study, which are presumed to exist—entirely—in material forms. As scholars such as Matthew Kirschenbaum, Sarah Werner, and Whitney Anne Trettien have shown, however, the digitized book, magazine, or newspaper is not simply a surrogate for the material object, but instead is a new edition of the text that opens some interpretive possibilities while foreclosing others. Dirty OCR is not simply an interminable series of errors to be redressed. Instead, these unedited archives necessitate careful bibliographic and book-historical attention that both leverages the powers of our current digital archives while historicizing their creation and use. Just as cheap, pirated (and often error-filled) American editions of nineteenth-century British novels now teach modern scholars much about the economics, print technology and literary culture in that period, dirty OCR illuminates the priorities, infrastructure, and economics of the academy in the late twentieth and early twenty-first centuries. Moreover, I will argue that a broad, if subtle, notion of "not good enough" is hindering scholarly thinking about what these new editions might reveal, both through projects that more fully embrace scale as a way of compensating for local transcription errors and even, perhaps, through creative engagement with OCR "glitches" as potential sites of critical attention or intervention.

10:00am - 10:30am **Entrance Hall to LMX122**	**Coffee Break 1: Coffee Break**
10:30am - 12:00pm	**ACCUTE/CSDH: Joint Lecture: Wendy Hui Kyong Chun, "New Media: Wonderfully Creepy"**
12:00pm - 1:00pm	**Lunch Break 1: Lunch Break**
1:00pm - 4:00pm	**LAC/CSDH: Joint Event LAC**
4:00pm - 4:30pm **Entrance Hall to LMX122**	**Coffee Break 2: Coffee Break**
4:30pm - 6:00pm **Lamoureux 122**	**2a: Panel: A Big Data Challenge in the Humanities**

A Big Data Challenge in the Humanities: Twenty-eight hundred years of the European Textual Record

Gregory Ralph Crane, Bruce Robertston, David A. Smith, Marie-Claire Beaulieu, Bridget Almas
E-mail: gregory.crane@tufts.edu, bruce.g.robertson@gmail.com, dasmiq@gmail.com, Marie-Claire.Beaulieu@tufts.edu, balmas@gmail.com

A Big Data Challenge in the Humanities: Twenty-eight hundred years of the European Textual Record

Panel Proposal for the Joint CSDH/SCHN & ACH DH Conference 2015

http://csdh-schn.org/2014/10/20/cfp15/

Overall Panel Abstract: Digital methods have changed the potential space of research in two complementary directions. On the one hand, millions of scanned documents, produced over thousands of years and in hundreds of languages are now available under various open licenses for download, analysis, modification and republication. At the same time, we have methods by which to represent our understanding of particular linguistic structures, social networks, textual interrelations and other phenomena in newly precise, machine-actionable formats. We can now trace intellectual developments across not only time and space but also language and culture at greater depth and breadth alike, but a relative handful of advanced researchers and library professionals, even with the best automated methods, cannot optimally exploit this new environment. We need to draw upon expertise from beyond traditional Western centers of higher education (e.g., linguistic competence in Classical Arabic and Chinese from Cairo and Beijing) and upon contributions from citizen scientists around the world. Such a new culture of intellectual production poses significant social as well as intellectual challenges but addressing these challenges opens the door not only for improved research but also for a dialogue among civilizations that is itself of fundamental importance. This panel documents work already completed in advancing this larger task. The presentations take Greek and Latin as a starting point, but, in addressing the full range of textual data in, and also about, Greek and Latin, we need to

address billions of words produced over thousands of years and affecting every aspect of modern life.

[Many of those listed as contributors contributed to work in more than one of the papers, but. with the exception of the final paper, we have tried to list authors only in a single paper.]

Historical OCR -- Optimizing OCR for Greek, Latin, Fraktur, and Early Printed Books

Bruce Robertson (presenter), Frederik Baumgardt, Federico Boschetti (CNR/Pisa), Matt Munson, Uwe Springmann, Nick White (Durham)

Image front searching allows users to search textual data automatically generated by Optical Character Recognition (OCR) and then to view the original scanned page images. This technique became practical in the 1990s, when scanning technology and storage costs became low enough that collections of scanned documents could be assembled. Fundamental services such as JSTOR.org and Google Books are based upon this technology. Millions of books from the historical record are now available in scanned form and under various open access licenses in Europe and North America. Copyright laws on both sides of the Atlantic complicate global access, but millions of scanned books in the public domain are already available for download.

The capabilities of commercial OCR software have, however, constituted a limiting factor. Commercial OCR works very well for documents in contemporary, economically important languages but not for many textual materials of critical cultural importance. No large scale OCR solution has existed for Classical Greek. OCR for German Fraktur has also required a specialized and restrictive commercial license. Standard OCR on early modern printed books in major languages such as English and French produce very poor results. Even OCR of Latin can be improved when scholar developers have access to the full OCR system.

This paper reports on (1) different efforts to optimize three open source OCR systems, Gamera, OCRopus and Tesseract, one or more of which has been used to address challenging printed materials, (2) the development of automatic error detection and decentralized correction, (3) the value of running a single OCR engine on multiple scans of the same book to improve overall accuracy, and (4) practical challenges in working with different fonts, especially as we work with older, less standardized printed materials.

The paper covers results from the Lace OCR Project at Mount Allison University, the Ancient Greek OCR Project at Durham and Tufts Universities, the Open Greek and Latin pipeline at Leipzig, and OCR of early modern books at Ludwig Maximilian University in Munich.

Open Greek and Latin -- Representing texts in many versions and over thousands of years

Alison Babeu (Tufts), Monica Berti (Leipzig), Lisa Cerrato (Tufts), Stella Dee (Leipzig/Tufts), Greta Franzini (Goettingen), Neven Jovanovic (Zagreb), Anna Krohn (Tufts), Simona Stoyanova (Leipzig)

Students of Greek, Latin and other languages have had access to digital corpora for decades but these initial corpora often suffer from structural weaknesses: (1) they are available only under proprietary licenses and their providers depend upon monopoly access and legal sanctions; (2) they include only transcriptions without the original scans upon which the transcriptions depend; (3) they include only the reconstructed texts and not the textual notes; (4) they include -- and their design only allows them to include -- a single edition for any given work and lack the infrastructure to represent textual history in depth; (5) they are docu-islands (or at best docu-continents) closed in scope (e.g., all Greek literature or all of the Patrologia Latina) whereas our large digital collections cover much of the surviving textual record -- for Greek and Latin, that covers thousands of years and billions of words.

This paper builds upon the fundamental advances in scholarly OCR described in the first paper and covers work at Leipzig University, Germany, Tufts University in the United States, and the University of Zagreb in Croatia to create an open, scalable collection of Greek and Latin textual data that can include thousands of versions of any given work, as well as quotations, citations, paraphrases and other instances of textual re-use. These efforts include: (1) the Perseus Catalog of Greek and Latin uses the Functional Requirements for Bibliographic Records to track multiple versions of individual works, with at least one version for every major Greek and Latin source text surviving from antiquity; (2) the Leipzig Open Fragmentary Texts series, which has begun organizing authors whose works are lost and whom we know only through hypertextual meta-editions that organize quotations, paraphrases and citations in works that do survive; (3) the Open Philology CTS/CITE repository, that stores full texts of primary sources in, translations from, and reference materials about Greek and Latin; (4) the challenges of creating a large but relatively well-defined corpus (e.g., print editions of Greek and Latin works that survive through c. 600 CE) vs. an open ended corpus that cannot be well defined and for which only samples can be organized (e.g., all Classical Greek and especially Latin, in all its forms, produced after c. 600, much of which is not in a form from which OCR can generate transcriptions).

Scalable Services -- Analyzing Millions of Scanned Books

David A. Smith (Northeastern)

Projects such as Open Greek and Latin produce manually curated textual data but manual methods do not scale when we begin working with millions of digitized documents. This paper shows how automated methods can both build upon, and contribute to, the curated data described in the previous paper. Developed over five years, the UMASS.edu Proteus Project used OCR output for 3.6 million digitized books from Archive.org as a test bed. The project developed results for a number of services:

Language Identification: The project identified the language of 3,628,227 OCRed books at the Internet Archive (results are available on the project Language Identification page).

From those 3.6M books, they identified where the language identification differed from the books metadata at the Internet Archive. They found 31,947 books that they believe to have the incorrect language. In addition, there were 17,272 books where the Internet Archive identified them as a language that they did not check for: results are available on their Language Identification Differences page. Books that have the incorrect language at the Internet Archive may benefit from being re-OCRed with the correct language to improve the resulting text.

Duplicate Detection: Canonical text for English and Latin works have been acquired from the Perseus Digital Library. There are 803 English works and 401 Latin works.

These works were compared with the text of the English or Latin OCRed books from the Internet Archive to find full and partial duplicates[1]. Partial duplicates are books where there may be extra material such as footnotes or other works - for example Hamlet within The Tragedies of Shakespeare. Duplicates are quickly identified by looking at only the unique words within a book.

Duplicate Alignment: Once duplicates are identified, they were aligned to identify corresponding portions of the works[2].

Entity Extraction: Named Entity Recognition (NER) was run on 1,072,356 books from the IA book collection. This identified several entity types: People, Places, Organizations, and Things.

Quotation Detection of Canonical Works: While duplicate detection allows us to find matches of complete works, finding matching quotations is more fine grained. By searching for a quotation, for example Rosencrantz's "Take you me for a sponge" (Hamlet, Act IV Scene II), we can find all occurrences of that quotation even in books that are not copies of Hamlet. This allows us to see which passages attract the most scholarly interest over time.

Citizen Science, Education, and Digital Philology -- Transforming the culture of scholarly production

Bridget Almas, Marie-Claire Beaulieu (Tufts), Giuseppe Celano (Leipzig), Thibault Clérice (Leipzig), Harry Diakoff (Alpheios.net), Maryam Foradi (Leipzig), Gernot Höflechner, Thomas Köntges (Leipzig), Uta Kremer (Leipzig), Christine Rougham (Leipzig)

The results of the first three papers now present several challenges. First, Optimized OCR for historical sources (classical Greek, Fraktur, early printed books etc) and the analytical services from large scale systems such as Proteus create data that is useful but that also contains errors and is incomplete. Second, the scale of data now available is so vast that no single project can examine and curate it all. Even if we restrict ourselves to Latin or Greek, we are working with materials produced over more than two thousand years and in a dizzying array of genres and cultural contexts. Third, neither traditional publications expressed in expository prose nor emerging, machine-actionable micro-publications expressed in forms such as RDF triples are by themselves adequate. We need

publications that include both: the machine actionable micro-publications can represent many of our conclusions (e.g., the Alexandria mentioned in this ancient text designates the Alexandria in modern Afghanistan not the famous one in Egypt) but we always need to be able to include the reasoning behind our conclusions. We thus need an environment that can scalably support hybrid contributions from many different contributors.

This paper describes cumulative work done by Alpheios.net, Papyrus.info, the Pelagios Project, the Perseids platform, and the Arethusa annotation framework, each of which contributes to creating an open, decentralized, scalable ecosystem for the scholarly contributions by student scholars and citizen scholars well as by advanced researchers and library professionals. This emerging ecosystem allows students and scholars to collaboratively create vetted open source digital editions by transcribing, editing, and translating texts, manuscripts, and inscriptions.

This paper presents results in addressing a number of use cases for this ecosystem: (1) classroom collaboration on digital editions; (2) scholarly curation of texts from scan through publication of TEI XML (3) developing dynamic syllabi using managed resources. These use cases aim to serve the needs of a broad audience including formal classes, ad-hoc research teams, and individuals working in Classics as well as in other fields of the Humanities.

Infrastructures 2015: Getting beyond individual projects to a sustainable intellectual Ecosystem

Bridget Almas (Tufts), Alison Babeu (Tufts), Frederik Baumgardt (Leipzig), Lisa Cerrato (Tufts), Sayeed Choudhury (JHU), Gregory Crane (Leipzig/Tufts), Tim DiLauro (JHU), Mark Patton (JHU), Maxim Romanov (Tufts)

We cannot have a sustainable Digital Humanities unless we think also about the Digital Libraries that must take over and sustain all of work of persistent value -- as John Maynard Keynes tellingly pointed out, in the long run, of course, we are all dead. The Digital Humanities may have emerged as a major topic, certainly a buzz word, arguably an intellectual bubble, but a search in Google News retrieves almost eight times as many articles for "Digital Libraries" as it does for "Digital Humanities" (501,000 vs. 64,100 on December 7, 2014) Those of us who work as Digital Humanists (or, more generally, as Humanists finding their way in a Digital Age) still think in terms of individual projects associated with particular collections and services. The ecosystem of which we are a part is not yet sustainable because the services and collections that we are developing are too closely tied to particular champions or groups. At the same time, researchers in all fields rarely appreciate the difference between a collection or a piece of software that serves a particular purpose from collections and software that third parties who have never met the developer can take over and sustain over time. Eager developers still all too often think little about what will happen when the current grant or their current enthusiasm runs out.

Funding agencies are more far sighted and they do worry -- they have, indeed, learned to worry -- about what happens in the long run to the projects that they support. A great deal of work has gone into infrastructure projects over the past five years, both in North America and especially in Europe, where support from Germany and the European Union stands out. At the same time, it can be difficult to separate the well-meaning promises that appear in the proposal and initial website from working services upon which third parties can build. There is inevitably a great deal more discussion about infrastructure than there are working, sustainable services that have evolved beyond the individual project stage.

This final paper compares the use cases and needs that have emerged in analyzing projects such as (but not limited to) the work described in the opening sessions of the panel. There are at least thirteen projects in the general area of Digital Classics that are sufficiently mature to shape concrete work with existing, and plans for new, digital infrastructure: they are at least five years old (and thus have survived past a second round of funding), they have reached the point where they know that they should not be managing their own proprietary systems (and thus are genuinely committed to open data and open source), and they have developed tangible user communities who depend upon their work (and thus have not simply created resources to advance a particular project or research question). The Digital Classics requirements may not cover the full range of Humanities needs but they are all essential to any infrastructure for the Humanities in a Digital Age.

This paper will review solutions to these requirements that are actually available as of summer 2015 from big European efforts such as Europeana, Clarin, and Dariah, as well as within the more fragmented North American landscape. The goal is to identify concrete foundations upon which we can immediately begin to build.

4:30pm - 6:00pm **Lamoureux 215**	**2b: Social Studies**

Decolonizing Digital Humanities in Africa
Titilola Babalola Aiyegbusi
E-mail: titilola.babalola@uleth.ca
see file attached

Influence in the Digital Humanities: A Social Network Analysis
Kim Martin, **Anabel Quan-Haase**
E-mail: kmart5@uwo.ca, aquan@uwo.ca

Influence, in the academic context, is difficult to define; to some it is a direct reflection of one's citation counts, while to others it is a much more far reaching social process, consisting of mentoring, discussion, and conversation. This paper summarizes the findings of a preliminary study into the concept of influence within the DH community.

DH @ Work and The Social Studies of Digital Humanities
Tanya E. Clement, **Daniel Carter**
E-mail: tclement@ischool.utexas.edu, carter.daniel.w@gmail.com

In 2009, Christine Borgman asked "Where are the social studies of digital humanities?" More specifically, she inquired, "Why is no one following digital humanities scholars around to understand their practices, in the way that scientists have been studied for the last several decades?" Arguing that such research has significantly shaped scholarly infrastructure for the sciences as "a central component of cyberinfrastructure and eScience initiatives," Borgman urged DH to learn more about its own practices.

For this study, we are "following around digital humanities scholars" from a range of backgrounds who function in a variety of roles within higher education, academic institutions. We have chosen "humanists" who work with the "digital" primarily because we are interested in the perspective of humanities scholars who find themselves working at the intersection of humanistic principles and the development of scholarly information infrastructures (defined by Borgman as "the technology, services, practices, and policy that support research in all disciplines"). While we understand that digital humanists do all kinds of work including administration, teaching, service, and writing (to name a few), we have chosen in this study to focus specifically on tasks and perspectives that are revealed in the process of project-based work. Digital humanities projects are often touted as the site of work that not only defines DH (Drucker 2009; Svensson 2010; Ramsay , Manifesto 2009) but also the site of work in which the development of information infrastructure has the most potential to be impacted by theoretical perspectives imbued in humanist critique and vice versa. Better articulating what kind of work infrastructure development entails is particularly illustrative for considering information work in digital humanities because it is a "research area where the interests of humanists, technology researchers, and others converge" (Friedlander).

Specifically, this paper will discuss our findings considering two essential topics: (1) the nature of the "information" work that digital humanists do; and (2) how we go about observing and studying such work to better understand ourselves. We have a mixed methods approach including topic modeling five years of "Day of DH" data as a glance into how DHers in general describe their work, interviewing approximately 20 digital humanists on their daily DH practices and the values they attribute to these practices as well as observations of "digital humanists at work."

Johanna Drucker reminds us that "humanistic theory provides ways of thinking differently, otherwise, specific to the

problems and precepts of interpretative knowing—partial, situated, enunciative, subjective, and performative" and that DH is defined by its "emphasis on making, connecting, interpreting, and collaborating" (Drucker 2012). She notes that "[o]ur challenge is to take up these theoretical principles and engage them in the production of methods, ways of doing our work" (Drucker 2012). This study will provide a snapshot of these ways and consider how and if these theories and methods are reflected in the quotidian practices of the digital humanist? What is the nature of studying DH @ work?

4:30pm - 6:00pm **Lamoureux 217**	**2c: Games and Gaming**

Digital Games as a Form of Cross-Cultural Knowledge Translation

Stephen Mathew Wilcox
E-mail: swilcox@uwaterloo.ca

Knowledge integration and translation—the practice of making scholarship available and intelligible to various audiences—is an increasingly-important component of modern academic research. What I explore in this presentation is the potential for digital games to function as a form of knowledge translation.

I begin with a demonstration of *Allergies & Allegories* , a digital game I am developing in collaboration with the University of Waterloo's Games Institute—a SSHRC Partnership initiative for the advancement of knowledge in game-related interactions and technologies—and GET-FACTS (Genetics, Environment and Therapies: Food Allergy Clinical Tolerance Studies), a Canadian Institutes of Health Research-funded knowledge mobilization initiative that is intended to raise awareness of food allergies in Canada. The game is based on patient-centered research on the social and cultural barriers faced by children with food allergies (Fenton et. al.). By incorporating this research into the procedures of the game, players become personally familiar with the social and cultural challenges food allergies present, thereby moving scholarship on food allergies into the public domain through an accessible, persuasive format.

I go on to make the case that games like *Allergies & Allegories* can be understood as a form of cross-cultural knowledge translation. This understanding of games is based on the rhetorical theory of invention—a form of knowledge translation rooted in culturally-defined topics or commonplaces. The commonplaces can be thought of as common relationships between ideas that are intrinsic to a particular culture or sub-culture. It is from this perspective that some scholars have conceived of art itself as a form of invention. Giambattista Vico, for instance, argued that art trains the senses in the commonplaces of a particular culture, facilitating the audience's capacity to locate knowledge embedded therein (*The New Science*).

Treating games as a form of *cross-cultural knowledge translation* extends this inventive understanding of art. It suggests that the procedures of a game persuade the player to discover common relationships through playful interaction, meaning that during play games become *interactive commonplaces* . In *Allergies & Allegories* , for instance, players take on the role of a food-allergic child as they build friendships, strengthen support networks, and establish safe and supportive spaces; in doing so players recreate the commonplaces of food-allergic children, providing players with a context in which to situate knowledge of food allergies and their risks. This same principle can be applied to other bodies of research as well, and it is especially suited to translating knowledges between distinctly situated persons.

Works Cited

Fenton, Nancy Elizabeth, et al. "Illustrating Risk: Anaphylaxis Through the Eyes of the Food-Allergic Child." Risk Analysis 31 (2011): 171-83. Print.

Vico, Giambattista. *The New Science* . Ithaca, N.Y.: Cornell University Press, 1968. Print.

Playing Queer: Locative Media, City Space, and Game Mechanics

Maureen Engel
E-mail: mengel@ualberta.ca

This paper theorizes the possibility of building not just a queer gaming experience, but a queer game mechanic – that is, a game whose very structure of play can be interpreted as queer.

Attention in Game Studies scholarship to analyses of race, gender, class, or sexual orientation most frequently focus on:

1) the content of games and their characters

2) game developers and industry influencers

3) players and how they're treated by "dominant" players

This is important and necessary work as we develop our critical apparatuses to analyze the roles of gender, race, sexuality, class, etc. in the gaming environment. But there is one central component of games that is too often left unexamined in these analyses: the game mechanic itself. I am interested in whether it is possible to build a game that capitalizes on the cultural logic of queer affiliation in order to make the game play itself queer.

Building on the foundational work of queer theorists and historiographers who have demonstrated how queer subjects have re-appropriated a range of urban, public spaces in order to build community (c.f. Warner and Berlant, "Sex in Public" and Chisholm, Queer Constellations among many others), my proposition is that the particularly queer affordances of the city can be deployed to achieve a queer gaming mechanic. If, as Michael Sibalis notes, "urbanization is a precondition for the emergence of significant gay culture," then locating game play in the city itself can provide particularly queer affordances. Jason Farman didn't have queer history in mind when he called locative media a "creative misuse of space," but it is a compelling echo nonetheless.

"Queer Edmonton" is a locative media experience that misuses the space of the city by embedding its queer history in its contemporary everyday. Based in the queer history of a particular city, it is a locative scavenger hunt where players must literally travel the spaces of the city in the hopes that they will encounter queer history -- now disappeared, redeveloped, forgotten. Each new discovery leads the player further, hails her, teaches him about how to look, queerly, for clues. As such, game play relies on the player thinking, and playing, queerly: Where are the queer areas of the city? Of any city? Why?

Algorithmic Generation of Global Racial, Cultural, Religious, Architectural and Linguistic Variation

Mark Richard Johnson
E-mail: mrj503@york.ac.uk

This paper explores the game currently under development by the author – *Ultima Ratio Regum* – and the game's algorithmic generation of fictional world cultures. Set approximately within the era of the Scientific Revolution, the software in question is able to generate millions of potential civilizations in exhaustive details. This is done via extensive "procedural generation" – the writing and creation of game content algorithmically rather than by a specific human designer. The details the game generates include variation in skin-tone and genetic markers (eye colour, hair colour, etc); variation in cultural preferences in a wide number of spheres including war, birth, death and burial, leadership, law and justice, foreign relations, literature, myth and legend, clothing styles, identity markers (tattoos, piercings, hair styles, etc) and many others; variation in religions including the names and gods of those religions, their religious practices, festivals, holy books (if any), forms of worship, relationships with other religions, and religious rules and edicts; consistent algorithmically-generated architectural styles of all buildings, ranging from ordinary houses to cathedrals and palaces; and linguistic variation in sentence structure, alphabetic/syllabic /logographic languages, and common sayings, phrases and expressions. Many of the racial and cultural algorithms are specifically designed to undermine the Western "othering" of cultures which adopt practices of body modification or cultural expression viewed as extreme behaviours among Western audiences and propose the possibility for such

practices being the "norm" across in-game society. Others are designed to generate civilizations of sorts never seen in the real-world but that remain within the realm of realistic possibility.

This paper consists of two interrelated parts. The first part of this paper explores the forms of data structure and algorithm that generate these civilizations, and ensure consistency within them in everything ranging from the architectural style of cathedrals to the forms of worship citizens practice within their homes. It will explore a range of the different "generators" within the game, and how the game has been programmed to identify civilizations that are more "interesting" than others and to focus upon generating them. The paper covers how this metric of "interest" was written into the game code and how the game is designed around ensuring maximum variation between civilizations, without ever turning civilizations into unrealistic "extremes" at the edges of these algorithms.

The second part of this paper explores the forms of gameplay that are designed to emerge from these detailed worlds. This will show the player is able to interact with these generated cultures in the playing of the game, and how these cultures have been "foregrounded" as important parts of the in-game experience rather than merely nonessential background information to the player to peruse if interested. It will lastly explore some of the unusual gameplay mechanics this is intended to facilitate, such as "researching" the game world in order to gain information not explicitly presented to the player, and how this type of gameplay aims to innovate away from contemporary game design wisdom emphasizing clarity and explicit goals.

| 4:30pm - 6:00pm | **2d: Édition** |
| Lamoureux 219 | |

Indexicality, Visual Poetics, and the PetrArchive: A Scholarly Digital Edition of Petrarch's Songbook

John A. Walsh, H. Wayne Storey
E-mail: jawalsh@indiana.edu

The PetrArchive (http://petrarchive.org) is a new digital archive and "rich text" edition of Francesco Petrarca's iconic fourteenth-century songbook *Rerum vulgarium fragmenta* (*Rvf* ; *Canzoniere*). A primary goal of the PetrArchive is to document, investigate and illustrate the graphic codes and structures—especially the "visual poetics"—of the work. Our paper will discuss and demonstrate specifically the broad issue of indexicality in the context of the digital editing and encoding practices and strategies adopted and exploited in pursuit of this goal.

The *Rvf* is both in its manuscript tradition and our new edition a highly indexed and indexable book. An index often contains a list of words, subjects, titles and addresses, as well as pointers and locations of references. These lists and addresses provide a representation, map, or model of a document. A comprehensive, hierarchical, multifaceted index to, for instance, a large edition of letters is of tremendous practical value as a guide through the collection. An index may also be a remarkable work in itself as a structured conceptual model of the contents of a collection. Often indexical structures are embedded in the document as we find in the Bible and other religious texts, with book titles, chapter and verse numbers, and cross-references embedded throughout the text. Petrarch's adherence in his model holograph MS Vatican Latino 3195 to his 31-line graphic canvas and his designs of various combinations of verse forms to fill that canvas generate, among other things, a visual index to the document, with the textual and graphic shapes of the manuscript serving as a visual map of genre and generic juxtaposition. Our project will build a graphic representation, or visualization, of the manuscript that will allow readers to browse and scan—by shape and structure—the distribution, combination, and juxtaposition of genre and form throughout the manuscript.

Another aspect of our visual and schematic indices to the edition will be the animation of Petrarch's own poetics of erasure and transcription, through which he revises his texts but also deforms the patterns of his own indexical practices to highlight the importance of the work's visual-poetic structuring. We will demonstrate an example of this deformation in our animation of the canzone Quel' antiquo mio dolce empio signore (*Rvf* 360). In his own holograph MS, by then a service copy, Petrarch is forced to abandon his ideal layout for the prosodic form of the canzone. Only in subsequent MSS will the canzone revert to its ideal, authorial form not in the author's hand. Our representation will allow readers to view the poem morphing from one layout to the other, requiring the encoding of both the actual and ideal layout in the document and the interpretation of those codes in the digital design and publishing layers of the edition. Beyond their instant utility in allowing users an overview of the design of individual MS pages and of the work's complex system of combining forms, these indices reconfigure the equally complex layers of indexical structures inherent in a scholarly edition.

Lettres et numérique : passerelles et précipices…

Cécile Meynard, **Thomas Lebarbé**
E-mail: cecile.meynard@gmail.com, thomaslebarbe@me.com

La profusion de sources modernes, générée par une volonté légitime de mettre à disposition du public les fonds patrimoniaux, représente une réelle opportunité scientifique et culturelle. Mais de telles initiatives peuvent manquer leur objectif (toucher, intéresser et cultiver un public, spécialisé ou non) du fait du peu de documentation dont ces fonds numérisés sont parfois accompagnés.

L'enjeu pour le chercheur est donc de partir de ces fonds numérisés pour les transcrire, les enrichir et les valoriser. La problématique reste de produire des ressources numériques riches (sans que leur mise à disposition ne souffre de la technicité induite par certains outils), et de structurer de la connaissance qui soit partageable et moissonnable (sans s'encombrer de normes et standards dont l'appropriation n'est pas toujours évidente).

Les auteurs sont les porteurs du projet Manuscrits de Stendhal (www.manuscrits-de-stendhal.org). Reprenant un travail de longue haleine initié par Gérald Rannaud, ils ont proposé un regard nouveau aussi bien sur les données que sur les formes à leur donner (édition papier, édition numérique, édition de corpus). Le modèle numérique et méthodologique qu'ils ont proposé à trouvé écho dans des projets qu'ils ont coordonnés ou auxquels ils ont été associés : ces projets diffèrent certes par leurs objets (manuscrits, inventaires, formes éditées), les points de départ de leur construction scientifique (sources manuscrites, éditions premières) et leur finalité scientifique (édition numérique, édition papier, édition complémentaire, approche pluridisciplinaire du contenu, interprétations de la notion de philologie). Ils portent sur des auteurs de notoriété plus ou moins grande et d'époques diverses. Toutefois, le point fondamental d'intersection entre de telles initiatives est la volonté d'éditorialisation et la mise en place d'une méthodologie, pensée à partir de la prise en compte des différents acteurs de la chaîne éditoriale, depuis les transcripteurs jusqu'aux « lecteurs », invités à se faire eux-mêmes acteurs dans ce processus.

Par ailleurs, dans le monde 2.0 permis par internet, la question se pose des approches collaboratives de la constitution et de l'enrichissement des ressources patrimoniales. Les auteurs se penchent précisément sur cette question à partir du projet centré sur le fonds Benoîte Groult. Outre la valeur culturelle, sociologique et scientifique de ce fonds, lui sont associés une communauté de pensée et un mouvement de revendication (le féminisme) et a fortiori des forces vives de la société civile prêtes à s'investir pour contribuer au projet d'éditorialisation des manuscrits de cette écrivaine. Ceci néanmoins soulève d'autres problèmes, tels les faiblesses et imperfections de transcriptions et enrichissements produits par des non spécialistes.

Travailler à ces différents objets de recherche dans le cadre des humanités numériques des textes oblige ainsi à redéfinir ces humanités à l'aune du numérique : il s'agit de redéfinir les fonctions et les métiers de l'auteur, de l'éditeur scientifique, de l'éditeur matériel, comme du lecteur. L'ensemble de la chaîne de production éditoriale doit alors être repensé à la fois au regard des technologies numériques et des humains impliqués, quels que soient leurs rôles ; et au regard de la relation de ces derniers aux outils du numérique et aux formes numériques.

Pour une version numérique de l'Anthologie Palatine

Elsa Bouchard, **Marcello Vitali-Rosati**, **Servanne Monjour**
E-mail: elsa.bouchard@gmail.com, vitalirosati@gmail.com, servanne.monjour@wanadoo.fr

L'Anthologie Palatine est une collection byzantine de poésie grecque dont l'ancêtre remonte à la période

hellénistique. C'est grâce à cette anthologie que nous sont parvenus la quasi-totalité des poèmes épigrammatiques grecs que nous connaissons aujourd'hui. Les épigrammes sont de courts poèmes destinés à diverses occasions en fonction desquelles on distingue entre épigrammes érotiques, funéraires, sympotiques, etc. L'Anthologie est basée sur un manuscrit (le Codex Palatinus 23) retrouvé en 1606 par Claude Saumaise à Heidelberg.

Le texte grec tel qu'établi par Paton est accessible sur le site du projet Perseus (Crane 1992). La traduction française la plus récente est celle des Belles Lettres, mais elle commence à vieillir et n'est pas disponible en ligne. La base du projet consiste précisément à produire une traduction inédite de quelques livres de l'AP dont le contenu est d'origine essentiellement hellénistique (livres 4, 5 et 7) et de la mettre en ligne. Notre traduction aura la particularité (absente de toutes les éditions existantes de l'AP) d'être accompagnée des nombreuses références marginales qui apparaissent dans le manuscrits : scholies, titres, remarques sur les auteurs, etc. Or ce type de contenu est particulièrement adapté à un mode de diffusion numérique.

Le constat de départ du projet est le suivant: la culture hellénistique qui a produit ce recueil a un lien très fort avec des traits de la "culture numérique" (Doueihi 2011). En particulier, les compilateurs hellénistiques développent l'idée d'anthologie: un ensemble de fragments qui essaie d'être représentatif d'une totalité. L'Anthologie est une selection qui veut rendre compte de l'ensemble de la poésie grecque, en d'autres termes une sélection paradoxalement exhaustive. Cela pose en premier lieu la question du classement: comment ordonner ces textes? Comment les relier entre eux? Il est évident, en étudiant le manuscrit qui nous est parvenu, que cette question était à la base de la rédaction. L'emploi des scholies en est une preuve. Il s'agit d'annotations - de véritables commentaires - qui permettent de forcer l'ordre linéaire du manuscrit en proposant une sorte de classement non linéaire: on renvoie aux autres poèmes du même auteur, on ajoute des "mot-clés" pour décrire le poème (le nom de la personne à qui le poème est dédié, ou son sujet, etc.).

Nous présenterons notre projet d'édition numérique de l'Anthologie grecque, une édition qui rendra compte de ces caractéristiques qui étaient embryonnaires dans le manuscrit et qui peuvent se réaliser pleinement grâce aux technologies numériques. Notre projet est de mettre en ligne l'ensemble de l'Anthologie (en commençant avec les livres IV, V et VII), en proposant une nouvelle traduction en français, alignée avec le manuscrit et qui permette de classer les fragments de façon non linéaire (par thématique, époque, mots-clés, auteurs, noms cités). L'édition que nous proposons aura à la fois une valeur de recherche indéniable - appareil critique, références, nouvelles traductions - et un grand impact sur le public non universitaire qui pourra accéder en open access à ce trésor de notre tradition littéraire.

Date: Tuesday, 02/Jun/2015

9:00am - 10:30am **Colonel By A707A**	**3a: Collaborative Development**

Curatorial Collaboration: Gamification of a Historic Building

Mitchell Paul Ogden, Dave Beck

E-mail: ogdenm@uwstout.edu

This paper examines the case study of an innovative collaborative undergraduate project at a public, regional polytechnic university (the University of Wisconsin–Stout) to develop a gamified virtual recreation of a historic building and curate a rich digital collection of interactive characters, objects, and historical narratives within the virtual space. Lessons learned from this case study include an approach to deploy an agile development model with undergraduate students together with a pragmatic approach to balancing the authenticity of archival content research from a digital humanities perspective with the interactivity of game design. Finally, the paper suggests an approach to designing undergraduate DH research opportunities that takes advantage of local curricular resources and builds deep collaborations between faculty and students.

A cohort of seven seniors in the Digital Humanities program at the University of Wisconsin–Stout joined in collaboration with twenty 3D art and animation students in the School of Art and Design on a year-long intensive collaborative project to build the gamified virtual environment of Harvey Hall. Harvey Hall was constructed on the University of Wisconsin–Stout campus in 1916 and stands as a showpiece of Progressive Era design and a regional landmark. As the building undergoes renovation in anticipation of its centennial anniversary, various research projects have been undertaken across campus to feature its particular history and role in development of hands-on education in the region.

A year-long senior thesis capstone course was co-taught with an advanced 3D art and animation studio course, bringing together high power game design skills with the curatorial sensibilities of our DH program. This intensive model proved to be a successful collaboration that allowed undergraduate DH students with limited coding and animation skills to witness the actualization of a robust and sophisticated virtual environment. For undergraduate DH programs to succeed, this paper argues that they must create and sustain deep relationships with programs on campus that can contribute tool-based technical skills to large-scale projects. Failure to build such collaborations limits the opportunities for undergraduate DH students to build serious DH projects, resulting in senior projects that struggle to surpass the usual conceptual and theoretical limits. If the field of digital humanities is, in part, defined by its intensive collaboration across disciplines, then undergraduate DH programs need to create structures that will support cross-disciplinary collaboration that actualizes the vision for building and making in collaborative teams.

Laboratories of Modernity

Dean Irvine

E-mail: dean.irvine@dal.ca

In the spirit of Emile Zola's Le roman experimental (1880) in which the naturalist novel becomes a kind of laboratory subject to the laws of scientific method and experiment, the modernists set up their own laboratories for the production of literature, film, radio, the performing arts, and the visual and plastic arts. From Hugo Munsterburg's psychology lab at Harvard and Alfred Stieglitz's 291 gallery in New York at the turn of the twentieth century to the European and North American art and design labs of the 1920s and 1930s, the modernist period witnessed the emergence of institutional formations that brought together artists, writers, film makers, architects, sociologists, psychologists, anthropologists, economists, scientists, and engineers in a transatlantic cultural movement that traversed disciplinary boundaries and fostered new modes of collaboration. What these laboratories of art and design consistently demonstrate is the extent to which the institutional structures of the artistic avant garde were informed by their critique of scientific and corporate models of research; their critique of the very models they reference varied from ideological rejection of industrial capitalism and Western science to aestheticist reform of industrial design. With the creation of studio-laboratories in the 1960s and 1970s, the late twentieth century welcomed a new generation of collaboration among artists, scientists, engineers, and industry that modeled itself on avant-garde labs of the early twentieth century. The opening of MIT's Media Laboratory in 1985 announced its experiments with digital media being "as much like the Bauhaus as a research lab," which at once moved toward the formation of digital-humanities and new-media laboratories of the late twentieth and early twentieth-first centuries and, at the same time, returned to the avant-garde labs of the modernist period.

My investigation of these modes of collaboration and their institutional formations seeks to articulate a genealogy of laboratories—aesthetic, scientific, and corporate—from the late nineteenth century to the contemporary moment of digital-humanities and new-media laboratories and collaboratories. The trajectory of this paper, broadly conceived, traverses multiple disciplines and technologies—from architecture, art, and industrial design to computer markup languages, source code, and digital tools—in the historical contexts of cultural, industrial, and postindustrial modernities. The fluid disciplinary scope of this inquiry finds its focus in the intersection between the production of modernist art as aesthetic experiments conducted under laboratory conditions—sometimes literally, other times figuratively—and the positioning of digital humanists in lab environments performing text-mining and visualization experiments with data and metadata derived from digitally remediated materials originally produced in print and analog media by the historical avant garde. Over the past two decades, as I will demonstrate, laboratories and collaboratories have emerged as institutional locations at which critical theorists, textual scholars, art historians, librarians, archivists, programmers, scientists, and engineers have converged in the implementation of collaborative methodologies and technologies that have transformed and tested the increasingly expansive scale and versatile disciplinary mix of digital modernist studies.

Wrinkleface and the Social Iguana: New Perspectives on What Students Can Do

Miriam Katherine Posner, Noemi Titarenco, Lindsie Levinson, Linta Kunnathuparambil, Jane Chang Mi

E-mail: miriam.posner@gmail.com

In June of 2014, the co-authors of this presentation completed Andean Narratives (mustardy.com/andeadpots), a web-based project that offers new insights on an important collection of ancient Andean ceramics. The team members gathered metadata from UCLA's Fowler Museum of Cultural History for the museum's collection of about 800 ancient Andean pots. They then used existing scholarship to build a "dictionary" of the pots' animal iconography, and used the dictionary as a guide to enhance the Fowler's metadata. The team also augmented the Fowler's metadata with information on the pots' colors, dimensionality, and historical phase. They used this enhanced metadata to build a series of data visualizations, a map, and an improved browsing interface. They also used photogrammetric software to build a number of interactive 3D models of selected pots. Among the team's conclusions: Anthropomorphic Iguana, a commonly depicted Moche creature, served as a kind of bridge icon; he is one of the few creatures that can be found in combination with many other animals (including his frequent companion, Wrinkleface).

The team's work would be impressive under any circumstances, but this group's project is especially distinctive because it was completed in 10 weeks by three undergraduate students and a graduate student, working alongside a faculty member. In this presentation, we propose to introduce this project, but also give a snapshot of our workflow, demonstrating how a university can develop a pipeline from a cultural institution to a DH curriculum. We believe this model has great potential, both as a method of teaching and learning and as a workflow for DH projects themselves.

After demonstrating the key components of the project, we will outline a core set of principles that we believe should govern an interdisciplinary, cross-hierarchy collaboration of this nature. The students who developed this project will explain what they learned and how they learned it. We will share best practices for co-learning while developing a

digital humanities project, and suggest ways that digital humanists can develop productive working relationships with cultural heritage institutions.

9:00am - 10:30am Colonel By B205	**3b: Panel: The Canadian Writing Research Collaboratory**

The Canadian Writing Research Collaboratory: An Experiment in Infrastructure

Susan Brown, Jeffery Antoniuk, Michael Brundin, Mihaela Ilovan, John Simpson, Megan Sellmer
E-mail: sbrown@uoguelph.ca, jeffery.antoniuk@ualberta.ca, brundin@ualberta.ca, ilovan@ualberta.ca, john.simpson@ualberta.ca, sellmer@ualberta.ca

This panel addresses a number of facets of cyberinfrastructure for the humanities through the lens of work related to the Canadian Writing Research Collaboratory/Le collaboratoire scientifique des écrits du Canada—CWRC (pronounced "quirk" for short)—an online environment for the study of writing in and about Canada. CWRC will be launching in fall 2015, and will be beta testing its core affordances and newly developed interface by spring of 2015, so this panel represents a soft-launch engagement with the DH community regarding several key strategies employed by this infrastructure project and some insights gained in its development.

Papers:
- "Infrastructure and/as Research", Susan Brown, CWRC project leader (University of Alberta and University of Guelph)
- "Quirky Paths to Quirky Tools", Jeffery Antoniuk (University of Alberta), Susan Brown (University of Alberta and University of Guelph), Michael Brundin, and Mihaela Ilovan (University of Alberta)
- "CWRC user-documentation: a case for DITA in a digital humanities context", Mihaela Ilovan and Megan Selmer (University of Alberta)
- "An entity-based approach to interoperability in the Canadian Writing Research Collaboratory", Michael Brundin (University of Alberta), Susan Brown (University of Alberta and University of of Guelph), Jeffery Antoniuk, Mihaela Ilovan, and John Simpson (University of Alberta).
- "Tangible Illustrations of the Fusion of DH Research and DH Infrastructure", John Simpson and the CWRC team

9:00am - 10:30am Colonel By D207	**3c: Digital Drama**

Leveraging Performance for the Digital Humanities

Jennifer Roberts-Smith, Paul Cegys, Stan Ruecker, The INKE Research Group
E-mail: sruecker@id.iit.edu

In this presentation, we consider the intersection of performance and the digital humanities in the context of the encouragement or deprecation of interpretive stances toward particular cultural objects of study. In particular, we discuss a production of Richard III, where technology, both on- and off-stage, was used to highlight the understanding of the titular character as a manipulator of performative media.

Digital Acting Parts: Learning and Understanding Shakespeare's Plays

Luis Meneses, Laura Estill, Richard Furuta
E-mail: ldmm@cs.tamu.edu

In the early modern period, rather than having access to a full-text play, actors learned their lines using "Actors' parts," hastily handwritten documents that provided them with only their cues and lines. Digital Acting Parts (DAP) is an online environment that both mimics and enhances the early modern acting experience in order to help users to learn the lines from Shakespeare's plays. DAP is the first project to give users an interactive experience with an early-modern-inspired "actor's part," which encourages both active reading and memorization, in turn leading to a better understanding of the texts themselves.

Digital Acting Parts combines data from several Shakespearean projects. Site visitors who explore The Winter's Tale and The Comedy of Errors—which were both extracted from the XML files and schema provided by the MLA New Variorum Shakespeare—have the added bonus of entering their text in modern English spelling (or speaking it aloud) and seeing their lines displayed in early modern English. Upon completing a scene from these two plays, users are rewarded with links to the First Folio facsimile available through the Internet Shakespeare Editions. We have also parsed and integrated the XML from Open Source Shakespeare, which allows DAP users to interact with all of Shakespeare's plays and poems.

In Orality and Literacy: The Technologizing of the Word, Walter Ong theorized that we were in a "second orality" based on the "use of writing and print" and considered how memorization worked in oral and written circumstances; In Shakespeare studies, the function of memory has been an important area of study for years, as scholars debate whether certain texts are "memorially reconstructed," that is, printed from actors' memories rather than a playwright's written text. The motivation behind our work is to synthesize the relationship between understanding memorization when it comes to cognition and the comprehension of literary and theatrical works— all while making use of emerging online technologies. More so, the use of DAP brings forth two questions: How successful are the users of our platform in memorizing and understanding plays? And how can we measure their progress? For this purpose, we propose to measure and evaluate the learning and memorization patterns of users by inspecting the access and interaction logs of our system. Taking into account the audience and different backgrounds of the users of our platform, we also plan to investigate why some plays are more popular in different geographical regions.

Digital Acting Parts can be accessed at http://digitalactingparts.tamu.edu.

9:00am - 10:30am Colonel By E015	**3d: Mobile Considerations**

Creating a Mobile and Augmented-Reality Scholarly Edition: Swinburne's Poems and Ballads, 1866

Bethany Nowviskie, Graham Wayne
E-mail: bethany@virginia.edu, wsg4w@virginia.edu

Scholarly editing, in a literary context, involves the painstaking examination of scores of printed and hand-written documents for minute textual differences, in an effort to understand, represent, and sometimes combat change over time—the accidental and intentional transformations history wreaks on poetry and prose. These processes are physical as much as intellectual, and so is the work of the bibliographer; but you'd never know it to look at our scholarly editions—particularly at contemporary digital editions, which are stuck in 1990s paradigms of design interaction. Our project takes the physicality of bibliography and textual criticism as its subject and invites readers and fellow editors to reach out, and step away from their desks. An in-progress scholarly edition of Swinburne's scandal-rocked and fundamentally-unstable **Poems and Ballads** (1866) becomes a playground for tablet-based presentation, augmented-reality interaction, and experimentation with textual collation and the work of scholarly editing using immersive VR. We will offer a brief textual history of **Poems and Ballads,** discuss the theoretical and design contexts in which we are undertaking this work, and then describe four prototypes or experiments-in-progress. These are: 1) a parallel-segmented TEI edition with an API-based interface designed for the special affordances of tablet and mobile displays; 2) a simple, clean, print-on-demand classroom reading text in which supplemental material and scholarly notes are made available through the augmented-reality viewport of a mobile phone; 3) a set of interactive timelines drawing on concepts from the Temporal Modeling Project, which depict both

evidence-based and conjectural textual histories of **Poems and Ballads,** including alternate histories suggested by the forgeries of T.J. Wise and the thought experiments of Swinburne's biographers; and 4) work in progress toward the optical collation of high-quality page images using head-mounted virtual displays like the Oculus Rift.

Authenticity and Play in Mobile Apps
Keith M Lawson
E-mail: klawson@dal.ca

Institutions of cultural memory and tourist bureaus are now making use of applications for mobile devices to connect visitors or tourists with museums or with historical districts. A large body of research exists on the motivation and goals of tourists and, more recently, on their use of mobile devices, particularly inexpensive digital cameras. However, little work exists on tourists' use of mobile applications to navigate scenes of cultural memory. Some applications lead users through a scripted series of locations or interactions to a (desired) goal of understanding or community. Others require users to explore and even to solve physical or mental challenges in order to create a sense of meaning or authenticity. The purpose of this paper is to examine how mobile apps designed to guide users through scenes of cultural memory take different routes to try to solve the problem of delivering authenticity.

MacCannell's seminal work on tourism argues tourists search for "authenticity of experience" in a reaction to the "shallowness of their lives and inauthenticity of their experiences" (1973). For MacCannell, tourists visit "places of social, historical, and cultural importance" in the same way that pilgrims "visit a place where an event of religious importance actually occurred" (1973, p. 593). More recent writers on the tourist experience have reacted against MacCannell's conception of tourists and of the authentic. For Cohen, the "post-modern tourist" is predisposed to "playfulness" (1988, p. 20). Molz regards authenticity, "as a discursive construct through which certain technological practices and mediated connections are made meaningful by tourists…" (Molz). She concludes that "mobile technologies … do, to a certain extent, re-enchant a disenchanted world" (Molz).

This paper discusses a number of specific applications in relation to their conception of the tourist/user and the idea of authenticity. For example, the 9/11 Memorial Museum's Explore 9/11 app takes a pilgrimage approach, leading users on a scripted path toward a goal of insight and community. But this approach raises questions about the experience it seeks to create, as Adam Gopnik notes in his review of the 9/11 museum, the project embodies contradictions: "an insistence that we are here to remember and an ambition to let us tell you what to recall" (Gopnik, 2014). On the other hand, the Museum of London's Streetmuseum app has no pre-defined path or goal, but engages the user in the sometimes frustrating challenge of locating graphic representations of historical buildings or events geolocated around the streets of the city, and where any feeing of authenticity or re-enchantment must be created by the user through an interaction with the technology.

Cohen, E. (1988, January 01). Authenticity and commoditization in tourism. Annals of Tourism Research, 15

Gopnic, A. (2014, July 7). Stones and Bones: Visiting the 9/11 memorial and museum. The New Yorker. Retrieved from http://www.newyorker.

MacCannell, D. (1973, November 01). Staged Authenticity: Arrangements of Social Space in Tourist Settings. American Journal of Sociology

Molz, G. M. (2012). Travel Connections: Tourism, technology and togetherness in a mobile world. London: Routledge.

Public Heritage at Scale: Building Tools for Authoring Mobile Digital Heritage & Archaeology Experiences
Ethan Watrall
E-mail: watrall@msu.edu

The spaces and places we inhabit and interact with on a daily basis are composed of layers of cultural activity that are, quite literally, built up over time. While museum exhibits, historical and archaeological narratives, and public archaeology programs can communicate this heritage, they generally do not allow for rich, place-based, and individually driven exploration by the public. In addition, public heritage programs rarely explore the binary nature of material culture, the preserved record of human activity, and heritage: the presented information about the heritage and the scholarly process by which that knowledge was generated. In short, the scholarly narrative of material culture, heritage, and archaeology is often hidden from public exploration, engagement, and understanding. Further, traditional public heritage and archaeology programs often find it difficult to support rich and vibrant multivocality, social interaction, narrative co-creation, or citizen scholarship.

In recent years, the maturation of mobile technology and augmented reality have offered both platforms and models for mobile heritage applications that at least partially address these issues. Projects such as The Museum of London's Streetmuseum, Histories of the National Mall, and the CHESS Acropolis Museum mobile application facilitate place-based public interaction with heritage and archaeology. Unfortunately, what we are not seeing are mobile heritage experiences that facilitate multivocality or that explore the process by which cultural, heritage, and archaeological knowledge is generated.

It is within this context that this paper will introduce and explore mbira. Developed as a collaboration between Michigan State University's MATRIX: The Center for Digital Humanities & Social Sciences and the Cultural Heritage Informatics Initiative, mbira is an open source platform designed to empower individuals, projects, and institutions to create and sustain compelling mobile heritage experiences. In addition to introducing mbira's authoring architecture, the paper will explore and interrogate the ways in which mbira was purpose-built to address aforementioned critical shortcomings in many mobile heritage applications.

10:30am - 11:00am **Colonel By A707A**	**Coffee Break 3: Coffee Break**
11:00am - 12:30pm **Colonel By B205**	**4a: Repositories and Communities**

Digital Humanities Projects and Digital Repositories
Elli Mylonas
E-mail: elli_mylonas@brown.edu

Digital Humanities projects have been aware of the importance of longevity and reusability, and have been practicing a form of data curation, for a number of years. This wasn't always a result of knowledge and best practices, but arose out of the expectations on the part of humanists that their materials would be used continuously and for a long time, and because of the effort and expense invested by humanists to carry out these projects. Although significant counterexamples can be found, it was in the humanities that SGML (later XML) were adopted and the TEI developed, and where projects like Perseus and the Women Writers Project have been running for over 20 years, although their delivery systems and capabilities may have evolved.

Institutional repositories, deployed by university libraries and IT departments, are most commonly intended to provide preservation and discoverability. However, they are increasingly emerging as a platform where research data can be stored and retrieved, and more intriguingly, re-used. There is an emerging synergy between digital humanities projects that create or use digital resources, and institutional digital repositories that can both preserve these resources and provide APIs for real time re-use. There are also inherent conflicts because of the differing fundamental purpose of each and despite the fact that using a repository promises to separate the creation of resources from a particular use. Repositories are centralized, benefit from uniformity and are relatively static. Digital Humanities projects can benefit from this, but are less predictable, change frequently and focus on the unique.

The Brown Digital Repository (BDR) is provisioned primarily by the library repository team and metadata specialist,

and houses special collections, archives, faculty research and some digital humanities projects. Its data models and API are transformative for Brown's DH projects, which can draw special collections materials directly from the repository, and store their own digital data in it. It can also be present challenges; data models that are too broad the inability to change data and metadata. A close look at how DH projects have been using the BDR and interacting with the BDR team can provide some examples of emerging models and interactions.

As Brown DH projects have begun to collaborate with the BDR, they have had to forge new relationships, develop new workflows and sometimes start to use software frameworks that they would not have chosen otherwise. They may also find that they cannot be as nimble as they might have liked, because they don't control the development process, and are sharing a space with other resources. In some cases, a parallel framework had to be developed. Furthermore, DH projects already require collaboration between faculty, DH experts and various technical people. The repository team can in some cases replace the DH team, but in others needs to be included. These various accommodations and new practices are not only enhancing collaboration between the BDR and the DH projects using it, they are also improving the repository's features and intervening in the way DH projects are carried out.

Iterative Prototyping of Small-Scale Experimental Digital Humanities Infrastructure

Matthew Hiebert, William Bowen, Ray Siemens
E-mail: hiebert8@uvic.ca

At the founding of the non-profit partnership in 1994, Iter interpreted its mandate to support Medieval and Renaissance research and teaching in terms of the development of finding tools and of infrastructure for the dissemination and publication of resources. By 2006, when user-created data affordance principles associated with Web 2.0 technologies were receiving early large-scale application, Iter had earned a sufficiently substantial user base to begin discussing the facilitation of scholarly communications and user-created resources within the frame of its mandate, and to be prototyped as an online social "collaboration resource" called "Iter Community." This paper discusses how the original vision behind the Iter Community platform came to involve the iterative prototyping of small-scale experimental infrastructure for digital humanities projects. In keeping with the assertions of Geoffrey Rockwell (2010), that small-scale infrastructure experiments are "valued research in humanities computing" and that infrastructure must serve not only "professional researchers at universities, but the amateur researchers in the community," our paper will relate how the development of the Iter Community platform as a social knowledge creation environment for a community of practice necessitated an experimental approach to "the cloud" distinct from corporate and academic institutional models.

We argue Iter Community infrastructure is a unique contribution to non-empirical modes of inquiry, reflecting an implication of the trajectory towards problem-based models of knowledge representation in digital humanities: specifically, the increasing number of small-team or individually generated and often experimental, boutique digital humanities research projects out there. A sandbox space for digital project creation was among the features most requested by an advisory group within their 2013 assessments of an earlier iteration of Iter Community. We perceive the proliferation of small-scale digital projects to arise from the same forces that continue to sustain the general trend of what Kathleen Fitzpatrick has called "big tent" digital humanities (2010). The international expansion of hybridized problem-based pedagogical models that adopt successful models developed by such organizations as the DHSI, HASTAC, and THATCamp contribute especially to the manifestation of such grassroots DH projects. These training models, often developing student skills through experimental play in small-scale project- creation, have only recently begun to foster reflection upon requirements for shared humanities- specific infrastructure to support and sustain the digital

{OBJ}{OBJ}{OBJ}

outputs associated with such learning activities, and to allow their further iteration and dissemination post-training. This early awareness towards the recognition of iterative humanities-oriented infrastructure design as a scholarly activity (alongside tool and metadata prototyping which have already partially won such status) is perhaps typified by the recent inclusion of an infrastructure creation course at the DHSI.

Iter Community offers a long-term home for boutique projects created by emerging and lay scholars seeking to find collaborators, support, and a community of practice inclusive of academic, alt-academic, and non-academic members; in addition, to larger, established projects requiring development, production, and/or archival resources. In reflecting on this year of backend experiments at small-scale, we hope to productively contribute to the thinking around and development towards larger- scale infrastructure projects for DH.

User Assessment with Emblematica Online: A Case Study of Analyzing Scholarly User Engagement with Digital Humanities Research Initiatives

HARRIETT E GREEN, MARA WADE, TIMOTHY COLE, MJ HAN
E-mail: green19@illinois.edu

Digital archives today not only provide unprecedented access to rare archival materials, but also enhance the exploration of these materials with tools for image analysis, data mining, and the like. But there are still large gaps of knowledge on how these digital resources are used in humanities research and pedagogy, and into the efficacy of digital collections for targeted user communities. In response, this paper reports the results of a user study of Emblematica Online, currently in a second phase of funding by the National Endowment for the Humanities, which provides access to digitized copies of Renaissance emblem books from the leading rare book collections around the world.

Studies of humanities scholars' interactions with digital resources frequently note the importance of scholars' expertise in the development of digital collections. [1] User studies of digital resources, such as those by Meyers (2011), which resulted in the Toolkit for Impact of Digitised Resources (TIDSR), emphasize the importance of scholarly users to the sustainability of digital projects.[2] And for specialized collections, Daly (2002) cites the importance of librarian-scholar collaborations in the ongoing curation of digital emblem books and similar digital special collections. [3] And with its particularly global scope, Emblematica Online provides an opportunity to assess the engagement of a broader, international community.

The paper will discuss the results of user interviews and observational usability testing conducted with scholars and students at a public, four-year institution and at the 2014 Society for Emblem Studies conference. The analysis of these user responses and interactions with the Emblematica Online portal seeks to shed light the current and potential uses of Emblematica Online, and on a larger scale, how scholars and students make use of digital collections in their research.

This analysis of user engagement with Emblematica Online offers lessons of interest to other researchers developing digital tools and research collections. The paper will offer insights on how digital humanities researchers can engage users in the creation and sustainability of their research projects, and maximize the impact of their work on the scholarly community.

References:

[1] Audenaert, Neil and Richard Furuta. "What humanists want: how scholars use source materials." In *Proceedings of the 10th annual joint conference on digital libraries* , 283-92. New York: ACM, 2010; Palmer, Carole L., Oksana Zavalina, and Katrina Fenlon. "Beyond size and search: Building contextual mass in digital aggregations for scholarly use." In *Proceedings of the American Society for Information Science and Technology* 47, no 1. (2010): 1-10. doi: 10.1002/meet.14504701213.

[2] Meyer, Eric T. *Splashes and Ripples: Synthesizing the Evidence on the Impacts of Digital Resources. Joint Information Systems Committee Report* . London: JISC, 2011. http://ssrn.com/abstract=1846535; *Toolkit for the Impact of Digitised Resources* , http://microsites.oii.ox.ac.uk/tidsr/welcome.

[3] Daly, Peter M. *Digitizing the European Emblem: Issues and Prospects* . New York: AMS Press, 2002.

11:00am - 12:30pm	4b: Panel: Médiation et performativité

Colonel By A707A

Médiation et performativité : l'auteur s'éditorialise

Marcello Vitali Rosati, Servanne Monjour, Julie Tremblay-Devirieux, Élisabeth Routhier

E-mail: marcello.vitali.rosati@umontreal.ca, servanne.monjour@umontreal.ca, julie.tremblay.devirieux@gmail.com, elisabethrouthier@hotmail.com

La fonction auctoriale, telle que nous la connaissons aujourd'hui dans le modèle de l'édition papier, s'est institutionnalisée à partir du XVIII siècle avec la mise en place des premières lois sur le copyright (Rose 1993). Cette fonction répond à une triple exigence: celle de garantir la fiabilité des contenus, d'en certifier l'originalité et d'en définir les conditions de propriété. Lorsqu'il annonce la "mort de l'auteur" en 1968, Roland Barthes entérine l'affaiblissement de cette fonction auctoriale déjà préssentie par les structuralistes. Aujourd'hui, alors même que le modèle papier est concurrencé par le web, cet affaiblissement de l'auteur en tant que fonction s'accompagne de l'émergence d'une autre structure : l'auteur en tant que "profil". Il nous semble en effet que, dans l'espace numérique, l'auteur ne doit plus se chercher du côté de la personne, mais du côté des stratégies de construction auctoriales introduites par le Web. Pour reprendre la formule de Benoît Bordeleau, blogueur québécois, « L'auteur est peut-être mort, mais il écrit encore » : sur les réseaux sociaux, sur les sites web, etc. L'écriture numérique devient le moyen par lequel l'auteur se produit, habitant dans l'espace du texte et s'exprimant dans cet espace. En conséquence, la distinction entre auteur et personnage – très claire dans le modèle papier – a tendance à s'estomper. L'auteur devient profil, ou mieux, profils au pluriel. Pour désigner ces pratiques de production du profil, ou de l'identité auctoriale à la fois dynamique (puisque sa construction n'est jamais définitivement achevée) et hétérodéterminée (par les lecteurs notamment), on pourrait dire que l'auteur s'éditorialise. L'éditorialisation (Bachimont 2007, Vitali-Rosati 2014) désigne en effet l'ensemble des pratiques et des dispositifs à travers lesquels un objet apparaît dans l'espace numérique. Il s'agit d'un processus qui est toujours ouvert dans l'espace et dans le temps : dans l'espace, parce qu'il ne se limite pas à une plateforme ou une page particulière, dans le temps, parce qu'il s'agit d'un processus toujours en devenir, toujours en mouvement. Dans le cas spécifique de l'auteur, on peut identifier une série de pratiques qui consistent à exprimer la figure de l'auteur sur des plateformes différentes : Twiter, Facebook, le blogue. De fait, un auteur est à la fois sa photo de profil, son statut Facebook, sa signature, mais aussi sa notice biographique et l'ensemble des traces qu'il laisse et qu'il continue toujours de produire. L'éditorialisation désigne donc une série de stratégies de médiations de soi par lesquelles l'écrivain performe (Butler, 1990) son identité en tant que profil-auteur. Ce sont ces stratégies que nous voulons questionner dans les quatre communications de ce panel.

11:00am - 12:30pm **Colonel By D207**	**4c: Blogs and Twitter**

Utilising 'Tweet My Street' to understand geographical nuances in support and opposition to the 2014 International Day Against Homophobia and Transphobia (IDAHOT)

Graeme William Mearns, Rebecca Simmonds

E-mail: graeme.mearns@ncl.ac.uk, r.m.simmonds@ncl.ac.uk

It was in *The City and the Grassroots* that Manuel Castells (1983) first revealed the importance of space for lesbians and gay men in San Francisco's Castro District. This seminal work documented how social prejudice, legal repression and political violence forced the invisibility of homosexuality. In turn, Castells noted how this creates a 'major obstacle to finding sexual partners, discovering friends and leading an unharassed open life' (1983: 4). Following this study, many geographers, sociologists and others working in the field of sexuality and space have made clear that visibility has been key to LGBT identity politics for over three decades. Visibility demands recognition of societal norms that construct public space as heterosexual, heterosexist and heteronormative (Bell and Valentine, 1995). Residential clustering (Adler and Brenner, 1992), the use of urban 'gay villages' in city marketing (Rushbrook, 2002) and the politics of Pride (Browne, 2007) have been key foci in geography. However, the field has also suffered Anglo-American and Western European bias whilst neglecting provincial and rural locations (Engebretsen, 2008). This has meant that many queer experiences remain unaccounted for (Oswin, 2006). As Stella (2012: 1825) contends, 'contextualizing queer space as a phenomenon linked to multicultural citizenship and cosmopolitan consumerism is not particularly helpful in contexts where there is no visible "gay scene" or where sexual citizenship does not extend beyond private consumption'. Research on transnational sexual solidarities is gaining traction as one way to interrogate the absences and ethnocentric assumptions about the empowering potential of visibility in these respects (Binnie and Klesse, 2014).

A collaborative effort between geography and computing, this paper draws on the emerging sexual solidarities literature to discuss the politics of visibility in relation to the spread of locative technologies, social media and an interrelated marketplace for 'big data'. We describe the custom development of a social media analytics tool that can aid researchers in mapping geospatial data, spatially and temporally, at different geographical scales. With reference to findings of network and content analyses performed on a subset of 10.5 million geo-located 'tweets' that referenced sex-gender minorities during IDAHOT, the paper aims to open up a space to reflect on the digital tools humanities, social science and computing scholars are now able to use to help advance understandings of sexuality and space, the ethical dilemmas inherent to doing so and some of the new questions we might wish to pose. With particular attention to geospatial data derived from Twitter in the 'worst places in which to be gay' (Withnall, 2014), we question whether there is now anywhere for queers to be invisible at a time when geospatial data is so readily accessible and computer algorithms can predict gender and sexuality by mining online metadata (Kosinski et al., 2013). Notwithstanding the benefits of legal reforms that have occurred on the back of LGBT visibility, improving the lives of many, we contend that as opportunities to mitigate the effects of bias in scholarly and activist work on sexuality, there are also potential dangers and place-centred geographical nuances remain central.

Digital Humanities and 140 Characters: A uses and gratifications study of Twitter

Anabel Quan-Haase, Kim Martin, Lori McCay-Peet

E-mail: aquan@uwo.ca, kmart5@uwo.ca

DH scholars, those who use "computational tools to do the work of the humanities" (Unsworth, as quoted in Gold, 2012, p. 67), have been identified as early adopters and avid users of social media (Kirschenbaum, 2010; Ross et al., 2011). Ross et al. (2011) have described how Twitter plays a role during DH events as a form of backchanneling. It has also been shown that Twitter can be used to amplify DH scholarly communication (Howard, 2009), and that the use of Twitter to share your own work will result in a climb in its readership (Terras, 2012). This paper examines Twitter use by DH scholars from a uses and gratifications (U&G) perspective, investigating (a) how digital humanities scholars are using Twitter and (b) what gratifications they obtain from its use. While the U&G approach has been utilized to study college students' use of social media (Quan-Haase & Young, 2010) and specifically Twitter (Chen, 2011), it has not, until now, been applied to the study of scholars.

LGBT nation and Otaku brotherhood in Cuba. A study of two communities excluded from the Cuban national culture discourse thought their blogs

Yasmín Silvia Portales Machado, Sheila Padrón Morales

E-mail: yasmin@cubarte.cult.cu, sheila@bpvillena.ohc.cu

Cuba has a very low internet access. Cuban government explains this due to the lack of resources for infrastructure development and the USA Embargo. To compensate this, the government had establish that most of Cubans only have access to the national intranet (.cu web sites) and charges with very high fees the cybercafés with free internet access, owned by the state telecommunication monopoly, ETECSA.

Blogs edited by Cubans residents in the island exist since 2005, in free blogs platforms like Blogia, BlogSpot and WordPress. Most of them devoted to politics and news. In September 2013 the blog platform cubava.cu open to free

inscriptions. Being the only space of online self-publication technically accessible from the all country, the democratization impact was absolute: in less than a year the Cuban blogosphere doubled in size and spread in subjects. Between the groups that take advantage of this were LGBT activist and Otakus (Japanese anime and manga fans).

This two groups are marginal subcultures in Cuban society, which is still strongly homophobic, have cultural roots in Africa and Spain, and a strong dependence of USA cultural products. This hegemonic model of the national culture is defended in the state public policies and the discourse of most of the cultural specialist. LGBT and Otaku advocates are annoying because they challenge the model. While the LGBT groups fight against cultural, legal and social homophobia, Otaku groups argue their right to consume pop Japanese product (anime, manga, videogames, roll games) without been labeled has "childish" or with "proclivity to violent behavior". Booth groups are starting to create independent spaces for documentation, dialogue and action. These places for interchange are mostly virtual –e-zines, chats, blogs, Facebook groups–, due to the strict state control of the public spaces and the media.

Cuban new cultural policies most recognize these emergent social identities and add it to the "official" national culture, in order to maintain the significance and bonding meaning of this idea for the new generations, or they will risk a fracture in the national identity.

Studies about Cuban online communities are scarce. Most of the research about Cuban blogosphere (Ted Henken, Elaine Diaz, and Yudivian Almeida) is focused in the impact of political blogs contents in the government actions. Only two papers had research the tensions between the post of individual blogs and a community political agenda in LGBT activist (Sandra Alvarez) and women bloggers (Yasmín Portales). They are not research about the Otaku Cuban Community.

The study of the Cuban LGBT and Otaku blogs will give evidence based knowledge about the editors profiles, the communities internal dynamics, the strategies they had develop to establish national range networks, the extension of this groups in the Cuban society, and the arguments they use to defend their belonging to the XXI century Cuban national culture.

| 11:00am - 12:30pm Colonel By E015 | **4d: Data** |

Improving Research Data Visibility and Accessibility, for Better Data Sharing and Higher Research Profile.

Peter Webster

E-mail: peter.webster@smu.ca

Thesis. This paper puts forward the thesis that methods and resources are available to make digital humanities data more visible, accessible and sharable online.

This presentation will focus on easy steps that digital humanities data creators can take to make sure that their data can be found online, and can be readily used and cited by other researchers.

We will examine the online landscape of available data repositories and data directory services, as well as basic tools for managing, describing and retaining research data.

Topics discussed will include:

- Internet discovery, deposit and subject access.
- Project and dataset description and standardized metadata essentials.
- Persistent identifiers for projects, data sets, and researchers.
- Requirements for data citation. Gaining scholarly credit for cited data.
- Addressing copyright and other rights issues.

Methodology: Review and features comparison of over 50 international central data repositories and data directory services. Functionality review of Canadian Instructional repositories. Literature review.

Conclusion: Greater centralized online infrastructure, training and human resources are badly needed, and many access and rights issues remain to be solved. But standardized metadata, and use of online registration and repositories can greatly improve shared access to digital humanities data resources.

Data Stewardship in the Digital Humanities

Sonja Christina Sapach, Geoffrey Rockwell, Catherine Middleton

E-mail: sapach@ualberta.ca, grockwel@ualberta.ca

In 2013 Canada's research funding agencies released a consultation document titled "Capitalizing on Big Data: Toward a Policy Framework for Advancing Digital Scholarship in Canada" that proposed a "realignment" to establish a "culture of stewardship." The document argued that our ability to preserve and manage research data is not keeping up with the promise of big data analysis. The digital humanities (DH) have long dealt with issues around text/data stewardship, in particular issues around text encoding and scholarly editing. In this paper we will make the case that what is needed is more than good data management. We need to treat the management of research data as a scholarly activity and recognize the data deposits that follow best practices as a form of publication for the purposes of tenure and promotion. Making available research data so that future scholars can make use of it should be considered as an activity in a scholarly continuum. Specifically our paper will do the following:

- Summarize the case made for a changed culture of stewardship in light of the potential for big data.
- Survey the state of research data management in Canada.
- Present recommendations for the social sciences and humanities.
- Make the case for stewardship as scholarly activity.

Canada's research councils (SSHRC, NSERC, CIHR, CFI and Genome Canada) have started a process to consider ways to preserve, archive, and make accessible data generated through publicly funded research projects. As one of the action items of the 2nd Annual Digital Infrastructure Summit organized by the Leadership Council for Digital Infrastructure (http://digitalleadership.ca/) we prepared a Digital Management Plan (DMP) Evaluation Summary Report summarizing the data management approaches of different worldwide funding agencies. We followed this up with a draft DMP Recommendation prepared for SSHRC and other interested agencies. This paper will be based on these two documents.

Currently efforts to plan and manage publicly funded data resources in the fields of the social sciences and humanities lack consistency. Research data is not available on an open access basis, nor is there widespread acceptance of the importance of data management and archiving practices. Data stewardship should be the responsibility of researchers, institutions, and funding agencies in Canada and Canadian scholars need to come together to develop a set of standards and practices that will satisfy the OECD Declaration on Access to Research Data from Public Funding that Canada and 33 other countries adopted in 2004.

The evolving field of DH is well-equipped to take the lead in developing standards and practices that will standardize and naturalize the practice of effective data stewardship in the social sciences and humanities in Canada. This paper, following the lead of DH scholars, outlines some recommendations for fostering data stewardship practices among Canadian humanities and social sciences researchers.

This paper ultimately argues that we need to think of research data contributions as scholarly contributions when they are well documented, follow community standards, are deposited for preservation and are made available in a timely and open fashion.

SylvaDB: A Framework for Research Productivity in the Digital Humanities

Javier de la Rosa, David Michael Brown, Juan Luis Suarez, Adriana Soto-Corominas

E-mail: jdelaro@uwo.ca, dbrow52@uwo.ca

This paper presents SylvaDB as a framework for the management of digital humanities data that employs a hybridized relational and graph approach to data modeling and transformation. SylvaDB was developed by the CulturePlex Lab to facilitate the process of collecting and managing highly connected and complex data collected during the Hispanic Baroque project (Suárez & Olid-Peña, 2007). Specifically, we needed to create entity–relationship (ER) style data models that are well suited to metadata curation (Batini, Ceri, & Navathe, 1991; Chen, 1976), but also implement techniques borrowed from graph theory and network analysis. SylvaDB's core functionality provides flexible ER modeling while allowing users to leverage the graph database's transformational power (Bastani, 2014), thus simplifying the processes of curation, statistical analysis, and data manipulation at the time of retrieval. This workflow has been used with high success rates in our projects appearing in a wide variety of both high profile digital humanities conferences and peer reviewed publications (Suárez, Sancho Caparrini, & de la Rosa, 2011; Suárez, Sancho, & de la Rosa, 2012; Suárez et al., 2013; Caldas, 2014).

12:30pm - 2:00pm **Colonel By A707A**	**Lunch Break 2: Graduate Student Networking and Mentoring Lunch**
2:00pm - 3:30pm **Colonel By B205**	**5a: Reading and Narrativity**

Connecting the Dots: Integrating Modular Networks and Narrativity in Digital Scholarship.

Amy Robinson, Jon Saklofske, Research Team INKE
E-mail: ajmmy.robinson@gmail.com, jon.saklofske@acadiau.ca

Lev Manovich posits that new media and the World Wide Web are modular or layered in nature, similar to structural computer programming in that their distinct elements are combinative while retaining independence. Modular structures and systems (such as Lego, mobile apps, computer software, and even language itself) exchange precision, specific connectivity, narrative stability, and a focus on progressive products for flexiblility, general compatibility, adaptiveness and a focus on aggregative processes. These attributes are well suited for new knowledge environments: Peter Schillingsburg sees modularity at the heart of dynamic digital collaboration and Susan Brown (et al) liken the modularity of digital projects to the cumulative nature of academic periodicals but caution that such projects are still often judged "as if they were a book," or by their apparent finishedness. While scholarly journal articles and monographs, as finished products of scholarly activity, aren't often constructed to demonstrate modularity or to function as modular components in broader arenas of scholarly communication, they could be reimagined as such (beyond citation). This is something that INKE's NewRadial environment encourages through its modular design and in the kinds of modular scholarly communication that it encourages.

When NewRadial was first conceived, "edges" or connection points between database object nodes were imagined as opportunities for users to link and constellate distinct data objects together. This feature was both limiting and enabling: limiting users to only being able to insert commentary as part of a connection or comparison, and enabling them to find and argumentatively justify unique patterns in cultural datasets. In this case, edges simply represented connection without direction, encouraging users to generate networks of association. While recently extending NewRadial's functionality to include compatibility with relational datasets (in which edges require a specific direction to make sense of the relation), researchers had to introduce the opportunity for users to specify unidirectional edges. While this seems like a small "tweak" to the features and functionality of the application, introducing unidirectional edges actually introduces an opportunity for users to create narrative "routes" or paths through the data, which is something that earlier iterations of NewRadial prohibited through its design choices. Enabling the creation of specific, higher-resolution routes through modular networks promotes the productive co-existence of modularity and narrativity in digital scholarship. Whereas traditional scholarly communication obscures its modular aspects through narratively constructed illusions of finishedness and completeness, through work defined by production and consumption, NewRadial foregrounds, embraces and extends the playful making opportunities related to modular potential while retaining narrativity as an optional form of construction and assertion.

On the Value of Narratives in a Reflexive Digital Humanities

Lai-Tze Fan
E-mail: lychee@yorku.ca

This paper will discuss the digital narrative as a critical framework for research and pedagogy in the digital humanities, arguing that it offers a new methodological scope through which to identify how narrative can form resonant, interpretive meaning out of bodies of data.

The digital narrative is a narrative text that is digitally informed, whether by way of being digitized, transcribed, embedded with code, or born-digital. In these ways, it exemplifies Alan Liu's theory of a "logic of connectivity" a cultural phenomenon that occurs "artistically as intertextuality, appropriation, sampling, and so on" (Local Transcendence 2; original emphasis). Examples of digital narratives thus include: texts that connect and converge media (intermedial and transmedial texts); born-digital texts (electronic literature); emerging textualities and genres in online culture (online narrative games and the "visual novel"); and texts in older media that represent digital media and their properties.

Through the introduction of transmedial, trans-spatial, and trans-temporal qualities to cultural texts, the digital narrative raises questions of what it means to "read" digital media. These questions are especially pertinent for digital humanities researchers, who must approach their projects with theoretical and analytical methods that account for the unique textualities of new media. In undertaking these methods, digital humanities researchers must also ask whether relying on methodological prowess through new media can lead to a potential loss of humanities-based reflection and interpretation.

In order to consider how the digital narrative can acknowledge and account for the new theoretical and analytical strategies of this researcher, and by extension, of the teacher who requires pedagogical methods specific to the digital humanities, this paper explores the digital narrative as a critical framework in the digital humanities. This framework is developed out of current research that calls for investigative analysis that can account for reflexive and interpretive elements in the digital humanities (Hayles 2012). Addressing this need, this paper begins to build on arguments that data requires narrational logic in order to produce resonant meaning. For example, literary scholar N. Katherine Hayles contends that any scientific and engineering research presented through data and facts requires narrative for "the interpretation of the relations revealed by database queries" (182). I argue that this observation illustrates the praxis and necessity of narrative even for research that is grounded in data.

In order to demonstrate a digital narrative framework in use, this paper will identify the use of narrational logic in representing texts in a digital framework. A digital narrative framework, I argue, permits the analysis of documents and projects that are linked in small and large collections and projects. Such projects often involve what Franco Moretti (2005) and others call "distant reading," a method through which researchers must make meaning out of thousands of digitized texts and documents by mapping patterns. I will analyze how the mapping of macro-level narratives can develop distant reading and research theory and praxis, with the objective of complementing and expanding the digitization, storage, and access of humanities documents.

Entre édition papier et édition augmentée : la collection Parcours numériques

Marcello Vitali Rosati, Michael Sinatra, Hélène Beauchef, Giuseppe Cavallari
E-mail: marcello.vitali.rosati@umontreal.ca, michaelesinatra@gmail.com, lnbeauchef@gmail.com

Dans cette intervention, nous présenterons la collection Parcours Numériques créée aux Presses de l'Université de Montréal par Marcello Vitali-Rosati et Michael Sinatra en 2014. La collection propose des textes en accès libre et tente de développer un nouveau modèle de publication scientifique associant deux formes différentes de publication : une publication numérique augmentée en accès gratuit et une autre plus traditionnelle, imprimée. Les trois premières publications sont sorties en 2014 : le manuel collectif Pratiques de l'édition numérique dirigé par

Marcello Vitali-Rosati et Michael Sinatra, l'essai Âme et iPad du philosophe Maurizio Ferraris (traduction de son ouvrage publié en italien en 2011) et Mémoires audiovisuelles. Les archives en ligne ont-elles un sens ? de Matteo Treleani. Trois autres publications sont prévues pour 2015.

La collection est basée sur l'idée qu'il doit y avoir une complémentarité entre l'édition papier et l'édition numérique, ces deux formes de publication présupposant différentes idées de lecture et deux approches à la réception des contenus.

Le livre papier – et l'on entend ici également le numérique homothétique (epub ou pdf) qui reproduit à l'identique le livre papier sur un support numérique – permet une lecture linéaire. Une thèse peut y être présentée et argumentée de façon complexe. Le lecteur sera capable de suivre de manière linéaire le développement de l'argumentation, de faire un cheminement avec l'auteur en se laissant accompagner d'un bout à l'autre du discours. C'est pour cela que nous avons fait le choix de publier des livres assez courts (120/200 pages) : c'est la longueur adéquate pour présenter une thèse et la démontrer à l'aide d'une argumentation unique et cohérente.

L'édition numérique augmentée, en revanche, présuppose une lecture non linéaire, qui procède par approfondissement. On commence par lire un premier texte sur un sujet, on souhaite en approfondir un aspect, puis on glisse sur un contenu qui se trouve ailleurs et qui nous permet d'en savoir plus sur ce qui, au départ, ne semblait qu'un détail. On navigue, on flâne, le parcours emprunté n'existe pas avant la navigation, il n'a pas été prévu par un auteur ou un éditeur. Ces derniers ont suggéré des pistes, ils ont ouvert des portes… puis la navigation est laissée aux lecteurs, à leurs envies, à leur créativité. Dans ce sens, il n'est pas vrai, comme le voudrait Nicholas Carr, que le numérique détruit notre capacité d'attention : il s'agit d'une attention différente, disséminée, qui permet l'approfondissement mais empêche de suivre un discours plus long et unitaire.

La collection est aussi en train d'expérimenter un modèle alternatif à l'évaluation par les pairs traditionnelle. Grâce à un partenariat avec la plateforme Media Commons, nous sommes en train d'expérimenter une évaluation ouverte pour un manuscrit proposé pour la publication. Lors de notre intervention nous présenterons les premiers résultats de ces expérimentations et discuterons des statistiques d'accès et du modèle économique de la collection.

2:00pm - 3:30pm **Colonel By A707A**	**5b: Panel: From Documents to Data**

From Documents to Data

Harvey Noel Quamen, Paul Hjartarson, Veronica Belafi, Andrea Johnston, <u>Nicholas van Orden</u>, **Constance Crompton**, Michelle Schwartz, Hannah McGregor
E-mail: hquamen@ualberta.ca, nicholasvanorden@gmail.com, constance.crompton@gmail.com

"From Documents to Data" is a panel that examines the dynamics of three DH projects, all of which are currently making data, rather than XML documents, the central core of their respective workflows. These three projects -- one from periodical studies, one from cultural history, and one from social network studies -- will explain how they have approached and solved the various research problems that are encountered in a move towards what we might call "data-centric" thinking in the Digital Humanities.

2:00pm - 3:30pm **Colonel By D207**	**5c: Navigating Memories**

Memories/Motifs, Historical Memory & the unexpected inspirations of Digital Humanities

<u>Rachel Deblinger</u>
E-mail: rdeblinger@gmail.com

Memories/Motifs (http://memoriesmotifs.com/) is an online exhibit that examines the use of multiple media forms to construct stories about Holocaust survivors in the immediate postwar period and consider how American Jews first learned about the Holocaust after the war. The exhibit traces the stories of three survivors as they were transformed from one media form to another and uses Scalar to showcase the variety of materials employed by Jewish communal groups to collect and disseminate these narratives in the wake of the war, including archival documents, fundraising materials, radio broadcasts, short sponsored films, and newspaper articles. Through hyperlinked text and multi-media materials, Memories/Motifs connects postwar Holocaust narratives with contemporary testimony collections and builds a network of meaning that extends and enriches our understanding of how Holocaust memory has been constructed in America.

This paper will explore three aspects of building Memories/Motifs that call attention to the use of digital tools in examining the history of collecting, preserving, and disseminating stories about the Holocaust through new media technologies. First, the paper details how linking historical materials, including text, film, audio, and visual artifacts, reveals new patterns about postwar Holocaust narratives and identifies continuities between the postwar and today – suggesting a long history of Holocaust memory construction in America. The paper will then explore the potential for Memories/Motifs as a form of public history that invites user interaction with historical materials and empowers both scholars and non-scholars to rethink acquired knowledge about silence in the postwar period. Finally, I will confront the limits of digital tools to address ethical questions about memory. Memories/Motifs calls into question the use of new media to transform survivor narratives for wide dissemination in the aftermath of the war through yet another mode of new media for yet another public audience. Thus, in doing this work, am I perpetuating the same behavior that I'm critiquing? And, where is my own culpability in the construction and appropriation of these stories?

Rather than retreat in the face of these questions, I have expanded the online world of Memories/Motifs by launching a Tumblr blog and Twitter handle. These online platforms allow me to further the conversation about postwar memory in America, the construction of Holocaust narratives, and the employment of Holocaust memory for a range of contemporary issues. Thus far, I have written about Obama's latest immigration speech, September 11th memorial practices, Rosh Hashanah in DP camps, and what makes Holocaust memory "appropriate." Social media and digital tools allow me to experiment with new forms of public facing historical thought to reflect on the implications of my own historical research and the legacies of Holocaust memory America.

By juxtaposing issues of methodology, digital meaning making, and ethical access, this paper explores how building Memories/Motifs offers unexpected avenues for imagining and reimagining the process of constructing postwar Holocaust narratives and how the project continues to provoke new kinds of questions about the ways we create memory.

Navigating through Memory Island of Stanford Encyclopedia of Philosophy – a demonstration of the Memory Island Technique

<u>Bin YANG</u>, Jean-Gabriel GANASCIA
E-mail: bin.yang@lip6.fr

Recently, Knowledge maps are promising tools for navigating through knowledge spaces in Digital Humanities and Digital Libraries. With the help of knowledge maps generated by a domain expert or an Information Visualization (InfoVis) technique, users can discover new information contents through navigating the knowledge space. A well designed knowledge map can make sense of the knowledge by using metaphors, and help users to achieve their information seeking tasks. However most of existing knowledge maps were infographics, they were manually created by the knowledge experts for some special knowledge. Manually creating of knowledge map is a challenge when the dataset becomes large.

Knowledge Visualization is a worthwhile topic for researchers in digital humanities. Ontology is one promising tool for knowledge management, as it can support the semantic reasoning and the visualization. Knowledge can provide more power if they are presented in ontology. Therefore, a visualization that try to use the power of ontology for a dataset could bring more insights for its users.

In this paper, we describe one case study on visualizing a large knowledge dataset (encyclopedia) with our Memory Island technique. We discuss how to generate a meaningful Memory Island for the Stanford Encyclopedia of Philosophy (SEP) with the help of Indiana Philosophy Ontology (InPhO) and InPhO API. The Memory Island technique was inspired by a method called "loci" (plural of Latin locus for place or location) of the ancient "Art of Memory" technique, which described how people in the antiquity and the Middle-Ages used spatialization to increase their memory capacity. Memory Island technique consists of creating a virtual insightful map and associating each entity to its designated areas. We defined some geographic metaphors and interactive functions in this Memory Island technique to help users with their knowledge understanding and visual discovery. According to our users' experiments, most of them can easily achieve their navigation with SEP's Memory Island, like using the Google map. Results show that the use of Memory Islands provides advantages for non-experienced users tackling realistic browsing, and that it helps with memorizing knowledge and information for most users. An interactive demo of this visualization is accessible at http://www-poleia.lip6.fr/~polyle/2014-12/result-inpho/index.html.

The Astrolabe Project

Anwesha Bhattacharya, Richard Furuta, Filipe Castro
E-mail: furuta@cs.tamu.edu

Introduction

Western European mariner's astrolabes evolved from Islamic Medieval instruments, which in turn evolved from late Roman astronomic instruments, which were a result of the long history of geometry and celestial interpretations. These instruments are simple and reliable, and allowed sailors to know their latitude by measuring the height of the sun at noon and consulting mathematical tables with the position of the sun on every day of the year. Nobody knows when the first mariner's astrolabe – probably made of wood – was used, nor who wrote the first tables with the declination of the sun, which are much older than the first references to mariner's astrolabes. The use of mariner's astrolabes is documented between the late 15th and the late 17th centuries, when they were replaced by more accurate instruments. Today marine astrolabes are rare collector's items, and only around 100 are known to exist, of which more than half are badly eroded.

Astrolabes' styles vary according to their provenience and date of making, but so far no in-depth study of their shape and decorations has been made. The challenges are several: the sample is small, many specimens are unprovenienced, the majority is eroded or lacking components, and there may be a few unidentified fakes in the collection under analysis. Because of these difficulties, a digital representation becomes attractive, both as a means of making a rare resource available to a broader audience but perhaps more importantly in order to allow development of techniques to conduct detailed analysis, categorization, and reconstruction. Towards these goals, we are initially working to establish a taxonomy based on the selection of a certain number of recorded stylistic characteristics of the known astrolabes. This is a first step that will enable a more complex undertaking, including shape recognition and multivariable analysis.

The traits considered in the first phase of this study were the astrolabe's size, weight, and shape of the wheel, ballast, posts, and crown (see Figure 1). A visualization from the results of the first iteration in this study is presented as Figure 2. There are other variants, such as the suspension ring or the alidade, and we intend to consider a wider range of morphological characteristics in a second study of this collection. In a few cases the astrolabe's makers are known and can be clustered around family ties, cities, countries, or regions. Consequently, will work to relate the developed taxonomy to the known data pertaining to each astrolabe's provenience and date of making. The expected outcome of this project is a stylistic grid that will help identify the provenience and chronology of astrolabes that entered the market without provenience.

2:00pm - 3:30pm **Colonel By E015**	**5d: Canadian Archives**

(Re)construire les archives de Gabrielle Roy

Sophie Marcotte
E-mail: sophie.marcotte@concordia.ca

voir pdf

The Fred Wah Digital Archive: Affective Labour and Grassroots Sustainability in Digital Humanities Initiatives

Deanna Marie Fong, Ryan Fitzpatrick, Janey Dodd
E-mail: deannaf@sfu.ca, rfitzpat@sfu.ca

A recent Association of College & Research Libraries (ACRL) email thread on sustainability identifies a common concern about DH initiatives: respondents expressed frustration at "seeing a number of really fabulous digital projects wither and die after supporting faculty retired or the grant funding ended." This tension was a core issue in the development of the digital archive of Canadian poet Fred Wah. The project began as a traditional bibliography, and evolved into a digital working environment where scholars could access, organize, and interact with Wah's texts. The site was originally developed at York University on Drupal 5, where federal funding allowed the team to create a custom interface for the archive's multimedia content. When the grant ended and investigators moved on to other projects, the site was not migrated to successive versions of Drupal and the content went dark. The project's suspension suggests that sustainability in DH extends beyond the scope of funding and technical support into larger political and institutional structures. As Jerome McGann has noted, one way of initiating structural reform is to shift the driving question from "what is the problem?" to "who are the agents?" (2)—an approach that places the human, rather than the product or project at the centre of DH research.

Our paper will detail our recent work on the reboot of the Fred Wah Digital Archive at Simon Fraser University. As it is one of the first in a wave of DH projects of at SFU—and the only project initiated and directed by students—we have found ourselves in the unique position to develop a template for digital sustainability that may be implemented by future projects. Based on our experience rebuilding this site, we will suggest four best practices toward a grassroots paradigm of sustainability: 1) fostering transparent, equitable relationships between all of a project's agents. Beyond the four major players that McGann identifies (scholars, publishers, librarians, and funding agencies) we suggest that this model be expanded to include systems technicians, research assistants, artists and writers (or representatives of their estates), and students, who may also belong to any of the categories listed above; 2) clearly delegating technical and editorial maintenance tasks; 3) documenting back-end and content development, and depositing these documents in an established archival space; 4) diffusing the project funding structure across institutional levels and bodies, and cultivating long-term commitments with these agents. These practices rely to a certain degree on what Michael Hardt has termed the "affective labour" of human contact and interaction that produces "a sense of connectedness or community" (96). Affective labour builds durable professional relationships and provides a framework for mentoring and generational exchange over time. While this practice is largely the norm in DH circles, like Hardt, we advocate for a formal recognition of this work within the community as a means of organizing "biopower from below" (99). Such a gesture responds to McGann's call to protect the interests of educational and scholarly communities against the encroachment of for-profit publishing models in digital cultural spaces.

First steps in integrating diverse data sources to tell a fuller story of the treatment of Japanese Canadians in WW2

Stewart Arneil, Holmes Martin
E-mail: sarneil@uvic.ca, mholmes@uvic.ca

The Landscapes of Injustice project draws on various types of sources to document the treatment of Japanese

Canadians in World War 2: their forcible confinement, the seizure and disposal of their assets, and the destruction of their neighbourhoods. We will use community directories, photographs, fire insurance maps, oral histories, land title records, and cabinet minutes to reconstruct these events. The project aims to produce narrative websites, archives, public exhibitions and other outputs. Each data source is most effectively dealt with using a specific toolset and stored in a particular format or system. On the other hand, a unified dataset is needed to support the range of outputs. This presentation addresses the issues that arise in integrating data from disparate historic sources to produce a data structure which supports a range of outputs.

This talk focuses on the community directories, insurance maps and title records. These three clusters have disparate data types and formats: xml-encoded texts, the semi-proprietary Access-based format used by ArcGIS, and tabular data in a relational database, respectively. We will briefly discuss how the characteristics and workflow for each source determine which data format we used.

Next we will talk about the reasons for our decision to use a unified data medium: XML. For example, converting all our data to XML while retaining information implicit in tabular relationships is trivial, whereas the converse is not. Integrating new and updated data into the central model is straightforward. XML is not just a markup language; it is now surrounded by a constellation of related querying, transformation and rendering languages (XQuery, XSLT, CSS), which make it an ideal platform for integrating our diverse data.

We have created an XSLT transformation which builds indexes of building and address information from the GIS data, title and owner information from the land title data, and information about residents from the community directories. It builds a unified dataset and renders it in XHTML5 which is both human-readable (in a web browser) and formally queriable (since XHTML5 is XML). We can scan the result looking for anomalies which suggest errors or omissions in the source data; we can easily do calculations and we can read in detail the history of a particular address, and the owners, residents and businesses associated with it. We will demonstrate this during the presentation.

As our source data becomes more complete and detailed, we will design a broader range of outputs, including street plans, timelines, and heatmaps, using the same techniques.

This work, although it appears to be primarily technical, is in fact a way of illuminating the injustice done to the Japanese community in Canada. For example, at 510 Alexander Street, the 1930 Vancouver directory lists only "Orientals", as does the 1942 directory; but from a combination of the 1941 directory and the Japanese community directories, we can place 16 Japanese individuals at that address in 1941. Restoring these individuals to the record is the first step in illuminating and describing what happened to them two short years later.

3:30pm - 3:45pm **Colonel By A707A**	**Coffee Break 4: Coffee Break**
3:45pm - 5:15pm **Colonel By B205**	**6a: Theory**

Art as Device / Criticism as Computation: What Russian Formalism Has to Say about the Digital Humanities

Glen Worthey
E-mail: gworthey@stanford.edu

This paper is an exploration of the theory and practice of contemporary Digital Humanities in the light of an important historical antecedent: early twentieth-century Russian Formalism. These two critical schools share a surprising set of structural and theoretical features.

Until recently, DH scholars have not often engaged the Russian Formalists as methodological or cultural forebears. The principal aims of this paper is to begin to fill that gap by discussing three broad areas of similarity between "us" (the international DH community) and "them" (a small, but disproportionately influential and memorable, group of critics of about a century ago): the focal point of our studies, the use of technology as metaphor and tool, and an uncanny similarity in our cultural histories.

Locus of Study

Contemporary DH shares with the Formalists an explicit focus on language as the stuff of literary study, and on linguistics as one of its proper modes of inquiry. This cross-historical "collaboration" in stylistics continues to be a fruitful one. Rather than taking psychoanalytical, biographic, rhetorical, or ethnographic approaches, both practices hold that the meanings of a text are couched in its words alone; what contemporary DH has done that the Formalists could not do is to train our computational tools on those words.

Likewise, as the Formalists advocated for expanding literary history beyond just the "generals of literature," so contemporary DH techniques such as distant reading aim to comprehend "the great unread. This radical reconsideration of the literary canon undertaken by both Formalists and DHers does not simply remake the canon (e.g., by making it more inclusive); rather, we explode it in order to explore it.

Technology and the Literary Machine

Like DH, the Russian Formalists were notably technocentric. The Formalists used technology more as metaphor than as method, and DH is defined a priori as technological in its methods. But in both cases, our engagement with technology is more than superficial, a deep focus on the nuts and bolts of poetic texts, in contrast to the metaphors, tools, and methods of the critical schools from which both we and the Formalists differentiate ourselves.

Believing the Formalist trope of "Art as Device" is not so different from our using technological devices to interrogate art. We both approach culture as something that can be taken apart, studied closely (or distantly), and interpreted; both make conscious efforts to employ some version of the scientific method and rigor in evidence-gathering, rather than the more subjective, anecdotal approaches of other critical schools; likewise, both are often subject to criticism as overly positivistic or reductionist.

Our Place in the Cultural Zeitgeist

Both DH and Formalism have very specific historical sources for our relative technophilia: Russia's industrial and political revolutions, and our own century's digital one, are important in both our histories. Both our work and the Formalists' depend on these revolutionary upheavals.

A study of these commonalities will both enlighten our own disciplinary self-awareness and reveals ways that DH might accomplish some of the Formalists' goals for which they lacked tools.

L'éditorialisation et la frontière entre le réel et l'imaginaire

Marcello Vitali Rosati, **Servanne Monjour**
E-mail: marcello.vitali.rosati@umontreal.ca, servanne.monjour@wanadoo.fr

La séparation entre réel et imaginaire se fonde traditionnellement sur la possibilité de distinguer de façon claire l'espace du discours de l'espace hors du discours. L'imaginaire s'oppose au réel en ce qu'il se trouve, justement, dans l'espace du discours. Les relations entre ces deux espaces ont fait l'objet de nombreuses interprétations, bien que l'on puisse mettre en évidence deux démarches décisives (Guidone 2003). D'une part, l'imaginaire peut être pensé comme un non-être - ainsi que le voudrait Sartre, notamment dans L'imaginaire (1940). D'autre part, et de façon inverse, l'imaginaire peut être pensé comme structurant le réel - c'est ce que propose notamment Valéry (Jarrety 1981). Dans un cas comme dans l'autre, réel et imaginaire sont opposés. L'hypothèse que nous souhaitons défendre cherche à se démarquer de ce rapport d'opposition : il nous semble en effet que le fait numérique brouille les frontières entre imaginaire et réel ou, mieux, qu'il produit une superposition entre les deux termes, au point de rendre leur caractérisation quasi non pertinente.

Notre argumentation s'appuie sur les quatre points suivants :

1. La réalité tend à s'identifier de plus en plus avec ce que Luciano Floridi (2014) appelle "infosphère", dont l'espace

numérique fait partie.

2. L'espace numérique n'est pas un discours sur le monde mais le monde lui même. Le concept que nous utiliserons pour expliquer cette correspondance est celui d'"éditorialisation" (Vitali-Rosati 2014a et b).

3. Dans l'espace numérique se trouve une série d'objets discursifs que l'on peut assimiler à ce qu'on appelle traditionnellement "imaginaire" : il s'agit notamment des objets littéraires - récits, nouvelles, etc.

4. En raison de la coappartenance à l'espace numérique d'objets discursifs différents, mais aussi en raison de l'assimilation de l'espace du discours à la réalité, la distinction entre ce qu'on appelle imaginaire et ce qu'on appelle réel devient de moins en moins pertinente.

Considérons par exemple le processus d'éditorialisation d'une ville, regroupant notamment le mappage Google, les recensions Trip advisors, les données sur Wikipédia ou sur DBpedia, les images, les sites institutionnels (le site de la ville, de ses musées)... Quand on marche dans la ville, on se trouve dans un espace produit par ces pratiques d'éditorialisation : la carte Google, les informations sur les restaurants ou les horaires d'ouvertures des musées cohabitent avec les murs, les bâtiments, l'architecture. La ville est constituée de l'ensemble de ces éléments. Or au milieu de ces éléments nous retrouvons aussi des "informations" qui relèvent de la fiction. Dans le cas de Paris, nous pouvons trouver des données ajoutées à la carte Google, qui "mappent" la ville dans le cadre d'un récit fictif : c'est ce que fait par exemple Cécile Portier dans Traques traces. Ici, ces données côtoient, dans une carte Google (http://www.petiteracine.net/traquetraces/map/node), les autres "informations" sur le lieu dont il est question. En d'autres termes, si la ville de Paris se superpose à l'infosphère qui la concerne, alors, puisque cette infosphère est aussi composée de Traques Traces, la fiction Traques traces est aussi Paris. Notre communication approfondira l'étude de cet exemple, afin d'appuyer notre hypothèse liminaire.

Towards a Quantum Digital Humanities

Grant Glass
E-mail: grant.glass@kcl.ac.uk

"My life amounts to no more than one drop in a limitless ocean. Yet what is any ocean, but a multitude of drops?" — David Mitchell, Cloud Atlas

Project Proposal: Towards Quantum Digital Humanities

In this project, I argue that the concept of quantum theory has the potential to address language in a different way than humanities has understood it. What if a system could be created around a single word that would describe all of its possible quantum states (meanings, interpretations, contexts)? Traditionally, humanities discourse can describe (for Jacques Lacan), the symbolic and imaginary aspects of language. Quantum theory offers us the chance to discover what the real is by creating a system outside of interpretation. As Schrödinger's Cat as taught us, we cannot measure (or in humanities terms, read) a text without distorting it. The solution could be to create a system outside of interpretation within a computer.

I propose to create a system within a computer to aggregate all the possible meanings and interpretations around the word, abandon which meaning has changed significantly over the years. Its current use is to "give up completely", like abandoning hope or abandoning a family member. However, in 14th century Middle English it meant "to subjugate or subdue" someone or something. It had stemmed from from the French phrase "mettre a bandon" meaning "to give up to a public ban". It is my hope to measure the trajectory of the word to predict how it might be further evolved in the future through capturing all its current contexts within a computer program. By completely capturing the model for a single word, it is my hope to make a model for a real understanding of abandon.

The Method of this project will involve a discussion on quantum theory's influence on the humanities. Specifically seen through the theories of Jacques Derrida and Jacques Lacan, a discussion on the nature of interpretation of language will be provided in order to shape the need for a system outside of the traditional humanities to gain a greater understanding of language. A machine model is needed to examine language as absolute rather than undecidability.

In order to capture all of the meanings and possible meanings of abandon, I will be using the Predictive Model Markup Language (PMML). By using PMML, I will be able to create a complex and complete model of abandon. This module will represent the word, and it will become the basis for teaching a computer to read and understand language much like a human can. This will be the start of providing computing a more complete understanding of language and reading, and hopefully more complex and rich methods of distant reading.

3:45pm - 5:15pm **Colonel By A707A**	**6b: Panel: Digital Cultural Heritage**

Digital Cultural Heritage

Brent Nelson, **Craig Harkema**, **Daniel O'Donnell**
E-mail: brent.nelson@usask.ca, craig.harkema@usask.ca, daniel.odonnell@uleth.ca

This panel of papers arise from the work conducted by the Digital Cultural Heritage (DigitCultH) sub-group of the Digital Humanities research group in GRAND-NCE: Canada's Digital Media Network of Centres of Excellence (http://www.grand-nce.ca/). Recognizing that the Galleries, Libraries, and Museums (GLAM) world has pioneered the use of Web technologies for representing cultural heritage objects, we are exploring the degree to which the current state of the art serves the needs of scholars and ways in which we as scholars can contribute new affordances to the representation of digital cultural heritage objects on the Web.

*SHO and Tell: Digitized Cultural Heritage Objects for Research and Public Engagement

Craig Harkema (University of Saskatchewan)*

Increasingly, digital humanities work involves both collaborations with scholars and activities that are more public-engagement oriented. These partnerships, along with an increase in digitization work that concerns cultural heritage objects rather than print materials, requires a better understanding of the range of activities taking place in museums at the intersection of public/academic and curatorial/scholarly. To this end, my colleagues and I began a bibliography and whitepaper outlining current projects, available tools, and recommendations of best practices for scholar-curator collaboration for public engagement with digitized cultural heritage objects. This paper outlines the findings of our environmental scan in an attempt to better understand some of the following research questions:

1. How can digital methods help to bridge between the gap between curatorial and scholarly practice?

2. How can we both invite public engagement with historical materials and also encourage the citizen scholar to engage more deeply with the content, to take the user beyond curious interest to deeper investigation of historical materials.

3. How can we develop and present both curated and scholarly materials together in a way that enables multi-level engagement while maintaining appeal for both the curious tourist and the invested scholar?

It is important to emphasize that environmental scan is not for a GLAM audience; rather, we are talking to humanities scholars and collaborators about the current state of the literature on treating cultural heritage objects.

To offer practical application on the themes we uncovered in the bibliography, this paper will also draw on my experience with Sask History Online (SHO: saskhistoryonline.ca). SHO is a Saskatchewan Ministry of Education funded cultural heritage digitization project that involves over 60 GLAM partners and has generated over 300,000 digital records from across the province of Saskatchewan. It contains several examples of the sort of public/academic and scholarly/curation intersections we are seeking to better understand. These collaborations required a variety of technological and theoretical approaches . This paper will highlight scenarios involving social media, imaging, metadata, digital asset management, and digital curation among others.

*Est ce qu'il y a de hors-edition? or, Can you edit everything?

Daniel O'Donnell (University of Lethbridge)*

Traditionally, academia has maintained an institutional and disciplinary distinction between "cultural" and "textual" heritage, or, very broadly speaking, between the study and preservation of a given culture's physical remains and the study and preservation of its literary tradition. The former activities--the curation, preservation, and presentation of its sculpture, buildings, manuscripts, paintings, and landscape--has been the preserve of the GLAM (Galleries, Libraries, Archives, and Museums) professionals; the latter--the content, history, and meaning of its books (and, to a lesser extent, its paintings and other forms of art)--has been reserved for the literary scholar or textual or art historian. There have always been individuals (particularly among cultural heritage professionals) who are comfortable in both worlds--the archivist-historian, for example, or the curator-critic. But these have, for the most part, been understood as people who are adept at two different tasks, rather than people who represent the convergence of these two broad disciplines. On the whole, we have been content to preserve our distinction between those who study and preserve objects and those who study and preserve content--between the curator and the philologist.

This distinction is not completely without reason. The discovery, curation, and preservation of physical objects require different training, techniques, and levels of trust and access from those required for scholarship of their content. And while scholars tend on the whole to write primarily for other scholars, GLAM professionals are often required to serve a mobilization and outreach function not normally associated with the university researcher.

Recent years, however, have seen a convergence of these roles. Scholars and GLAM professionals increasingly find themselves using similar (digital) tools and techniques to capture and represent the tangible and intangible aspects of cultural and literary heritage. There is, moreover, a convergence of function and audience as well. Mobilization mandates and the increasing visibility associated with web-based publishing platforms have meant that scholars increasingly operate in the public sphere at the same time as technological advances allow conservation professionals to engage in activities--such as the compilation of virtual collections from dismembra--that were previously the domain of the university researcher.

This paper looks at the implication of this convergence from the perspective of the scholar. How does the increasing visibility of scholarly research, and the ability to use techniques and tools from the world of cultural heritage affect the type of content-focussed work scholars have traditionally engaged in? To what extent can scholarship assist in the mediation of physical objects as well as their containers? Our answers to these questions are based on work done by members of the Digital Cultural Heritage research team within the GRAND National Centre of Excellence.

*The Early Modern Collection of Curiosities as Model for Multi-Layered Interfaces

Brent Nelson (University of Saskatchewan)*

Web curation of cultural heritage objects has been pioneered by GLAM practitioners, with the result that most of the available resources are tailored to the needs and interests of a public audience. In recent years, with the growth of interest in material culture among humanities scholars, many of us have begun to think about ways in which digital resources and processes can facilitate humanities research. Building on the context established by Craig Harkema and Daniel O'Donnell, this paper focuses on the central question of DigiCultH: in representing cultural heritage objects on the Web, how can the interests of the public-facing museum and the academic scholar relate. It begins by considering the modern museum as a knowledge environment, tracing the intersection of knowledge generation and dissemination with public education back to the earliest forms of the museum--the cabinet of curiosities in the late sixteenth and early seventeenth centuries. Almost from their inception, English cabinets of curiosities were open to the public, giving people from all walks of life access to a rapidly expanding world of knowledge and experience, but the principal function of these collections was to inform the research of those who collected and circulated these objects.

The bulk of this paper is an examination of the ways in which early exhibition and dissemination of collections created layers of access for multiple audiences and levels of engagement and uses. The early modern cabinet of curiosities will thus serve as a model for Web exhibitions that leverage the affordances of image-based Web interface to engage curious members of the public, while at the same time providing layers of archived information and historical documentation. These early museums are the subject of my own Web-curation project, "The Digital Ark," a virtual museum of collections of rarities and curiosities in England and Scotland from 1580-1700, comprising documentary and graphical representation of up to 10,000 specimens and artifacts collected in that period, some of them surviving today in museums in England. The archive includes some images from seventeenth-century publications, and some photographs of extant objects, but most of the material is textual: XML encoded transcriptions of historical documents (catalogues, inventories, correspondence, visitor books) keyed to a relational database containing extensive information on people, places, and bibliographic objects. My paper will conclude with a discussion of a prototype interface that attempts to implement this layered approach, focusing on the collection of John Bargrave (1610-1680), whose collection remains significantly intact in the Canterbury Cathedral archives. The objective of this interface is to use the visually-rich qualities of this collection to create a graphical interface that gives meaningful context and access to the historical documents contained in the Digital Ark archive.

3:45pm - 5:15pm **Colonel By D207**	**6c: Text Analysis Traditions**

Reproducing Text: Observations from a Pre-Digital Humanitie

Daniel Powell
E-mail: djpowell@uvic.ca

Writing in *Literary and Linguistic Computing* , Julianne Nyhan et al argue that "without a better understanding—a more appropriate term might be 'body of interpretations'—of the near and distant history of computing in the humanities, we are condemned to repeat the revolutionary trope *ad infinitum* " (2-3). Willard McCarty, amplifying this, has written that "rather than hypnotizing ourselves with supposedly unprecedented marvels, we must learn to see computing in its historical and social contexts, and so be able to ground modelling in something larger than itself. For computing to be of the humanities as well as in them, we must get beyond catalogues, chronologies, and heroic firsts to a genuine history. There are none yet" (255).

This presentation aims to, perhaps polemically, test the boundaries of accepted histories of digital humanities by considering an equally technologically sophisticated pre-digital humanities. Highlighting particularly those technologies of textual reproduction developed prior to the oft-quoted originary moment of 1949, I draw on the history of Early English Books Online (EEBO) to argue that while a *computational* humanities may indeed be limited to the last half-century, the *technological* humanities—in both materialist and cultural senses—have a much longer history.

EEBO contains over 125,000 titles, all "in full digital facsimile from the Early English Books microfilm collection." That microfilm effort began in 1931, intensified as World War II loomed, and continues today. Digital images of these microfilmed documents were made (and are still being made) available online first in 1998. The printed *Short Title Catalogue* (itself published in 1926) has determined what objects were photographed and, subsequently, scanned and put online ("About"). The history of EEBO crosses multiple media, was directly impacted by global war, involves private companies and public universities, and is both analog and digital. To bracket EEBO (and EEBO-TCP) as a only a digital project impoverishes our understanding of how digital technologies have impacted the reproduction, preservation, and use of texts in humanistic scholarship.

In part, this presentation seeks to answer claims (articulated by Alan Liu, Tara McPherson, and others) that digital humanities does not engage with socio-critical questions. Blending media analysis, historical perspectives, and knowledge of technical infrastructure, I hope to question the boundaries of what we consider digital humanities to be, how we write our histories, and how we move forward.

References

"About EEBO." *Early English Books Online* . < http://eebo.chadwyck.com/marketing/about.htm>

Julianne Nyhan, Andrew Flinn, and Anne Welsh. "Oral History and the Hidden: Histories project: towards histories of computing in the humanities." *Literary and Linguistic Computing* Advance Access published 30 July 2013. DOI: 10.1093/llc/fqt044, 2-3.

McCarty, Willard. "What's going on?" Literary and Linguistic Computing 23.3 (2008): 253-261. DOI: 10.1093/llc /fqn014, 255.

Characteristic Curve: Reinterpreting Early Analytics

Geoffrey Rockwell, **Stéfan Sinclair**
E-mail: grockwel@ualberta.ca, sgsinclair@gmail.com

(Please see attached abstract.)

In 1887 the polymath T. C. Mendenhall published an article in Science titled, "The Characteristic Curves of Composition" which is both one of the earliest examples of quantitative stylistics but also presents one of the first text visualizations based on the (manual) count of words. Mendenhall thought that different authors would have distinctive curves of word length frequencies which could help with authorship attribution.

This paper is about the recapitulation of Mendenhall's and other early analytical methods. In the paper we propose to:

• Argue for historical recapitulation of key methods in the history of text analysis as a way understanding the historical formation of the analytical imagination.

• Walk through three important early methods in the tradition of humanities text analysis.

• Show how methods can be shared and explained as Notebooks following a "literate programming" paradigm.

Exploring Cryptography: Issues for Digital Humanities

Quinn DuPont
E-mail: quinn.dupont@utoronto.ca

Prior to the 20th century movement towards technological instrumentalism, the study and use of cryptography functioned more like digital humanities today: working across disciplines, offering new perspectives on textual materials through digitization, and employing the latest computational technologies. Many of the issues now facing the study and use of cryptography were first approached within the humanist tradition. Throughout Renaissance and Early Modern history, cryptography was seen as the leading edge of linguistic research, spurring advances beyond the mere transmission of secret letters. This paper accomplishes two goals: 1) it introduces key issues facing the analysis of cryptography within a humanist tradition, and 2) it argues that digital humanities is an ideal home for new and emerging (or neo-humanist) analyses of cryptography.

I explore four key interpretive issues facing cryptography from the perspective of digital humanities: digitality, language and text analysis, combinatory and statistical thinking, and modular production. I open each issue archeologically, with an single illustrative example. First, I look at how Francis Bacon's biliteral cipher was an early way of understanding the world digitally. Second, I show how John Wilkins' key work of modern language scholarship, *An Essay towards a Real Character and a Philosophical Language,* had its seed in his earlier cryptography manual *Mercury.* Third, I discuss how the development of combinatory thinking machines from Raymund Lull to Gottfried W. Leibniz was based on co-evolving cryptographic primitives. And fourth, I show how Leon Battista Alberti's development of modular, replicable production was motivated by work in his cryptography manual *De Cifris.*

With the deepening role of cryptography in modern life, more attention will be given to its rich humanist tradition. This paper offers an early investigation into some of the contours and themes of an important technology in an emerging digital humanism.

3:45pm - 5:15pm **Colonel By E015**	**6d: Working with Poetry**

Immersive Data Environments: 3D Visualizations of Phonemic Data

Marc Plamondon
E-mail: marcp@nipissingu.ca

This paper will outline theoretical considerations with regards to the three-dimensional representation of the phonemic content of poetry and will present two innovative representations. A modicum of attention has been paid to issues surrounding the visualization of poetic data: see Benner, Coles and Gonnering Lein, Chisholm (1976; 1981), Clement, Forstall and Scheirer, and Plamondon (2005; 2009). The basic assumption is that poetry contains patterns of data that are not always readily apparent. With computational processing of phonemic content, the difficulty lies in the multi-dimensionality of phoneme characteristics: while phonemes represent a potentially useful set of data because there are only so many of them, a significant reduction of the high dimensionality of word-based text analysis, individual phonemes can be classified in multiple ways: consonants and vowels, open and closed, short and long, voiced and unvoiced, etc. Each dimension of a text representing a phoneme can itself be represented by multiple dimensions, defined by the type of phoneme under consideration by that dimension. Recent attempts to visualize the complexity of phonemic data have been unsatisfying because of these layers of dimensionality.

The first approach this paper presents consists of a more standard 3D object model approach, in line with what Michael L. Benedikt outlines when he talks about the "spatiology of information": extrinsic and intrinsic attributes are combined to produce virtual objects, each representing a particular phoneme or phoneme class. The combination of these objects represents a visualization of the phonemic content. For example, one can see at a glance by the number of quickly spinning objects how many voiced fricative phonemes are present; the number of slowly spinning object represent how many unvoiced fricative phonemes are present. This approach is very close to the data modelling presented by science fiction writers (such as Gibson and Stephenson), and represents a step forward in the visualization of phonemic content. The second approach is more innovative: it consists of an immersive 3D environment where the appearance of the landscape and the objects contained therein correspond to the phonemic content. An example of this will be demonstrated, where the phonemic content of each syllable of a poem produces a mini-environment within a natural setting. A reader or researcher can then explore the shifting landscape more directly, immersing herself into the sound constituents of the poem.

These approaches, especially the latter, attempt to balance the semiotic representation of the abstracted data with immersion. The natural environment approach is meant to obfuscate what is otherwise a too-obvious relationship between sign and signifier. While the semiotic representation still exists, it is much less apparent, thereby potentially allowing for a deeper understanding of the abstracted data. It is more reflective of the reading process, where discovery, rather than data, is key. It helps transform a poem (following Barthes's terminology) from "work" to "text"; the kinetic aspects in particular help us visualize the implied "grain of the voice" (Barthes again) of the text, leading to an understanding of the words that is less computational and more experiential.

Schooling Donald Allen: Re-Locating Mid-Century American Poetry Networks

Lisa Chinn, **Brian Croxall**, **Rebecca Sutton Koeser**, **Kevin Young**
E-mail: brian.croxall@emory.edu, rebecca.s.koeser@emory.edu

Drawing on extensive archival materials, "Schooling Donald Allen" analyzes networks of authorship and editing of mid-century poetry journals. In particular we are investigating how the "schools" that Allen created for organizing his influential anthology, *The New American Poetry* , correspond with the fuller, published record.

Semantic Notations: On Building "Belfast Group Poetry _Networks_ "

Brian Croxall, Rebecca Sutton Koeser
E-mail: brian.croxall@emory.edu, rebecca.s.koeser@emory.edu

Belfast Group Poetry *Networks* began with a question: what more could be done with library data? This presentation discusses lessons learned from a two-year project to enhance library data about the work of the Northern Irish writing workshop, as well as what our analysis of the transformed data has taught us about literary community and unintentional biases in archival data.

5:30pm - 7:00pm **Colonel By C03**	**Award Plenary: CSDH/SCHN Outstanding Achievement: Chad Gaffield**
7:30pm - 9:00pm	**Reception: at the Royal Oaks pub (on campus): 161 Laurier Avenue East**

Date: Wednesday, 03/Jun/2015	
8:30am - 10:00am **Louis-Pasteur 154**	**Digital Demos 1: Digital Demonstrations**

The VAT: Video Analysis at Scale

Virginia Kuhn, Alan Craig, Michael Simeone, Luigi Marini, Mona Wong, Dave Bock, Sandeep Satheesan
E-mail: vkuhn@cinema.usc.edu

The screens that surround us on a daily basis impact our lives in ways we have yet to adequately understand, making it vital to conduct research in this area. The VAT (video analysis tableau), is a software workbench for video analysis, annotation, and visualization, using both current and experimental discovery methods using the Medici framework. Led by PI Virginia Kuhn of the USC School of Cinematic Arts (with Alan Craig, Michael Simeone, as co-PIs), and supported by the NSF's XSEDE program (extreme science and engineering discovery environment), the VAT project team includes experts in media studies, computational analysis, computer vision, database formation and humanistic epistemologies.

The VAT joins the emergent field of cultural analytics, an approach that deploys computer technologies to analyze the formal features of art and culture, making them available to interpretive methods. Moving image media is particularly ripe for computational analysis given its increasing ubiquity in contemporary culture. Indeed, digital video—whether recorded digitally or digitized from film—is a rapidly expanding form of contemporary cultural production, one made possible by the proliferation of personal recording technologies and hosting platforms like YouTube, Vimeo and the Internet Archive. In short, video is one of the most compelling "big data," issues of the current cultural moment; its formats are diverse, rapidly transmitted, and boundlessly large in number.

Yet despite its scale and importance, video remains a daunting object for sustained research, for obstacles that are technological, institutional and conceptual in nature. The VAT endeavors to fill existing gaps for asking cultural questions about video archives using computers, while also experimenting with transformative methods in video research and analysis. The longterm goal is to allow researchers to move with agility from textual description and collection management, to manual inspection, to automated analysis, to visualization of discrete films as well as whole collections.

Work with CWRC

Susan Brown, Jeffery Antoniuk, Michael Brundin, Mihaela Ilovan, John Simpson
E-mail: sbrown@uoguelph.ca, ilovan@ualberta.ca

The Canadian Writing Research Collaboratory (CWRC, pronounced "quirk") is building a research infrastructure which will encourage the creation, reading, enhancement, analysis and management (Brown et al.) of writings in or about Canada.

As the platform is fast approaching its official launch in the fall of 2015, its core affordances that were designed and developed to support complex, collaborative research are entering beta testing this spring, together with a redesigned interface.

The proposed digital demo falls somewhat outside the habitual mould for such demos, in that we are not aiming to introduce conference participants to a certain digital tool, though different tools will be available for demos and "test-drives". Conference attendants will have the opportunity to "sample" different areas of the CWRC research space in a redesigned environment that gives a makeover to the default Islandora interface and facilitates the interoperability of tools like the CWRC-Writer XML/RDF editor, the entity management system, the XQuery-based find and replace module, the workflow management tool and the PlotIt Timeline and Mapping visualization. The demo will also offer a glimpse into the functionality gained from bridging out into the landscape of well established digital tools like the ones in the Voyant Tools suite.

The main purpose of "Work with CWRC" would be to allow digital humanities scholars to see in action and briefly experience the type of complex, scholarly tasks that a collaborative platform such as CWRC affords and to complement, through a hands-on approach, the more theoretical discussion of "The Canadian Writing Research Collaboratory: An Experiment in Infrastructure"

References

Brown, Susan et al. "From CRUD to Cream: Imagining a Rich Scholarly Repository Interface." Scholarly and Research Communication 3.4 (2013): n. pag. src-online.ca. Web. 6 Dec. 2014.

Voyant Tools 2.0: The New, The Neat and the Gnarly

Stéfan Sinclair, Geoffrey Rockwell, Michael Sinatra, Marcello Vitali Rosati
E-mail: sgsinclair@gmail.com

Voyant Tools is a web-based, general-purpose text analysis and visualization environment. Its primary modus operandi is to operate on a single, immutable corpus that is defined at the moment of text ingestion (using uploaded files or specified URLs). Voyant Tools was designed to support larger corpora than its predecessors (HyperPo, Taporware), though the current version (1.0) struggles to deal with more than a couple of hundred of documents and/or dozens of megabytes of text.

A new version of Voyant Tools (2.0) will be demonstrated that addresses several of the major shortcomings and irritants of the current version. In addition to performance improvement throughout, the search and filtering functionality have been vastly enhanced and now support proximity and n-gram operations. Documents can be reordered or added to corpora on the fly, and there is now a lightweight access management layer that differentiates between full access, full-text access, and expressive/consumptive access.

Voyant Tools 2.0 is also the progressively ripening fruit of preliminary collaborations with large-scale publishers and content providers (Presses Universitaires de Montréal, Érudit.org, Early English Books Online, Digital Humanities Quarterly, etc.). The demonstration will focus in particular on the ability of Voyant to manage very large text collections (tens of thousands of texts or more) and to use search and analytic functionality to easily and quickly create subsets of documents, or corpora-on-the-fly. This capability fills a gaping hole between existing content repositories and specialized analytic platforms.

Voyant Tools 2.0 strives to be familiar for existing users while offering compelling improvements. It remains a work in progress, and this demonstration will provide insights into some of the new and most interesting features, while also recognizing known issues and essential weaknesses.

Advocacy by Design

Jeremy Boggs, Purdom Lindblad, Bethany Nowviskie, Ivey Glendon, Lisa Goff, Take Back the Archive Team
E-mail: jkb2b@virginia.edu, jpl8e@virginia.edu

Take Back the Archive <http://takeback.scholarslab.org> is a public history project by University of Virginia faculty, students, librarians, and archivists developed to preserve, visualize, and contextualize the history of sexual violence at the University. Imperative to the success of our project is the need to attend to the collection and presentation of our materials in ways that advocate for rape survivors and challenge a community of complacency and ignorance. More than simply collect and share this material, we plan to ensure that practices of collection and presentation used across the project empower contributors and users of the materials to dismantle passive acceptance of rape culture. In doing so, we hope our project will help establish a national model for college and university communities wishing to memorialize, historicize, confront, and end sexual violence. Our demonstration will elaborate on our purpose for Take Back the Archive and the design decisions that went into it, and open broader dialog about how researchers

can use and modify existing tools and systems in ways more appropriate to their scholarship and audiences.

Modernist Commons

Dean Irvine, Graham Jensen, Alan Stanley, Nigel Banks
E-mail: dean.irvine@dal.ca

This digital demonstration will feature the Modernist Commons (modernistcommons.ca)—a digital repository, modular editorial workbench, and collection of critical editions designed by Editing Modernism in Canada (EMiC) in partnership with Islandora (islandora.ca) and its open-source software-services company DiscoveryGarden (discoverygarden.org). EMiC has extended the functionality of Islandora Core to implement additional open-source, semantic-web tools from multiple digital-humanities projects (Canadian Writing Research Collaboratory, Shared Canvas, Australian Scholarly Critical Editions, and Interedition). With the workbench, editors can ingest images, audio, and video (along with metadata); produce transcriptions via multilingual optical character recognition processing; as well as edit and perform structural and semantic markup of transcriptions and images using a single graphical interface, which supports overlapping Text Encoding Initiative (TEI) and Resource Description Framework (RDF) standards. Editors can perform algorithmically generated collations of transcriptions, which can be manually corrected and visualized in in comparison views, alignment tables, and network graphs. The reading interface is configurable so that editors can assemble images, audio and video, critical apparatus, and visualizations.

Topic Words in Context (TWiC)

Jonathan Ilan Armoza, Stéfan Sinclair
E-mail: jonathan.armoza@mail.mcgill.ca

Topic modeling has given humanists the potential for newfound insight at the level of the corpus and the individual text. One of the difficulties, however, in interpreting a topic modeler's outputs has been making sense of why the words of topic-word lists are produced by a modeler run with particular parametrization by topic count, optimization interval, and hyperparameters. We can look to the math – an exacting reasoning of why this correlation might have been observed. But a significant goal of humanistic inquiry is to derive meaning from the methods we use to examine texts. "Topics" may relate to an overarching semantic line that runs throughout the corpus and its texts in varying proportion, or it may relate to a particular linguistic register: an authorial stylistic choice, for instance. Topic words thus removed from their source contexts and juxtaposed with one another paint a suggestive image for interpretation, but that suggestion may be as misleading as it is revealing. How can we alleviate this confusion? One answer is to read the texts, searching for topic words and their potential locations, but there is no need to guess. MALLET provides a state file which lists which topic is associated with which word across the entire corpus it has modeled. Still, the difficulty remains in leveraging this information in a compelling and provocative way to reassociate context-derived meaning back to topic-word lists.

To aid the process of interpreting topic models of texts I am developing a new visualization called "Topic Words in Context" (or "TWiC") under the supervision of Prof. Stéfan Sinclair. It enables users to see topic words as they are situated in their source texts right alongside the model's topic-word lists, utilizing several highlighting strategies including topic-associated coloring and font weight based on probabilistic weighting from the model. The visualization allows for simultaneous viewing of this information at multiple scales to enable topic-word contextualization and pattern-recognition within an indivudal text and from above with many texts in view. From on-high texts become icons which contain pixel-scale topic words in context of their sentences as well as coloring indicative of the document's topic composition. Texts are grouped by topic composition metric(s) in force-directed graph form, and can be clicked on to show in another panel a readable version of the individual text with its highlighted topic words. Topic visibility can also be toggled to aid in the recognition of those patterns. Additionally, texts may be prefigured in groupings so that once the modeler has run over the corpus, such grouping may be contrasted with the modeler-produced document grouping. As Ben Schmidt demonstrated with topic models of 19th century shipping logs, errors in a model because of inaccurate parametrization will not be apparent by looking to the top topic words output by MALLET. It is sometimes the outlier words that help to better identify potential meaning(s) of a topic. What would be required then is the "ability to visualize the entire model at once." Not only does TWiC re-engage users of topic modeling with the contexts from which topic words arise at multiple scales of the model, but it can also serve as a corrective for interpretation of misleading topic word lists.

Power Up Your DH with Compute Canada

John Edward Simpson
E-mail: john.simpson@ualberta.ca

Compute Canada is more than just large-scale batch processing. Stop by this demo to learn about services and support that can power up your DH research, including:

- OwnCloud. 50Gb of shareable, dropbox-like space available across multiple devices.
- Globus. Fast, secure, fire-and-forget data transfer for large files and large file sets.
- Vidyo. Simple to use, fault-tolerant, video conferencing.
- Cloud. Virtual-machine based development space that includes an outward-facing IP address.
- Data Integrity. Data storage and back-up systems provide stability and security options over your desktop.
- General Analyst/Consultant Support. Consultations regarding project architecture and resource needs.
- DH Support. Dedicated digital humanities expert available.
- Visualization Support. Dedicated 3D visualization expert available.
- Training. Training sessions covering core skills offered regularly and custom courses available on request.
- Portals. Hosting for specialized data and tools for entire research communities.
- Archival Storage. Robust storage solutions, including tape backup.
- Specialized Software. Over 250 software programs and packages already integrated with Compute Canada systems.
- Computation. Expandable power ranging from the equivalent of a second desktop to access to machines with thousands of cores, terabytes of RAM, and a variety of system architectures, including GPUs.

8:30am - 10:00am **Louis-Pasteur 155**	**Digital Demos 2: Digital Demonstrations**

TAPoR 3.0

Omar Isidro Rodriguez-Arenas, Kevin Schenk, Milena Radzikowska, Kamal Ranaweera, Stéfan Sinclair, Mark McKellar, Geoffrey Rockwell
E-mail: orodrigu@ualberta.ca

The challenge of recognizing software contributions to the digital humanities has bedeviled the field for some time. Computers and the Humanities included software reviews from its first issue in 1966, but software reviews after the fact don't have the same value as peer-review for publication. In 2003 a panel chaired by Sinclair reviewed ways of reviewing software and called for research tool review processes, but no one seems to want to run them. In 2006 the report of a MLA Task Force on Evaluating Scholarship for Tenure and Promotion recommended that "institutions should recognize the legitimacy of scholarship produced in new media, whether by individuals or in collaboration, and create procedures for evaluating these forms of scholarship." (Report 2006, p. 11) The problem has always been on of developing a culture of informed discussion around tools and methods that goes beyond one or two review processes. (Rockwell 2011, p. 166) To that end TAPoR (Text Analysis Portal for Research) has been maintaining an online database of text tools (tapor.ca) that provides a stable place for the discovery of and comment on tools, including historical tools. In this digital demonstration we will show the next version of TAPoR (version3.0) that includes two new features that we hope will support community conversations around tools:

1. **Curated lists** that are maintained by Associate Editors of tools that are suitable for a particular purpose or audience.

2. **Related tools** are suggestions for other tools that a browser might be interested in and we are refining the statistics that we use to generate these.

Curated Lists

When analyzing text with automated methods, it is common to use multiple tools in order to read that text from different perspectives and extract the most information possible. To address this type of workflow TAPoR 3.0 will provide environment that allows people to curate lists of text analysis applications that meet meet some criteria. This simplifies the discovery of new tools since they are no longer presented as distinct entities but are shown as part of a collection with comments from an editor. This approach is on that other web sites dedicated to software review take like Merlot (www.merlot.org). It is allows us to avoid developing a full review process which would duplicate what venues like DH Commons (dhcommons.org) are developing while still being useful. In the digital demonstration we will show a first set of list developed with partners like Text Mining the Novel (novel-tm.ca) and show how lists are edited.

Related Tools

As stated before, one of the goals on TAPoR is to make it easier to find new tools; trying to fulfill this goal, usage statistics have become very valuable. TAPoR 3.0 has more granular metrics that allow us to analyze user behaviour within the system in finer detail. With the information obtained from the more granular usage metrics we improved the related tool suggestion system and are also now showing suggestions to editors (list curators) and other entering tools. This system can work now as an aid to administrators/editors, when linking related tools, by providing a list of possible related tools. This list is generated by comparing the tags, categories, usage statistics, and the rest of the metadata available for each tool. An expert system is used to generate this list comparing the tools as vectors in multidimensional space.

References

"Report of the MLA Task Force on Evaluating Scholarship for Tenure and Promotion". Modern Language Association. MLA, 2006. <http://www.mla.org/pdf/taskforcereport0608.pdf>

Rockwell, Geoffrey. (2011) "On the Evaluation of Digital Media as Scholarship." Profession. MLA. Pages 152–168.

Sinclair, Stéfan, John Bradley, Stephen Ramsay, Geoffrey Rockwell, and Ray Siemens. (2003) "Peer Review of Humanities Computing Software." Panel chaired by Sinclair at the ACH/ALLC 2003, University of Georgia, Athens, GA, USA.

The Big Red Book: A Demonstration of Serendipitous Tool for Augmenting Knowledge (STAK)

Brian Greenspan, Jiayu Li, **Kim Martin**, David Mould, **Anabel Quan-Haase**
E-mail: brian_greenspan@carleton.ca, kmart5@uwo.ca, aquan@uwo.ca

Serendipity is the word most often used to describe that moment when a scholar unexpectedly stumbles across a useful document. Humanities scholars often describe serendipitous experiences as central to their research, particularly when they occur in a library or archive. STAK is a web-based tool we are designing for mobile devices to augment the user's experience of browsing physical collections by creating opportunities for the serendipitous discovery of information resources relevant to a project or interest not currently in the foreground of her attention.

"Everything on Paper Will Be Used Against Me: Quantifying Kissinger"

Micki Kaufman
E-mail: mkaufman@gc.cuny.edu

Perhaps more than any other former Secretary of State or National Security Advisor, the public celebrity of Henry Kissinger was (and remains) based upon an array of paradoxes. As compelling as foreign policy insights, the internal contradictions within Kissingers political and personal (and ostensibly, moral) psychology have been the subject of much of the recent historical study of Kissingerology. Traditional scholars have faced significant difficulty in their efforts to understand what appear at first to be apparently incompatible motives and behavior, attempting to reconcile facets of a public figure who appears to embody a host of mutually exclusive dichotomies.

Needless to say, disagreements about the man and his motives persist just as new interpretations based on text analysis techniques like word frequency analysis, topic modeling, sentiment analysis and corpus linguistics, are becoming possible. Combined with novel forms of knowledge production like distant reading, and made comprehensible using new, interdisciplinary tools and techniques, including the language of visual design, new examinations into the Kissinger/Nixon White House using new methods provide not merely new answers to existing questions they allow entirely new questions to be asked.

As detailed on the project's web site (http://www.quantifyingkissinger.com) my work involves the application of a host of quantitative text analysis methods and data visualization designs to an analysis of the Digital National Security Archive (DNSA)s Kissinger Collection. The archive, comprising approximately 17,600 meeting memoranda (memcons) and teleconference transcripts (telcons) detailing the former US National Security Advisor and Secretary of States correspondence during the period 1969- 1977, has served as a basis for combining political/international relations history with the fields of linguistics and visual design.

As one example, the projects combination of the computational approach employed herein with an emphasis on emotional history has illuminated new avenues of inquiry that combine text analysis with more subjective questions around interpersonal dynamics and individual psychology. In addition to analyzing Kissingers social relationships, the analysis has studies aspects of his, and the administrations, behavior around matters of secrecy and violence in one example uncovering the highly selective, event-driven use of the telcons (which Kissinger expected to remain private) versus a more generic and non-committal use of the memcons for discussing matters regarding bombing and [Cc]ambodia.

Even more tantalizing than detailing existing knowledge about Former Secretary of State Henry Kissinger, the project has begun to surface deeper understandings and questions about how this new kind of distant knowledge is formed and the ramifications for historical analysis in general. By combining computational and emotional history, a number of interesting results have emerged about the man as much as the geopolitical focus of the administration he served - from his preference for outside channels to his proclivities for seduction, secrecy, humor and violence. The application of computational techniques to the study of twentieth century US diplomatic history has thus far generated useful finding aids for researchers, provided essential testing grounds for new historical methodologies, and prompted new interpretations and questions about Kissinger and the individuals, organizations and events of his tenure.

The Open Syllabus Project - Using Overview to search, analyze, and visualize a collection of 2 million college syllabi

David McClure, Joe Karaganis, Dennis Tenen, Jonathan Stray, Jonas Karlsson, Adam Hooper
E-mail: davidwilliammcclure@gmail.com

The Open Syllabus Project is an effort to build an online database of university course syllabi that can be used as lens to better understand intellectual trends, institutional histories, and disciplinary boundaries in higher education. Which texts are being taught in college and universities? Where are they being taught, and how often? How have those trends changed over time? What can be learned by extracting citation networks that trace out the connections among syllabi that include similar texts? Which texts are gaining in the intellectual mindshare of the academy, and which are losing ground? This information could lay the groundwork for new applications in research, teaching, and administration - it provides a front-row view onto the modern process of canon formation and disciplinary change, points the way to new types of metrics for evaluating the impact of scholars and publishers, and gives researchers the tools to compare curriculums across disciplines, regions, and periods.

With support from the Alfred P. Sloan foundation, the Open Syllabus Project is being built with a tool called Overview. Incubated at the Associated Press, Overview is a platform that makes it possible for scholars and journalists to search, analyze, and visualize large collections of text documents. Originally designed to make it easier for journalists to make sense of large dumps of documents like the Wikileaks cables, Overview provides the technical infrastructure to manage millions of documents and a modular plugin architecture that makes it possible to build custom tools for discovery and analysis - full-text search interfaces, interactive maps, citation network navigators, and more. This demonstration would present preliminary results from the process of analyzing and visualizing the syllabus corpus, while at the same time providing a general-purpose introduction to the capabilities of Overview for a wide range of text analysis and exploration tasks.

Game of Writing (GWrit)

Mark Pearse McKellar, Geoffrey Rockwell, Kamal Ranaweera, Aiden In, Melania Ru'Aini, Roger Graves, Heather Graves, Omar Rodriguez-Arenas
E-mail: mpm1@ualberta.ca

Gamification has gained popularity as a tool used to promote positive behaviour (as in the case of chorewars.com) or teach concepts (such as Rocket Fuel Productions Trajectory software). The Game of Writing (GWrit) was created in order to see if such gamification elements can be combined with analytics and used to promote writing and collaboration in an educational environment. Warnock (2013) found that modality of instruction (online, in-person, hybrid/blended) does not affect student outcomes for writing development, but he did not consider gamification or online gaming studies of the teaching of writing.

GWrit is a web based application in which users are able to log into a website in order to complete assignments, customize an avatar, and earn rewards, and points. Gamification has received a lot of attention from designers like Jane McGonigal (2011). Others like Ian Bogost (2011) have justifiably criticized it as a way of domesticating videogames for business. Others (Kapp 2013; Paharia 2013; Burke 2014) argue that gamification is useful, particularly in educational contexts. We argue that it can been seen as a mode of presenting information back to user that can be set against human feedback. Information about the completion of an assignment can be presented as a fact or as a badge; GWrit has been designed to allow us to experiment with different combinations of analytics and gamification.

Our Digital Demonstration will do the following:
- Demonstrate both student and instructor workflows.
- Discuss the scaling issues experienced.
- Describe the new architecture developed to support large writing classes.

GWrit has been developed at the University of Alberta over the past two years and has already been tested over two iterations of a writing course (WRS 102). The first iteration had 31 students, and ran over a period of 7 weeks, providing useful information on user interactions as well as an opportunity to remove minor bugs in the system. The second iteration was much larger, with 163 students over 15 weeks. Many of the problems from the first test had been resolved, but we experienced and dealt with significant load and performance issues in the second run. The issues were addressed with a software architectural revision. These tests have been beneficial in molding a system which is fun, easy, and useful for both students and instructors.

Our demonstration will walk through the basic workflows for both students and instructors. We will show the process of completing an assignment and how a student can influence the writings of others through comments; how the reward system works and how these rewards are displayed to the user. Finally, we will demonstrate how the instructor can interact with students, oversee student collaborations, track progress, leave and promote comments, and create their own rewards for the class.

Does gamification work in the teaching of writing? Comments and analytics tell us that the progress bar adds some value, since it was always in the student's field of view; students commented on their eagerness to fill it up completely. Badges and scores were added in the second iteration and their effectiveness is currently being reviewed.

References

Bogost, I. (2011). "Gamification is Bullshit." Blog essay at <http://bogost.com/writing/blog/gamification_is_bullshit/>

Burke, B. (2014). Gamify: How gamification motivates people to do extraordinary things. Brookline, MA: Bibliomotion.

Kapp, K. M. (2012). The Gamification of Learning and Instruction: game-based methods and strategies for training and education. San Francisco: Wiley.

McGonigal, J. (2011). Reality is Broken: Why Games Make Us Better and How They Can Change the World. New York: Penguin.

Paharia, R. (2013). Loyalty 3.0: How to Revolutionize Customer and Employee Engagement with Big Data and Gamification. New York: McGraw-Hill.

Warnock, Scott. (April 2013). Studies Comparing Outcomes Among Onsite, Hybrid, and Fully-Online Writing Courses, WPA-CompPile Research Bibliographies, No. 21. WPA-CompPile Research Bibliographies. http://comppile.org/wpa/bibliographies/Bib21/Warnock.pdf. May 5, 2014.

Shout-Out for the Humanities": A student contest sponsored by 4Humanities

Kim Martin
E-mail: kmart5@uwo.ca

The 4Humanities Collective is rolling out a new competition open to both graduate and undergraduate students. They are seeking answers (in a variety of media) to such questions as: Why is studying the humanities–e.g., history, literature, languages, philosophy, art history, media history, and culture–important to you? To society? How would you convince your parents, an employer, a politician, or others that there is value in learning the humanities?

| 10:00am - 10:15am | **Coffee Break 5: Coffee Break** |
| Louis Pasteur 154/155 | |

| 10:15am - 11:45am | **7a: Text Patterns** |
| Colonel By C03 | |

Making Sense of Abundance: Opportunity and Challenges Across Three Web Archive Case Studies

Ian Milligan
E-mail: i2milligan@uwaterloo.ca

The sheer amount of social, cultural, and political information that is generated and, crucially, preserved every day presents new exciting opportunities to historians. A large amount of this information is being contained within web archives, which contain billions of web pages. Scholars broaching topics dating back to the mid-1990s will find their projects enhanced by web data – military historians can use forum posts by soldiers, social historians can track aspects of everyday life through blogs and comments, political historians can study changing sentiment, tropes, and link structures, and economic historians can explore the rise and fall of businesses webpages. Yet this tremendous opportunity is mitigated to some degree by the sheer challenge of dealing with all that data: we have more information than ever before, but the scale is overwhelming.

My paper draws on three distinct case studies that I have been pursuing to explore this problem of abundance: the Internet Archive's March-December 2011 Wide Web Scrape (WARC files), the 2009 GeoCities end-of-life torrent (a

wget-compiled collection of mirrored websites), and the 2005-Present Archive-It collections of Canadian political parties, unions, and organizations (WAT files, which contain derivative data). For each archive, I discuss the usage, technical, and ethical challenges that such collections present for historians. While similar in content challenges, they provide different challenges in terms of process.

I highlight what we can learn, notably the activities of everyday people and their connections to each other on the Web, but also note the challenges: problems of too much data, processing time, and the difficulties in applying cutting-edge natural language processing methods in a humanities context. For each collection, I demonstrate various techniques to: cluster and identify recurring concepts (and how you adapt it to each collection type), extract and visualize significant images, convert to searchable plain text, and other forms of basic textual analysis and computer vision. These tasks all feed back into my central research questions concerning the fruitful extraction of research data. Yet all present different ethical dimensions: GeoCities users had different expectations of privacy and far less access to the always-critical robots.txt file which governs web archiving, as opposed to official websites. While we can meaningfully manipulate and extract this data today, we also need to tackle questions around how we can appropriately do so – or whether we even should.

This is significant because each of these formats calls on different skills: WARCs call on specialized web archiving tools, WATs lend themselves more to a traditional machine-reading approach, whereas the wget collections speak to a body of research around the "live web." This presentation will be of value to those who want to use web archives, in their various forms, as well as to those who are creating them – often on limited budgets. My paper ends on a note of hope, highlighting what we can learn, while recognizing that the challenges that lie ahead can only be solved through interdisciplinary connection and cooperation.

Police Officer-Involved Homicide Database Project

Brittany Suzanne Paris, **Irene Pasquetto**, **Jennifer Pierre**, Ashley Sands, Morgan Currie
E-mail: brittany.paris@gmail.com, irenepasquetto87@gmail.com, jp639@cornell.edu

Recent events in the case of Michael Brown, an 18-year-old unarmed African American man killed by a police officer in Ferguson, Missouri, have drawn attention to a gap in publicly available information regarding the number of Americans killed by police each year.

The FBI's Uniform Crime Reporting Program (UCR) collects data from the more than 17,000 law enforcement agencies to provide statistics about crime and law enforcement in the United States. In the seven-year period from 2008-2014, there were nearly 400 "justified" police homicides according to the Supplementary Homicide Report (SHR), another FBI report filed separately from the UCR. [1][2] The number of justified police homicides may seem alarming in itself; however, perhaps more alarming is the reported number is likely low. The UCR and the SHR rely on voluntary involvement of state and local police agencies. Just over one third of law enforcement agencies contribute to the FBI's justifiable homicide database. Even fewer agencies report to the more detailed SHR. Thus, the largest databases of police homicides in the U.S. are quite incomplete and inconsistent. [3][4]

Our project explores un- and under-reported incidents of law enforcement-involved homicides, both justified and unjustified, through a document analysis of extant federal and local databases with information pertaining to officer-involved homicides, mining and analysis of social media data in addition to participatory action research methods, primarily hosting hackathons, to fill gaps in existing government and local databases. We have chosen Los Angeles County as the first community to study. Los Angeles (LA) is a special example in that it is considered to be the metropolitan area with the best records on officer-involved homicides. Even in communities considered to have the "best records", we see that there are still important gaps to be explored.

This mixed method approach attempts to bridge and adequately address the variety of dimensions included as part of our project. Ultimately we seek to make this information not just useful to the public via hackathons, but also available to the public for further endeavors, whether they might relate to data journalism, the efforts of community groups interested in projects promoting social justice or to the interests of common citizens with questions about their law enforcement officials. In coming months, we will build a website that houses all of the information we gather and learn from this hackathon, as well make available tools, methods and models we used to derive this information so that other communities might use it to gain similar localized knowledge about police officer-involved homicides.

[1] "Easy Access to FBI's SHR," National Center for Juvenile Justice, August 1, 2014: September 30, 2014, http://ojjdp.gov/ojstatbb/ezashr/.

[2] "Uniform Crime Reports," Federal Bureau of Investigation, September 30, 2014, http://www.fbi.gov/about-us/cjis/ucr/ucr.

[3] Enten, Harry J, "Justifiable Homicides: the story behind the data and our findings," The Guardian, April 5, 2012: September 30, 2014, http://www.theguardian.com/world/2012/apr/05/jusifiable-homicides-guns-data-analysis.

[4] Bump, Phillip, "The Number of Justifiable Police Homicides Has Changed Since the 1990s," The Washington Post, August 15, 2014: September 30, 2014, http://www.washingtonpost.com/blogs/the-fix/wp/2014/08/15/how-the-number-of-justified-police-homicides-has-changed-since-the-1990s/.

The HTRC Extracted Features Dataset

Peter Organisciak, Loretta Auvil, Sayan Bhattacharyya, J. Stephen Downie
E-mail: organis2@illinois.edu

We introduce the Extracted Features (EF) dataset, consisting of textual features derived from over four million volumes in the Hathitrust Digital Library (HTDL). The EF dataset presents a familiar but non-consumptive way to interact with a large text corpus within intellectual property access restraints. We also present a toolset for working with the data.

10:15am - 11:45am **Louis-Pasteur 154**	**7b: Panel: Les éditions critiques en contexte numérique**

Les éditions critiques en contexte numérique

Michael E. Sinatra, **Simon Harel**, **Ollivier Dyens**, **Esma Aimeur**, **Jason Camlot**, **Jean-Claude Guédon**, **Isabelle St-Amand**
E-mail: michael.eberle.sinatra@umontreal.ca, simon.harel@umontreal.ca, ollivier.dyens@mcgill.ca, aimeur@iro.umontreal.ca, jason.camlot@concordia.ca, jean.claude.guedon@umontreal.ca, isstamand@gmail.com

Voir notre PDF ci-joint

10:15am - 11:45am **Louis-Pasteur 155**	**7c: Global Horizons**

10 Million Spanish Books: A Dynamic Networks Approach to Understanding the Diffusion of Spanish Texts

David Michael Brown, Juan Luis Suarez, Adriana Soto-Corominas
E-mail: dbrow52@uwo.ca

Overview

This paper takes a data driven approach to the history of publication in Spanish. Using OCLC metadata from over 10 million books, we investigate the diffusion of texts them from their point of production the their place of holding. To best understand this process, we have chosen to model it as an evolving migration network (Schich et al., 2014). By filtering the data model by time period, we generate a series of snapshots that represent the evolution of the transatlantic network formed by the book trade, starting with the earliest days of the Spanish empire through the end

of the 20th century. Using formal models borrowed from graph theory combined with historical analysis, we can begin to determine the factors that motivate the movement of cultural objects throughout their geotemporal landscapes.

Hypothesis

We predict that a text's trajectory from production to library holding varies by time and location of production. As shown in studies relating to colonial art (Suárez et al., 2013), we expect to see a massive exportation of colonial American texts away from their centers of production to libraries in North America and Western Europe. However, as we progress towards the 19th and 20th centuries, this exportation will decrease resulting in a lower average distance travelled overall, especially in Latin American countries.

Methodology

The initial metadata was processed using Open Refine and the Getty thesaurus to standardize place, text, and author names and was stored in a Neo4j graph database. Using a parallel web crawling system developed in IPython, the text metadata is used to query the WorldCat search API to determine current text holding locations. Due to the large quantity of metadata used in this project—over 10 million entries—and the limitations of the WorldCat Search API—50,000 daily requests—this is an ongoing process that will be completed in 2015. Initial processing suggests that the final graph projection will result in over 60 million edges. To facilitate analysis, data is chunked into periods of 25 years corresponding to the year of publication of each text in the dataset; however, preliminary results were too large to be processed in memory. In order to overcome this limitation, we leverage an architecture similar to Mazerunner (Bastani, 2014), employing Apache Spark (Zaharia et al., 2010) to calculate network statistics. The results are then analyzed within the framework of dynamic networks, taking into account topological factors, contextual homophily, models of diffusion, and relevant historical factors.

Results

The preliminary results demonstrate the movement of texts printed in a colonial environment towards Western European and North American cultural centers; however, as various Latin American countries were formed in the 19th century, there is a decided shift in the patterns of production and movement. Latin American production spikes during the 19th century, and while exportation continues to be a huge factor in their diffusion, increasingly these texts remain in libraries near their point of production.

RedHD: Open, Collective, and Multilingual Work Dynamics

Ernesto Priani, Isabel Galina, **Elika Ortega**, Miriam Peña
E-mail: epriani@gmail.com, eortegag@uwo.ca

The product of a series of workshops that took place between 2010 and 2011, RedHD has become a landmark organization in the global Digital Humanities. RedHD has grown from a clearly local initiative and become greatly successful at establishing links between emerging Spanish speaking digital humanists around the world and, through them, facilitating connections with other linguistic and geographic DH communities. By doing this, RedHD has transformed the landscape of the field in subtle but important ways.

The organization has its origin on the impulse of a number of scholars mainly located in Mexico who had found themselves somewhat isolated from a larger body of faculty reluctant to adopt computational methodologies and approaches in their research and/or teaching. The 2010-2011 workshops fostered a budding sense of community based on common issues and obstacles to carry out DH projects and practices. This community spread locally and internationally in a worth-of-mouth way thanks to the connections each of the members brought in.

In just four years since its beginnings, RedHD is now made up of a base of researchers located primarily in Mexico, but has excelled at maintaining productive collaborations with other scholars based in Latin America, North America and Europe. This has made RedHD a truly international organization that is not limited by a geographic territory nor a linguistic one. Instead, it has fostered open, multilingual, and collective work dynamics. RedHD's international and multilingual profile has been determinant to aid the bearing that RedHD currently enjoys in the global landscape of Digital Humanities, as well as to establish the impact and visibility of both its international initiatives and its inner projects.

In this paper we will showcase the results of the analyses of a handful of events organized by RedHD--DíaHD 2013, and 2014 and the 1st and 2nd Encuentro de Humanistas Digitales that took place in Mexico city in 2012 and 2014. These results show the progressive formation of the network and its zones of impact. Furthermore, we will explore the spread and impact of a handful of collaborative projects like *Checklist: Herramienta de evaluación para proyectos de Humanidades Digitales* , *MapaHD* , and *RedHD in translation* . These examples are not only illustrative of our history, profile and work dynamics, but also serve as a locus of reflection upon the growth of the network, its needs, and its mission as an organization in the process of consolidation both locally and internationally.

Remaking the Atlas, Unmaking the World: Towards a Cultural Atlas of Global Blackness

Roopika Risam
E-mail: rrisam@gmail.com

In "Hello Worlds," Matthew Kirschenbaum speaks to the role of digital humanities in shaping how we look at the world. He proposes that programming may be viewed as an act of world-making that requires articulating the observable rules and characteristics of an environment to create a system. Still other digital humanities methods engage in world-making in more literal ways, such as the use of maps or geographic information systems (GIS) to visualize data across time and space. Like the worlds instantiated by the colonial project, the worlds created within digital humanities projects do not exist independently of the values or assumptions that shape the worlds we inhabit. Echoing the arguments of postcolonial scholars who have linked world-making and ideology, Kirschenbaum reminds us that virtual worlds are not only empirical but also ideological: "They embody their authors' biases, blind spots, ideologies, prejudices, and opinions." [1] Indeed, the knowledge of the world produced within such work - like the colonial project that remade the world as we know it - is imbricated in a matrix of culture, politics, and economics, among others. In the same vein, critical cartography scholarship has suggested that affordances of GIS provide the possibility of reimagining the process of mapping beyond its colonialist history and of using maps as methodologies towards emancipatory ends. [2]

Considering world-making within digital humanities through a postcolonial lens, this talk examines my work on the prototype for *A Cultural Atlas of Global Blackness,* an interactive database and digital map that traces representations of blackness across geography and temporality. Influenced by the work of the Electronic Cultural Atlas Initiative, this project emerges from a research question articulated in my previous work: how has black radical thought traveled throughout the postcolonial world? My talk will situate this project within the work of the Electronic Cultural Atlas Initiative, with attention to the methodological difficulties posed by the goal of mapping by concept - particularly a floating signifier like "blackness" - rather than by place. Additionally, I will discuss considerations of platform, database structure, and metadata that informed project development. In doing so, this talk will focus on project design to explore how mapping blackness on a global scale unmakes the world. I suggest that the task destabilizes the map itself and produces new knowledge for how we understand cultural transmission for the African diaspora and beyond.

[1] Kirschenbaum, Matthew. "Hello Worlds," *The Chronicle of Higher Education,* January 23, 2009. Accessed December 8, 2014. http://chronicle.com/article/Hello-Worlds/5476.

[2] Advocated by geography scholars such as J.B. Harley, Denis Wood, and Jeremy Crampton, among others, critical cartography blends mapping praxis with critical theory. Critical cartography is grounded in the belief that maps are sites of knowledge and power, not simply visualizing knowledge but producing it.

10:15am - 11:45am **Louis-Pasteur 285**	**7d: #gamergate, Feminism and Digitally Constituted Communities (CSDH/SCHN & CGSA Joint Session)**

11:45am - 1:15pm	Lunch Break 3: Lunch Break

| 1:15pm - 2:45pm
Colonel By C03	**8a: Early Modern Considerations**

Preliminaries Project: The Social Network of the Duke of Lerma

Juan Luis Suarez, David Michael Brown, Javier de la Rosa, Adriana Soto-Corominas
E-mail: jsuarez@uwo.ca, dbrow52@uwo.ca

The Preliminaries Project aims to understand the networks of people, places, and institutions involved in the production of Spanish Golden Age literature. Using information gleaned from the preliminaries sections of a collection of early 17th century books, we constructed a model that permits the statistical analysis and visualization of this Early Modern cultural network. This paper focuses specifically on the period between 1598 and 1618, during the administration of the valido Francisco de Sandoval y Rojas, Duke of Lerma I (1553-1625) under the reign of Phillip III (1578–1621) of the Habsburg dynasty. Using techniques from social network analysis and literary studies, we discuss the influence of individuals and community structure on the publication and diffusion of cultural information in the Spanish empire. Furthermore, we historically contextualize this data to illuminate the societal and political forces that result in the trends and patterns present in our dataset.

DREaM: Distant Reading Early Modernity

Stephen Wittek, Stefan Sinclair, Matthew Milner
E-mail: stephen.wittek@mcgill.ca

Our proposed paper will provide an overview of the theory and methodology driving the creation of Distant Reading Early Modernity (DREaM), a digital humanities project that has made a massive corpus of early modern texts amenable for use with marco-scale analytical tools. Key focus areas include the technical challenges deriving from non-standardized spelling, the argument for our approach to the early modern archive, and the potential benefit to early modern scholarship of distant reading techniques.

eMOP's Printers and Publishers: Toward Crafting an Early Modern Print Database

Matthew Christy, Elizabeth Grumbach, Laura Mandell
E-mail: mchristy@tamu.edu

The Early-Modern OCR Project (eMOP), currently underway at the Initiative for Digital Humanities, Media, and Culture (IDHMC) at Texas A&M University, is a Mellon-funded project tasked with developing open-source tools and techniques to improve Optical Character Recognition (OCR) outcomes for early modern printed documents. The basic premise of eMOP is to 1) use book history to identify the typefaces represented in the collections and the printers that used them; 2) train open source OCR engines on those typefaces; and 3) OCR early modern document page images using an engine trained on the typefaces specific to those documents. As our dataset, eMOP is using the 45 million page images that comprise the Eighteenth Century Collections Online (ECCO) and Early English Books Online (EEBO) proprietary collections.

After two years of testing and development, we have shown that training Tesseract, Google's open-source OCR engine, to recognize early modern typefaces improves character recognition in our corpus. However, applying these techniques more precisely is not yet possible. We simply don't have a way, on a large scale, to identify documents based on their printer and date of publication, and then use that information to further identify the typeface or typefaces likely used in the printing of that document. To apply accurate OCR typeface training for early modern documents in this way would require a database containing early modern printers' names correlated with the typefaces they used and the documents that they printed in these typefaces. Currently such a database does not exist. But eMOP is taking steps towards this goal.

A stated deliverable of eMOP is a Publisher Imprint Database (PubImprintDB) culled from the ESTC-supplied imprint lines of the works available in EEBO and ECCO. We have developed algorithms that will parse these natural language imprint lines and identify information relative to:

- who the document was Printed By,
- who the document was Printed For,
- who was the Seller of the document,
- the Location of the print shop (i.e. "near St. Dunstans Church in fleet-street"), and
- the Place of publication (i.e. London).

By the time of the CSDH/SCHN & ACH Conference in June 2015, the PubImprintDB will be available online for scholars to use and edit, as necessary.

We see the PubImprintDB as the first step towards creating a more comprehensive database of early modern printers, publishers, typefaces, and works. Such a database would not only allow for more accurate OCR'ing of early modern page images based on the typefaces they use, but could serve as a research tool for the examination of printer networks, the geographic spread of typefaces, and more.

We wish to present on the PubImprintDB at CSDH/SCHN & ACH in June in order to introduce our new PubImprintDB, demonstrate some of its potential research uses, and seek collaborators for further development of a larger database or set of databases related to early modern printing in an effort to aid scholarly research and continue development of a valuable scholarly resource on early modern printed documents.

| 1:15pm - 2:45pm
Louis-Pasteur 154	**8b: Panel: Transactive Memory Keepers Database**

The Kanata Indigenous Performance, New and Digital Media Art Project and the Transactive Memory Keepers Database

Julie Nagam, Heather Igloliorte, Carla Taunton
E-mail: jnagam@faculty.ocadu.ca, heather.igloliorte@concordia.ca, ctaunton@nscad.ca

This project is the first major research study to theorize and map Indigenous theory as it has emerged in Canada within the fields of Visual Studies, Native Studies, New and Digital Media and Performance Studies. With this SSHRC funded project, we will investigate creative practices, aesthetics, performance, new and digital media. The research project represents a two-year research partnership between three Canadian universities (OCAD U, NASCAD U, and Concordia University). Our team will trace Indigenous practices and methodologies in the area of performance, digital and new media art by documenting performance-oriented scholarship and creative practices carried out in Canada. We have drawn on existing archives at V-tape, ImagineNATIVE, Obx Laboratory for Experimental Media, Isuma, Arnait Video, Unikaat and others. The main focus of this talk will be showcasing the online database, this will be the first interactive website that will store and allow people access to Indigenous performance, digital and new media art in Canada.

| 1:15pm - 2:45pm
Louis-Pasteur 155	**8c: Evolving Pedagogies**

An Agile, Robust Partnership Model for DH Training

Raymond George Siemens
E-mail: siemens@uvic.ca

This paper discusses work of our community toward the DHSI-based Graduate Certificate in Digital Humanities, as well as plans emerging toward an enlarged partnership structure that can include participant institutions across the country and continent, and around the world.

The DHSI-based Graduate Certificate in Digital Humanities began accepting applications in September 2014 for intake May 2015. The certificate program can be taken in conjunction with other graduate degrees at UVic, elsewhere, or on its own, allowing the hundreds of those who come each summer to DHSI, and its several thousand alumni, to receive graduate university credit for the work they've done at DHSI and in related curricula.

The certificate curriculum blends computational methods and theories with humanities research and pedagogy, specifically addressing the demand for graduates who are proficient in computing and will contribute to growth areas such as information management, multimedia communication, social computing, game design, digital preservation, and data visualization; at the same time, the program also prepares graduates for active participation in the digital dimensions of humanities research, including prototyping, encoding, and data processing. Graduates of the program are well positioned for project coordination and leadership roles in emerging digital, mobile, and database-driven projects, serving as informed liaisons between programmers, technical writers, new media artists, researchers, and user communities; and well beyond.

To complete the certificate, typically participants will attend the DHSI, held annually in June, to take credit courses. There is no residency requirement per se, but the requirement that three of the courses must be taken through the DHSI or UVic necessitates at least three weeks of attendance on campus. These may be accumulated one, two, or three in any given year, depending on the offerings of the DHSI. Two of five courses may be taken at affiliated institutions.

Works

Digital Humanities Summer Institute. http://dhsi.org/index.php

Graduate Certificate in Digital Humanities. http://english.uvic.ca/graduate/digital_humanities.html

---

First Year English as a DH Course

M. Nicholas van Orden
E-mail: vanorden@ualberta.ca

Discussions about digital humanities pedagogy are many, varied, and have featured contributions from dh's most prominent figures--Johanna Drucker, Adeline Koh, Brian Croxall, Matt Kirschenbaum, and Miriam Posner have all written famously or recently on dh in the classroom. These stimulating discussions generally focus on the differences between upper-level dh-specific courses and introductory-level survey courses, and they often feature advice on how instructors with minimal resources might integrate dh tools and projects into their curricula . As a PhD student designing and teaching my first dh-focused first-year English course, I am well situated to comment on and contribute to these already rich discussions.

In my paper I will report on the design, functioning, and results of the first-year English and Film Studies course I am teaching in the Winter 2015 semester at the University of Alberta. EFS 121 is titled "Literature in Historical Perspective" by the English and Film Studies department and it is generally taught as a scaled-down survey course (covering a wide-range of texts and time-periods and requiring students to produce reading reflections, presentations, and research essays). Instead of following the survey model, my course focuses entirely on recently published Edmonton-based texts and authors, connects with a range of community partners, and requires students working in small groups throughout the semester to complete dh projects.

The dh tools the class will use must be free and online to ensure that all students have access through their own devices or library computers; the tools must also be simple to use as I cannot devote class time to technical questions. While these requirements exclude many popular dh tools and restrict the projects students can undertake, tools like Voyant, CartoDB, Omeka, Layar, and JuxtaCommons provide a host of interesting possibilities for examining literary data. One of the main challenges in designing the course has been collecting texts for the students to analyze--we need material that fits within the reading and writing requirements of the course, but also material that lends itself to dh analysis. Some of the tools, like JuxtaCommons, require particular types of input to produce interesting results. To meet this need for material we have partnered with various individuals and community groups, including, among others, the English and Film Studies' Writer in Residence (Minister Faust), the Edmonton Public Library's Writer in Residence (Jason Lee Norman), the Edmonton Arts Council, the Alberta Writer's Guild, the Aspen Foundation For Labour Education, and the Seniors Association of Greater Edmonton (SAGE).

In my paper I will outline the tools, texts, and partners that have shaped the course, and I will provide a report on the progress of the students' group projects. I will describe the assignment structure of the course and explain my rationale for deviating from the traditional essay-focused survey course. Throughout the paper I will focus on the feasability of incorporating a range of dh tools into a first-year course and will describe the various problems that the students and I encountered throughout the semester.

The Pedagogical Hermeneutics of Humanities 100: or, How to Teach DH in a Liberal Arts Context (with Archival Materials)

Diane Katherine Jakacki, **Katherine Mary Faull**
E-mail: dkj004@bucknell.edu, faull@bucknell.edu

This paper will focus on the teaching experience of Drs. Faull and Jakacki to show how the planning, design, and execution of a new project-based course in Fall 2014 introduced students to the world of DH through the use of selected digital tools and methods of analysis. This course was designed specifically for first- and second-year students, working with archival materials in order to encourage the development of digital habits of mind at the earliest phases of their liberal arts curricular experience. The decision to root the course in a multi-faceted analysis of archival materials provided the rare chance for students to engage in the research process typical for a humanities scholar: namely, the discovery of artifacts, the formulation of research questions, followed by the analysis and synthesis of findings culminating in the publication of initial findings in a digital medium.

| 1:15pm - 2:45pm | **8d: Text Viz** |
| Louis-Pasteur 285 | |

Textplot: Visualizing the thematic structure of texts and corpora

David William McClure
E-mail: davidwilliammcclure@gmail.com

Textplot is a program that converts a text into a network of terms that teases out the underlying topic structure of the document. *War and Peace* becomes a massive triangle - war on the left, peace on the right, Tolstoy's essays about history on top; the *Divine Comedy* turns into a pillar running from heaven to hell, a visual cosmology with "christ" on top and "torment" on the bottom; *Walden* has the same shape, but with Walden swapped in for heaven and Concord for hell; the *Odyssey* is an opposition between land and water, civilization and solitude. When applied to large corpora, the same technique does something different - picking up on high-level chronological changes. For example, the Humanist discussion group (a 27-year-old, 12-million-word listserv dating back to 1987) can be visualized as a horizontal progression between the computing terminology of the late-80s and the modern milieu digital humanities in 2014. Under the hood, Textplot uses a novel approach to determine how the words should be connected in the network. Each term is converted into a probability density function across the width of the text, and then the edge weights in the network are computed by measuring the statistical distances between individual pairs of words. This results in a schematic organization of the document as a whole, instead of a more granular view of the relationships among words at the level of phrases or sentences.

Drawing inspiration from Franco Moretti's applications of social network analysis and mapping to literary texts, these

types of visualizations can serve as critical "deformances" (McGann, Samuels), alternate versions of literary objects that tease out relationships that might otherwise remain hidden. In this case, I will argue that the networks open the door to a particular form of *visual* close reading, in which the network layouts make it possible to confirm, challenge, or generate intuitions about the original texts that can then be re-applied to traditional interpretive tasks. The networks also point to new ways of comparing and contrasting literary objects. At a small scale, for example, it's possible to compare different translations of the same text - the network layouts can be sandwiched on top of each other, which makes it easy to compare the words that different translators use to represent the same regions of meaning. Or, at a much larger scale - could the graphs provide new types of metrics that would make it possible to pick out similarly "shaped" texts in large corpora? Would it be possible to trace the evolution of textual structure and organization over time?

Why We Are Designing Hardware Peripherals for Text Analysis

Stan Ruecker, Kim Erwin, <u>Celso Scaletsky</u>, Guilherme Meyer, Xinli Lin, Elizabeth Jernegan, Júlia Dias
E-mail: c.scaletsky@gmail.com

In this paper, we describe our recent attempts to design new forms of hardware peripherals that are dedicated to understanding and interpreting text. The projects are predicated on the following factors:

- a desire to leverage embodiment
- the continuing trajectory of ubiquitous computing toward ambient information systems
- a growing community of qualitative analysts interested in working collaboratively

To anchor our discussion, we provide two cases studies: one dealing with creating interpretive encoding; the other with examining existing encoding.

"Everything on Paper Will Be Used Against Me: Quantifying Kissinger"

<u>Micki Kaufman</u>
E-mail: mkaufman@gc.cuny.edu

Perhaps more than any other former Secretary of State or National Security Advisor, the public celebrity of Henry Kissinger was (and remains) based upon an array of paradoxes. As compelling as foreign policy insights, the internal contradictions within Kissingers political and personal (and ostensibly, moral) psychology have been the subject of much of the recent historical study of Kissingerology. Traditional scholars have faced significant difficulty in their efforts to understand what appear at first to be apparently incompatible motives and behavior, attempting to reconcile facets of a public figure who appears to embody a host of mutually exclusive dichotomies.

Needless to say, disagreements about the man and his motives persist just as new interpretations based on text analysis techniques like word frequency analysis, topic modeling, sentiment analysis and corpus linguistics, are becoming possible. Combined with novel forms of knowledge production like distant reading, and made comprehensible using new, interdisciplinary tools and techniques, including the language of visual design, new examinations into the Kissinger/Nixon White House using new methods provide not merely new answers to existing questions they allow entirely new questions to be asked.

As detailed on the project's web site (http://www.quantifyingkissinger.com) my work involves the application of a host of quantitative text analysis methods and data visualization designs to an analysis of the Digital National Security Archive (DNSA)s Kissinger Collection. The archive, comprising approximately 17,600 meeting memoranda (memcons) and teleconference transcripts (telcons) detailing the former US National Security Advisor and Secretary of States correspondence during the period 1969- 1977, has served as a basis for combining political/international relations history with the fields of linguistics and visual design.

As one example, the projects combination of the computational approach employed herein with an emphasis on emotional history has illuminated new avenues of inquiry that combine text analysis with more subjective questions around interpersonal dynamics and individual psychology. In addition to analyzing Kissingers social relationships, the analysis has studies aspects of his, and the administrations, behavior around matters of secrecy and violence in one example uncovering the highly selective, event-driven use of the telcons (which Kissinger expected to remain private) versus a more generic and non-committal use of the memcons for discussing matters regarding bombing and [Cc]ambodia.

Even more tantalizing than detailing existing knowledge about Former Secretary of State Henry Kissinger, the project has begun to surface deeper understandings and questions about how this new kind of distant knowledge is formed and the ramifications for historical analysis in general. By combining computational and emotional history, a number of interesting results have emerged about the man as much as the geopolitical focus of the administration he served - from his preference for outside channels to his proclivities for seduction, secrecy, humor and violence. The application of computational techniques to the study of twentieth century US diplomatic history has thus far generated useful finding aids for researchers, provided essential testing grounds for new historical methodologies, and prompted new interpretations and questions about Kissinger and the individuals, organizations and events of his tenure.

2:45pm - 3:00pm **Louis Pasteur 154/155**	**Coffee Break 6: Coffee Break**
3:00pm - 4:30pm **Colonel By C03**	**9a: CCA and CSDH/SCHN joint session** Round-table on "Open Access Awareness: What Scholars Need to Know" with John Willinsly, Martin Eve, Heather Morrison, Andrea Whiteley Chaired by Michael E. Sinatra
3:00pm - 4:30pm **Louis-Pasteur 154**	**9b: Tracing Ideas and Events**

Millican to Ferguson: Digital Representations of Activists Discourse Then and Now

<u>Toniesha Taylor</u>
E-mail: tltaylor@pvamu.edu

In 1868, two years after the end of the Civil War in the United States, many Southern states that had seceded from the Union found themselves needing to make a variety of social, cultural and economic shifts in order to re-enter the Union. Texas, which was not readmitted to the Union until 1870, found the transition slow and at times, only possible through United States Military occupation. Moments of high radicalized tension marked the years between 1865 and 1870. In 1868 Millican, Texas became known for one of the largest bloodiest race riots in Texas, and perhaps American, history. The rioting lasted for nearly a week and left between six and an unconfirmed number dead (historical documents vary greatly on the figure). Newspapers around the country and throughout the world carried stories of the event in 1868. The dataset for this paper includes recently digitized American newspapers that covered stories of the Millican riots, particularly those that use language to describe Black and White participants in the events as activist engaged in what was then called Negro Suffrage Movement, now called the Civil Rights Movement.

The Ferguson protest are happening in a cultural and historical time where newspapers are not the only means of mass communication. The advent of Twitter equals simultaneous transmission of information by individuals, organizations, media outlets including newspapers, and governments. We no longer rely on the gatekeepers —newspaper editors or government agents—for information. We are in a new age of activism and activist language. The dataset for viewing activist and activist activities in Ferguson include a robust twenty-four hour twitter cycle.

This papers takes a long view of history with the aide of digital tools for analysis. Rather than apply traditional rhetorical methods, which call for a limiting dataset, engagement with digital tools makes possible the search for activist language and terms to engage in content analysis comparing and contrasting the terms used in newspapers

to describe events in Millican in 1868 with those on twitter and in newspapers in Ferguson in 2014. The central claim of this paper is that digital tools for analysis make possible a comparative analysis of social movements by centralizing activist language. This project asks several questions including, is it possible to see the arch of activism through language and rhetorical analysis of newspaper coverage? Can digital tools for discourse analysis provide a clear landscape for mapping language over time and space? How does the change in the rate of information transition between Millican and Ferguson change the impact of social justice activism. Is there a connection through a singular social movement that can be traced through rhetorical acts of resistance and social justice? The author makes the claim that it is possible—with a long view of history and the use of digital tools to answer these questions.

The Business of Culture: A Single-Subject Community

Nandita Dutta, Bennett Mcarthur
E-mail: ndutta2@uwo.ca, bmcarth3@uwo.ca

Single-subject communities are 'hypertopical' journalistic platforms, approaching singular, specified topics through different media formats. Attracting a more loyal, engaged following than larger platforms due to the specialized nature of their content, these platforms may represent a solution to the plight of understaffed and underfunded newsrooms, which are often unable to devote the necessary human or material resources to exploring nuanced topics

Titled 'The Business of Culture,' our project incorporates journalistic concepts and practices associated with single-subject communities with the purpose of exposing the mechanisms of cultural production to the general public.So far, no such single source exists that may approach familiar concepts through an analytical lens, exploring the economic patterns that drive production and consumption of cultural products. This analysis will then be distributed in different forms via a digital interface—not unlike a traditional web-based media platform. Subjects of study have included but not limited to Music, Gaming, Publishing and Public Health.

To establish and maintain the platform, first comes the generation of content that entails an intersection of journalistic and academic research, most of which has been in development since August 2014. The content creation process has been a joint endeavour amongst some Graduate and Undergraduate students at Western University, some of whom are part of the CulturePlex Lab, while others' work on the project counts towards their Digital Humanities Minor.

Stories will require close reading and analysis, and will be accompanied with data visualizations—all in an effort to make our content accessible for non-academic readers. Content is structured as such: Research pieces with data analysis; short pieces highlighting emerging trends in culture production and consumer interaction; visualizations updated regularly to show global trends in the reach and transfer of products and revenue; an 'around the web' section that brings together relevant real time cultural news across major online news platforms; opinion pieces from industry experts; and a regularly updated ticker of the numbers related to a particular cultural product.

The second phase involves maintaining a WordPress website to host the content, allowing for easy alterations and access to metrics. The platform is meant to be a 'Weekly' and is expected to go live in end-January with an initial runtime of four months. Content generation will continue as new stories emerge and are investigated. The team will expand to include social media analysts who will lead concurrent analysis of how the target niche community of users interacts with both the platform and the content itself.

Since the platform is not live yet, what expectations we have of the platform's run have been projected from the best practices reports of other Single-Subject online environments. Most of these center on the experience of the user, using social media to deliver stories directly to their target audience and to encourage interaction with their content through commenting and sharing. Following these methodologies within the Digital Humanities may prove integral to fostering a community of learning that extends beyond the confines of academia.

Wicking Ideas

Andrea Budac, Geoffrey Rockwell, Zachary Palmer, Robert Budac, Stan Ruecker
E-mail: abudac@ualberta.ca

The Wikipedia is over 13 years old and its articles now represent a major resource for understanding the world. Many of these articles have been negotiated and edited for a decade or more and the history of that editing can provide insight into the recent history of ideas. This paper describes the development of a tool that works with the Wikipedia API (Application Programming Interface) to scrape the Revision History of any Wikipedia article to build a corpus for subsequent text analysis and visualization. Specifically we will provide some background about the Wikipedia and how researchers are using it as a research resource, show the WIcker (Wikipedia Idea scraper) and how it generates a corpus for visualization or study with other tools, and finally discuss how this can help study the recent history of the evolution of ideas by showing a couple of examples.

3:00pm - 4:30pm **Louis-Pasteur 155**	**9c: Consuming Culture**

Decomposing the Human Development Index with Respect to Music

Matthew Harold Woolhouse, Michael Barone, Nick Rogers
E-mail: woolhouse@mcmaster.ca

The desire of the chief architect of the United Nations Development Program's 1990 report, Mahbub ul Haq, was "to shift the focus of development economics from national income accounting to people centered policies" (UNDP 1990). That is to say, previous calculations of national development had concentrated on Gross Domestic Product growth at the expense of more human-centered, quality-of-life measures. Attempting to redress this imbalance, ul Haq and others proposed a single composite statistic, ranging from 0 (less developed) to 1 (more developed), composed of various dimensions critical to human wellbeing. These included life expectancy, educational attainment, and standard of living. In 2012, Human Development Index (HDI) values were calculated for 187 of the 193 Member States of the United Nations.

Given that the HDI is socially embedded, reflective of human activities and deprivations rather than merely industrial output, we are examining the extent to which the widespread cultural practice of music consumption correlates with this and other UN indices. A supposition is that if the HDI represents aspects of human wellbeing faithfully, and its widespread adoption would suggest that this is indeed the case, then patterns of music listening—an activity closely associated with human wellbeing—should be related to the factors used within "human development accounting" (Sen, 2000: 18). Moreover, given the detailed date/time information associated with each download within our database, patterns of leisure and work emerge that are expanding the interpretation and relevance of HDI into a hitherto unexplored human activity, namely music consumption.

The results of our research are highly promising. For example, statistically significant relationships exist between music download dispersions across various times scales and HDI. Briefly, there is a far higher variability in downloading in more developed countries than less developed. Although the reasons are complex, undoubtedly some social factors have greater causal relationships to these patterns than others, and it is these causal factors that we are exploring. For example, with respect to download dispersions across various time scales, as countries industrialize and living standards improve individuals typically do less unpaid work (Gershuny, 2003), freeing them to participate in consumer culture to a greater degree, including consuming music.

The database used in this study contains 1.4 billion downloads of some 17 million mobile-phone users accessing MixRadios's online music stores around the world (countries high to low on the HDI spectrum are represented.) Within the database, user attributes include date and (local) time of download, (anonymized) user ID, start date, download count, and country; track attributes include artist, genre, track name, duration, album, label, rightsholder, and so on. Methodologically, these data are arranged within a relational database management system and queried using the MySQL implementation of SQL. A Python Database API is used to run automated iterative queries,

extracting and comparing information from multiple tracks, artists, users and countries within the database.

Plotting Television Narrative Shapes

Jarom McDonald
E-mail: jarom_mcdonald@byu.edu

This paper offers a tentative thesis for what we might measure to plot the shape of contemporary multimedia narratives, and seeks to explore how this metric might play out in shaping a complex narrative -- specifically, we will look at Twitter activity during the act of viewership as a way to plot points of narrative engagement. Such a methodological starting point might seem overly superficial on first glance ... after all, we're just counting tweets over time. But when placed in juxtaposition to a narrative, the meta-commentary of a Twitter backchannel represents some key factors in the unfolding of complex TV narrative. Like a traditional text narrative, television stories are consumed linearly, over the course of a chronological period of time. Yet due to the unique nature of broadcasting, multiple people are watching an episode simultaneously, allowing us to aggregate responses. When sets of aggregations are compared to each other over the course of an hour, we can see narrative moments that elicited significant interactive response, and those that left things silent. To test this theory, then, we wrote Twitter aggregators and populated them with the hashtags of a collection of contemporary American TV shows -- some in the genre of Jason Mittell's "complex television" narrative, others more procedural driven (with some control television broadcasts measured as well that have no pre-set narrative to shape, such as NFL football broadcasts). We collected nearly 4 million tweets over the course of 2 months (for about a dozen shows), and after filtering out retweets, tweets that didn't happen during an actual live broadcast, and spam, were left with a still useful dataset of nearly 1 million objects.

We're still in the process of analyzing the shapes that result when plotting these data points, but it's already clear that shows have their own shape ... and by getting away from sentiment and emotion, and instead focusing on interactivity and narrative engagement, we hope that we can find patterns in these shapes (especially when stretching them out and looking and the shape of a single episode in conversation with one we might plot for a longer running story arc or season) that might reveal a little more about how complex TV narrative functions.

Bibliography

Mittell, Jason. "Narrative Complexity in Contemporary American Television." The Velvet Light Trap 58 (2006). http://muse.jhu.edu/login?auth=0&type=summary&url=/journals/the_velvet_light_trap/v058/58.1mittell.html

Jockers, Matthew. "A Novel Method for Detecting Plot." http://www.matthewjockers.net/2014/06/05/a-novel-method-for-detecting-plot/

Rockwell, Geoffrey. "Kurt Vonnegut on the Shape of Stories." http://theoreti.ca/?p=4382

Vonnegut, Kurt. "Kurt Vonnegut at the Blackboard." Lapham's Quarterly, 2005. http://www.lqbeta.org/voices-in-time/kurt-vonnegut-at-the-blackboard.php?page=all

Vonnegut, Kurt. "The Shapes of Stories." https://www.youtube.com/watch?v=oP3c1h8v2ZQ

Bullipedia: lessons from StackOverflow's reputation system

Antonio Jiménez-Mavillard, Juan Luis Suárez
E-mail: ajimene6@uwo.ca

Overview

elBulliFoundation seeks to be a center for creativity and innovation in high cuisine. Originating from *elBulli* , voted best restaurant in the world five times by *Restaurant* magazine, the foundation's main project endeavors to become a hub for gastronomic knowledge held within an online encyclopedia on cuisine, *Bullipedia* . However, this is an idea yet to be developed. Thus, the question to answer at this point is: *"What should the Bullipedia be like?"* We have identified several requirements that this project must meet, from which we should note *fostering crowdsourcing* and *encouraging community participation* .

Numerous authors have addressed researches that point out the advantages of crowdsourcing. Besides, *Bullipedia* is a project inherently 2.0 that can harness the collective intelligence and generate value from their users. To engage the community, some researchers suggest that efforts should be focused on rewarding schemas, such as recognition and reputation. In addition, willingness to collaborate is strongly dependent on the trust level in a community and the value of its information.

In order to achieve these goals, we propose to incorporate the reputation system of *StackOverflow* , a popular question and answer (Q&A) website, to the future *Bullipedia* . This website encourages participants to vote questions and/or answers up or down depending on the quality of their content, while their respective authors earn reputation points. This approach has a double purpose:

- It supplies information about levels of community involvement.
- It provides incentives for effective contributions and good behaviour.

Thesis

The challenge we face is to determine whether or not the community (even if it is highly involved) is creating valuable knowledge in the *Bullipedia* . Our assumption is that at *StackOverflow* , knowledge is created when a questioner accepts one of the answers, i.e., given the corpus of answers for a question, he/she marks one of them as the answer that solved his/her problem. A question and its answer form what we call a *unit of knowledge* . The behavior of *Bullipedia* is slightly different. The information has the same structure (question/answers – recipe/comments) as *StackOverflow* but not the same functionality (recipes are not questions and comments are not answers). Therefore, we need a correspondence between the *units* of *Bullipedia* and *StackOverflow* .

Methodology

To build up this mapping, we applied *sentiment analysis* to a dataset from *StackOverflow* . The dataset is formed by questions and their respective answers. For each question, we analyzed the sentiment of its answers and determined if the question got a positive or a negative evaluation. Also, we noted whether or not the question was answered (had an accepted answer). The results of this analysis revealed that in the general case, answered questions generally obtained a positive evaluation, whereas unanswered questions got a negative evaluation.

Conclusions

The reputation system on the *Bullipedia* will encourage users to contribute with their ideas, and by applying the same type of *sentiment analysis* to the recipes' comments, we will be able to determine if the recipe was *accepted* as valuable knowledge or not by the community.

3:00pm - 4:30pm **Louis-Pasteur 285**	**9d: Big Data**

Clustering-based Interestingness Measure for Linguistic Pattern Exploration: Application to French Classic Literature

Mohamed Amine BOUKHALED, Francesca Frontini, Gauvain Bourgne, Jean-gabriel Ganascia
E-mail: mohamed.boukhaled@gmail.com

We present a computational stylistic study of French classic literature texts based on a data-driven approach where discovering interesting linguistic constructs is done without any prior knowledge. More specifically we propose an objective methodology to assist linguists in the study of syntactic style. The developed method is meant to support stylistic textual analysis by:

1) Verifying the degree of importance of each linguistic pattern (syntagmatic segments) via a new clustering-based interestingness measure.

2) Automatically inducing a list of linguistic features that are significant, representative for an author's work.

In our study, we consider a syntagmatic approach. The text is first segmented into a set of sentences, and then each sentence is mapped into a sequence of syntactic (POS-tag) items. For example, the sentence "Le silence profond régnait nuit et jour dans la maison" is first mapped to a sequence of: <DET><NOM><ADJ><VER><NOM><KON> <NOM><PRP><DET><NOM> <SENT>

Then sequential patterns of a determined length are extracted, such as :

• <DET><NOM><ADJ>

• <NOM><ADJ><VER><NOM>

• <KON><NOM> <*> <DET><NOM>

As sequential pattern mining is known to produce a large quantity of patterns even from relatively small samples of texts, an interestingness measure should be applied on these patterns in order to identify the most important ones.

To the best of our knowledge, all the interestingness measures proposed in the literature to deal with this issue are based on the support of the pattern, that is, the frequency of occurrence of those patterns in the texts. However, frequency based methods are argued not to be precise in studying linguistic phenomena unless a huge quantity of texts is used.

In our approach we presented a new interestingness measure based on the position in which a pattern appears in the text, rather than its frequency. Our hypothesis is based on the fact that the most characterizing linguistic patterns should significantly reflect the author's stylistic choice which makes theirs occurrences controlled by the author's purpose, while the irrelevant linguistic patterns are distributed randomly in the text. Following this idea, the assumption made in this approach is that the higher the importance of a linguistic pattern, the more its occurrences cluster together, detaching themselves from a random distribution. By this methodology we search for patterns whose frequency is much higher in single portions of texts than in others, thus making each of them the locally most prominent pattern.

In the results drawn from a pilot experiment on Eugenie Grandet, by Honoré de Balzac the top 10 most frequent patterns are representative of structures that are typical of the French language, and thus not very informative on the style of the novel.

On the contrary, the patterns extracted by our measure are less frequent, and more peculiar: for instance some patterns involve proper names (NAM), which tells us that the novel represents social interactions between different characters.

In the final work a thorough analysis of the patterns extracted with this methodology will be given for a small selection of French novels in order to illustrate the kind of stylistic features that it enables to identify.

The Politics of Macroanalysis

Elisa Tersigni
E-mail: elisa.tersigni@mail.utoronto.ca

Scholars have long noted the bias of the OED and other dictionaries of the English language as better representative of canonical writers, such as William Shakespeare, from whose works lexicographers largely derive their examples of language and vocabulary use. I contend that the same canonical and gender bias of dictionaries extends to methods of macroanalysis. Specifically, I argue that studies of early modern women writers are significantly disadvantaged when compared with studies of early modern men writers, because women use a less standardized register of English that is more difficult to process. Because the characteristics of women's early modern English are not identified, and not identified as distinctly different from men's early modern English, algorithms are written to deal with men's English, making women's English more difficult to process. The results of that processing misrepresent women's English and further obscure the differences of that register.

While natural-language processing is rapidly improving, algorithmic analysis of text is still better suited to standard Englishes—modern American and British English, or similar dialects thereof, that are standardized and regularized. Currently, there is no adequate method of automating macroanalysis of early modern English. Few attempts have been made to accommodate the peculiarities of the language, such as the lack of standardized spelling and grammar, because the issues of variation are difficult to resolve. I argue that the issues are further complicated because what we currently characterize as early modern English is actually a language made up of multiple, distinct registers, which include women's early modern English and men's early modern English. While these two registers differ significantly in several different ways, I focus on spelling variation, as spelling is the primary way my lemmatization program processes text.

The data illustrates differences in spelling variation according to gender and date. In the sixteenth-century, spelling variation in men's early modern English decreases by 65% starting in 1553, whereas spelling variation in women's early modern English only decreases 23% beginning in 1587. The data shows not only that standardization of English begins much sooner than the eighteenth-century, but also that standardization of women's early modern English begins decades after the standardization of men's early modern English. Even into the seventeenth-century, women have 58% greater spelling variation than men—what I call the "standardization gender gap."

My corpus includes 50 English letter writers from 1450 to 1700, men and women representing different classes and occupations in the period. I constructed the corpus (5000 words from each author) using original letters transcribed from unedited manuscripts and using collections of recently edited letters. I wrote a semi-automatic lemmatizer program in VBA that uses a dictionary of 35,000 Early Modern words and their lemmatized forms, which I also compiled.

Temporal and gender biases will be inherent in macroanalysis until we design algorithms and dictionaries that not only acknowledge but also deal with different registers of language. But to design these, we first need to know these registers exist, and the details of their fundamental characteristics.

Reading Algorithms: Critical Approaches to the Computational Mediation of Humanities Resources

Glenn Roe, Mark Olsen
E-mail: markymaypo57@gmail.com

The exponential increase in the scale of humanities datasets is altering the relationship of scholars to the objects of their research. Scholarly interactions are increasingly mediated by sophisticated systems at almost every stage of our work, from initial search to final analysis. Technologies that facilitate this work are found in modern search engines, recommendation systems, classifiers and navigation aids. Text analysis can be done on machine-generated classes of documents. Reading is performed not only via human-created links but also when systems propose related materials based on similarity calculations. While effective and powerful, these algorithms and techniques are not a panacea to be adopted and employed uncritically for humanistic research. We should focus not only on the performance of specific algorithms and applications, but also on the theoretical and methodological assumptions that underpin them. In short, we should adopt the same critical awareness in regards to machine learning algorithms as we do to the study of our traditional objects of research.

Our critical position is not meant to undermine recent efforts to apply machine learning and data mining approaches to humanities questions. Quite the contrary: the ability to query collections that represent a significant fraction of the published record opens exciting new avenues of research, from distant to close reading, with an unprecedented ability to see how texts relate to context. Much of the power of modern machine learning algorithms, however, is paradoxically dependent on reductive techniques that seem antithetical to humanistic text analysis and reading practices, as well as to more general poststructuralist sensibilities. Features of documents employed in machine learning applications tend to be restricted to small subsets of available words. Document clustering based on relatively small feature sets and an often-arbitrary number of clusters similarly tends to focus on broad patterns. Lost in all of these operations are the marginal and exceptional, rendered hidden and invisible as it were, in classification

schemes and feature selection processes.

Our paper will explore some the assumptions that inform advanced text mining, classification and recommendation systems, as well as the various limitations that arise from their use in humanistic research, with specific examples from our own work. Our contention is that the use of smarter systems to sift through the ever-increasing amounts of electronic texts to detect meaningful patterns offers the ability to frame new kinds of questions. But these technologies also bring with them a set of assumptions, often hidden in the software we use, which should be subject to careful critical scrutiny. There is a significant need in the humanities, with practitioners of the digital humanities at the fore, to do this critical work; to open the black box and shine light on what is inside. Deconstruction in the digital library should be a reading strategy not only for the materials found therein, but increasingly of the systems being deployed to manage, control and make the contents of electronic resources accessible and intelligible.

4:45pm - 5:30pm **Colonel By C03**	**AGM: Annual General Meeting**
5:30pm - 6:45pm **Colonel By C03**	**Keynote: Amy Earhart: DH Futures: Conflict, Power, and Public Knowledge**