

2016 Research Showcase

Monday, October 24

12:00 - 4:30 p.m.

**Illini Union, Ballrooms A and B
1401 W. Green Street, Urbana**

Keynote speaker: Melissa Cragin (PhD '09), executive director, Midwest Big Data Hub, NCSA

12:00 - 1:00 p.m. Poster session I
1:00 - 2:00 p.m. Presentation session I
2:00 - 2:30 p.m. Break
2:30 - 3:30 p.m. Presentation session II
3:30 - 4:30 p.m. Poster session II

Refreshments will be provided during the poster sessions.

Does consumer product labeling actually protect consumers?
Can scientific models lead to greater workflow reproducibility?
How does digital literacy, traditionally defined, hinder digital inclusion?
Does implanting technologies inside the body alter how we define "human"?
Can social informatics principles transform big data?

<http://ischool.illinois.edu/research/showcase>

Overview of Sessions

1:00 – 2:00 p.m. • *Presentation Session I*

YESWORKFLOW: MORE PROVENANCE MILEAGE FROM HYBRID PROVENANCE MODELS AND QUERIES

Bertram Ludäscher (presenter); joint work with Duc Vu, Qiwen Wang, Yang Cao, Qian Zhang, Timothy McPhillips

ALL AND EACH: THE DYNAMICS OF SCALE IN DIGITAL HERITAGE CULTURES

Rhiannon Bettavia

INFORMED CONSENT AND CHEMICAL EXPOSURE FROM EVERYDAY CONSUMER PRODUCTS

Henry Gabb

DIGITAL INCLUSION, DIGITAL EQUITY, AND THE URGENT NEED TO RECONSIDER DIGITAL LITERACY

Sharon Irish (presenter); joint work with Martin Wolske

2:30 – 3:30 p.m. • *Presentation Session II*

THE SOCIOTECHNICAL IMAGINATION

Les Gasser

IMPLANTABLE BRAIN TECHNOLOGIES AND THE CREATION OF CYBORGS

Beth Strickland

THE ONION ROUTER: UNDERSTANDING A PRIVACY ENHANCING TECHNOLOGY COMMUNITY

Masooda Bashir (presenter); joint work with Hsiao-Ying Huang

INTRODUCING THE AUTHOR-ITY EXPORTER, AND A CASE STUDY OF GEO-TEMPORAL MOVEMENT OF AUTHORS

Vetle Torvik (presenter); joint work with Mikko Tuomela and Brent Fegley

12:00 – 1:00 p.m. • *Poster Session I*

3:30 – 4:30 p.m. • *Poster Session II*

CLASSIFICATION AND DETECTION OF MICRO-LEVEL IMPACT OF ISSUE-FOCUSED FILMS BASED ON REVIEWS

Rezvaneh Rezapour, Jana Diesner

TOWARDS AUTOMATIC DATA EXTRACTION FROM CLINICAL RESEARCH REPORTS: A CASE STUDY OF A SYSTEMATIC REVIEW OF ORAL PAIN RELIEF

Linh Hoang, Jodi Schneider

"WHAT IS YOUR EVIDENCE?" A STUDY OF CONTROVERSIAL TOPICS ON SOCIAL MEDIA

Aseel Addawood, Masooda Bashir

TALKING BACK TO THE PUBLIC LIBRARY: MEASURING AND CONCEPTUALIZING THE LITTLE FREE LIBRARY MOVEMENT

Kate Williams, James Whitacre, Noah Oluwafemi Samuel, Elizabeth C. Hartman, Alexandra M. Budz, Katherine A. Cawley, Salem A. Gebil, Kelly Y. Greeling, Victoria E. Henry, Brittany S. Jones, Jenna Kim, Amanda M. Luna, Emily C. Purcell, and Paul A. Wheelhouse

APP AUTHORS: THE LAUNCH YEAR

Deborah Stevenson, Rachel M. Magee, DeAnza Williams

EXPLORING J-DISC: SOME PRELIMINARY ANALYSES

Yun Hao, Kahyun Choi, J. Stephen Downie

INCLUDING MARGINALIZED VOICES: BUILDING A COLLECTION OF DALIT RESOURCES

Nisha Mody, Nicole A. Cooke

SHOPPING FOR SOURCES: AN EVERYDAY INFORMATION BEHAVIOR EXPLORATION OF GROCERY SHOPPERS' INFORMATION SOURCES

Melissa Ocepek

MASSIVE DATA, INDIVIDUAL LEARNERS: CHALLENGES FOR DEVELOPING HOLISTIC VIEWS OF MOOC PARTICIPANTS

Paige Cunningham

UNDERSTANDING THE NEEDS OF SCHOLARS IN A CONTEMPORARY PUBLISHING ENVIRONMENT
Katrina Fenlon, Maria Bonn, Harriett Green, Christopher R. Maden, Megan F. Senseney, Aaron McCullough

INVESTIGATING THE RELATIONSHIP BETWEEN GLOBAL MEASLES INCIDENCE AND SOCIO-DETERMINANTS OF HEALTH
Maria Bohri, Ian Brooks

TOWARDS A PROGRESSIVE MODEL FOR METACOGNITIVE SKILLS AND MAKERSPACE LEARNING
Jeff Ginger, Maya Israel, Lisa Bievenue, Rebecca Teasdale

TUMBLR AND GENDER PRONOUNS
Julia Burns Petrella

GAMIFICATION IN CITIZEN SCIENCE: A CASE STUDY FROM GALAXY ZOO
Lo Lee, Linda C. Smith

YOUNG RESEARCHERS
Rachel M. Magee, Margaret Buck

NEXT-GENERATION INNOVATION APPLICATIONS FOR URBANA-CHAMPAIGN SMART GIGABIT COMMUNITY
Martin Wolske, Chieh-Li "Julian" Chin

ISSUE-FOCUSED DOCUMENTARIES VERSUS OTHER FILMS: RATING AND TYPE PREDICTION BASED ON USER-AUTHORED REVIEWS
Ming Jiang, Jana Diesner

ACQUIRING AND REPRESENTING DRUG-DRUG INTERACTION KNOWLEDGE AS CLAIMS AND EVIDENCE
Jodi Schneider, Richard D. Boyce

DEVELOPING A CODING TEMPLATE FOR ANALYZING COLLEGE-LEVEL WORKS OF LITERATURE IN HIGH SCHOOL
Pompilia Burcica, Rachel M. Magee

STARTED WITH A HURRICANE: INCORPORATION OF LAW SCHOOL LIBRARIES SERVICES AND PROGRAMMING WITH LAW SCHOOL PRO BONO SERVICE AND PUBLIC INTEREST LAW TO SUPPORT NON-PROFIT COMMUNITIES
S.K. (Kayleigh) Van Poolen

TRENDS IN CENTURIES OF WORDS: PROGRESS ON THE HATHITRUST+BOOKWORM PROJECT
Peter Organisciak, J. Stephen Downie, Jacob Jett

RADICAL READING STUDIES: EVALUATING THE POSITION OF READER-FOCUSED CURRICULUM IN INFORMATION SCIENCE EDUCATION
Mikki Smith, Elaine Martaus

DIGGING, REACHING, AND LEARNING: AN UPDATE ON THE FIRST YEAR OF THE HATHITRUST RESEARCH CENTER'S LIBRARIAN TRAINING PROGRAM
Harriett Green, J. Stephen Downie, Eleanor Dickson, Ruohua Han

RESPONSIBLE CONDUCT OF RESEARCH IN HUMAN CENTERED DATA SCIENCE: COMPLYING WITH NORMS AND REGULATIONS
Chieh-Li "Julian" Chin, Jana Diesner

ONLINE INFORMATION SEEKING FOR CONTROVERSIAL TOPICS

Aseel Addawood, Masooda Bashir

TWO AND A HALF DIMENSIONAL PRINTING: USING 3D PRINTING TO MAKE RARE MATERIALS ACCESSIBLE

Jon Sweitzer-Lamme, Rachel M. Magee

MOBILIZING TAXONOMIC DATA: THE IMPORTANCE OF USER-CENTERED DESIGN

Andrea K. Thomer, Michael B. Twidale, Jinlong Guo, Roger A. Burks, Matthew J. Yoder

EXPLORING THE DYNAMICS OF INTERDISCIPLINARY COLLABORATION USING SOCIAL NETWORK ANALYSIS: THE EFFECTS OF FORMAL AND INFORMAL NETWORK STRUCTURES

Ly Dihn, Barley William

THE ONION ROUTER: UNDERSTANDING A PRIVACY ENHANCING TECHNOLOGY COMMUNITY

Hsiao-Ying Huang, Masooda Bashir

SIMULATING SOCIAL SYSTEMS AT SCALE

Les Gasser, Santiago Nunez-Corralles

LINKING SCHOLARS AND SEMANTICS: DEVELOPING SCHOLAR SUPPORTIVE DATA STRUCTURES FOR DIGITAL DUNHUÁNG

Jacob Jett, J. Stephen Downie

Presentation Abstracts

YESWORKFLOW: MORE PROVENANCE MILEAGE FROM HYBRID PROVENANCE MODELS AND QUERIES

Bertram Ludäscher (presenter); joint work with Duc Vu, Qiwen Wang, Yang Cao, Qian Zhang, Timothy McPhillips

The provenance of a data product aims to describe (a) its actual data lineage and processing history captured at runtime (retrospective provenance) or (b) the general "recipe" or workflow by which this kind of data is derived in general (prospective provenance). Scientific workflow systems provide (b) essentially "for free" and often have specialized recorders to capture (a). Script-based workflows and analysis tools (written, e.g., in Python, R, or MATLAB) can be instrumented in various ways to capture (a), but little support exists for (b). The YesWorkflow toolkit closes this gap by (i) providing a simple, annotation-based approach that allows users to model their scripts as workflows, (ii) incorporating any available retrospective provenance (recorded or reconstructed) from script runs, and (iii) allowing users to build "bridges" between the different kinds of provenance information, so that more informative "hybrid" provenance queries can be answered. We illustrate the utility of different kinds of provenance queries, using a number of simple, script-based workflows from different science domains (crystallography, climate modeling, and astrophysics).

ALL AND EACH: THE DYNAMICS OF SCALE IN DIGITAL HERITAGE CULTURES

Rhiannon Bettavia

Open-access web-based technologies provide new methods for fostering engagement between cultural heritage organizations and their audiences. However, the institutional imperatives that drive their development have sometimes given cause for unease about the maintenance of autonomy for

those using them (Andrejevic, 2007). Foucault's (1979) later work on 'governmentality', or the rationality of government, speaks to this concern. Foucault identifies the mutually reinforcing relation of all and each, 'to develop those elements constitutive of individuals' lives in such a way that their development also fosters that of the strength of the state'. The suggestion that these totalizing and individualizing effects work to bolster one another is an important insight for the theorization of scale with regard to the dissemination of digital cultural content. Furthermore, in the heritage context, the dynamics of all and each support a method of investigating, more closely, the negotiation of those audience relationships outlined above. This paper investigates the move of all and each as it relates to the digital cultural heritage project, Europeana. Europeana is one of the more recent attempts by the European Union (EU) to confirm Europe as a unified entity through the notion of a shared cultural heritage. Funded by the European Commission (EC), the EU's executive body, Europeana demonstrates this move very clearly; it is precisely the involvement of individual users and organizations that furthers its aim to promote a distinctly European space online for heterogeneous cultural objects and experiences. A similar political imperative is implicit in its technical functionality and examination of these elements reveals the myriad scales at which Europeana operates: from a single node of the data model to the larger, supranational scope of the web portal.

INFORMED CONSENT AND CHEMICAL EXPOSURE FROM EVERYDAY CONSUMER PRODUCTS

Henry Gabb

People today have a chemical load that is significantly higher than a century ago. Chemical exposure can come from many sources, e.g.: indoor and outdoor pollution, everyday consumer products, food storage containers, etc. This is not acute poisoning. The effects are more subtle. We are, in effect, the test subjects in an uncontrolled biochemistry experiment. This has become an accepted, or perhaps ignored, tradeoff of life in modern society. The Fair Packaging and Labeling Act (FPLA) requires manufacturers to list ingredients in "descending order of predominance" on product labels. The FPLA was a good step toward empowering consumers to make informed decisions about the products that they use. However, a recent analysis of product labels revealed a surprising degree of synonymy among ingredient names. For example, octinoxate, a common ingredient in sunscreens because of its ultraviolet filtering properties, appears in ingredient labels under the following names: octinoxate, octyl methoxycinnamate, and ethylhexyl methoxycinnamate. These names are synonymous according to PubChem, a chemistry database provided by the National Library of Medicine. Octinoxate is a suspected endocrine disrupting compound but a consumer who is actively trying to avoid this chemical must know all of its synonyms in order to make an informed decision regarding product selection. This is just one example among many. Chemical synonymy obfuscates the information in product labels and undermines the informed consent that the FPLA attempts to provide. A case can be made to amend the FPLA to standardize ingredient nomenclature, particularly with respect to potentially harmful ingredients.

DIGITAL INCLUSION, DIGITAL EQUITY, AND THE URGENT NEED TO RECONSIDER DIGITAL LITERACY

Sharon Irish (presenter); joint work with Martin Wolske

It's easy to imagine the eventual role information science professionals have in advancing people's digital literacy skills given the foundational role information and communications technologies play in supporting the transformation of information into knowledge so as to support human flourishing within an "information age" and a "knowledge economy". But is it possible that in our very efforts to "bridge the digital divide" and build "21st century digital literacy skills" that we are actually further deconstructing civil society and civic engagement? Is it possible that rather than fostering self-sufficiency and community collaboration we are instead furthering magical thinking about technology, our belief in the supremacy of the technocrat, and the centrality of market forces fostering passive

consumerism?

In this presentation it is argued that to achieve digital inclusion and equity, a radical reconsideration of digital literacy is essential. Such reconsideration borrows from both traditional digital literacy and computational thinking definitions, while also bringing in a critical sociotechnical perspective. Development of specific technical skills is thereby resituated to in-fill, just-in-time learning rather than the primary goal of digital literacy training. Instead, primacy is given to developing the cognitive, socio-emotional, and information seeking skills needed to achieve the application goals of people. In so doing, individuals and communities advance their own knowledge power and agency to effectively select, [co-]create, and use sociotechnical products to amplify their forces and capacity for human and community development.

THE SOCIOTECHNICAL IMAGINATION

Les Gasser

After more than 40 years of research, the modern field of Social Informatics has developed numerous insights and principles for explaining interactions among information/communication technologies (ICTs) and social systems such as organizations, societal institutions, and groups. Over time, consensus has grown around these scientific foundations, but there's still little perspective on how to use them to actually do "Sociotechnical Analysis of ICTs" - to expose specific enabling conditions and consequences of socially embedded ICTs, and to shape sociotechnical design and policy processes. This research has developed a set of core theoretical perspectives on social informatics and how they matter in the practice of sociotechnical analysis for policymaking, sociotechnical "forensics," and design. The theoretical perspectives range from function and economic analyses, to social power and political economy, to the aesthetics and social meanings of ICTs. The integrating principle for this research is "architectures": sets of nested dynamic constraints and opportunities that participants co-construct, and within which participants interpret situations and act upon them. We show how SI principles, analyzed in this framework, apply in substantive areas of concern for organizations and societies, such as transformations of work, privacy, "big data," and the new problematics of autonomous systems, IoT, and information / cyber-warfare. The research also has developed strategies for communicating, learning and exercising Social Informatics foundations.

IMPLANTABLE BRAIN TECHNOLOGIES AND THE CREATION OF CYBORGS

Beth Strickland

The use of implantable technologies within humans sparks debates about the ethical use of such technologies since they penetrate the traditional skin-and-skull boundaries of what constitutes the body. These boundaries work to designate a person as human versus cyborg when using technology, but what happens to our conceptualization of "human" when technologies are implanted inside the body? The ethical debate about the use of these technologies posits on one end of the spectrum the acceptance of certain implantable devices, those used for therapeutic purposes, and at the other end the unacceptable use of devices to enhance normal physical capabilities. Individuals who use implantable devices for therapeutic purposes are viewed as still human even though their physical bodies have been merged with technological devices. However, those who use technology as a way to upgrade their body get labeled as cyborgs. Why does this distinction exist and why does it matter? In this presentation I begin by considering how we use traditional physical boundaries as a way to define what's human. Next I examine how these boundaries get crossed by the medical use of deep brain stimulation to treat Parkinson's Disease related tremors. This technology is then examined using the therapy versus enhancement framework to consider how it represents the full physiological fusion between human and machine; yet, also demonstrates a social aversion to labeling users as cyborgs. This presentation concludes with a brief discussion about what this labeling aversion suggests for the discourse of implant ethics and the future of human embodiment.

THE ONION ROUTER: UNDERSTANDING A PRIVACY ENHANCING TECHNOLOGY COMMUNITY

Masooda Bashir (presenter); joint work with Hsiao-Ying Huang

Internet technologies have made mass surveillance prevalent and much easier to carry out, while at the same time making personal privacy more difficult to protect. The ubiquity of personal data processing has raised public awareness about the infringement of information privacy. To protect users' information privacy, several initiatives have been developed to provide Privacy Enhancing Technologies (PETs). One of the most well-known PETs is the Onion Router (Tor) network, which provides users with online anonymity. The Tor network is supported by a group of volunteers who contribute their resources to sustain the availability and quality of the service. However, Tor Volunteers may find themselves in a tough spot at times because Tor network is often monitored by law enforcement, which makes this PET community different from any other open-source initiatives. To explore this volunteer community's motivation for providing their services despite the risks, we conducted an online survey. Our study results reveal that one of the main motivations for these volunteers is to advocate and provide privacy for online users. In addition, Tor-relay operators report on their views about anonymous networks, the challenges they face, and how their belief in providing an opportunity for everyone to access information without interference or censorship is a key component of their volunteer participation.

INTRODUCING THE AUTHOR-ITY EXPORTER, AND A CASE STUDY OF GEO-TEMPORAL MOVEMENT OF AUTHORS

Vetle Torvik (presenter); joint work with Mikko Tuomela and Brent Fegley

We introduce a web service, Author-ity Exporter, that permits searching and exporting data from Author-ity -- a database that has PubMed author names disambiguated with a high degree of accuracy [1]. Each author is represented by a cluster of papers annotated by publication count, time-span, affiliations, topics, journals, co-authors, citations as well as imputed data from MapAffil [2], Genni [3], and Ethnea [4] and links to their NIH/NSF grants and USPTO patents; and we have plans for more. This service should enable and simplify new types of author-centered bibliometric analyses with a unique strength in funding, geography, and diversity (gender, ethnicity, and professional age). We also present an illustrative case study of modeling of authors' career movements to and from a specific city based on data retrieved from Author-ity Exporter. The service (and the R code used in the case study) are available at <http://abel.ischool.illinois.edu/cgi-bin/exporter/search.pl>.

Poster Abstracts

CLASSIFICATION AND DETECTION OF MICRO-LEVEL IMPACT OF ISSUE-FOCUSED FILMS BASED ON REVIEWS

Rezvaneh Rezapour, Jana Diesner

In this paper, we developed a new method to enhance the existing research in the area of review mining and impact assessment to extract the micro-level influence of information products (documentary films) from user-generated content (reviews). Documentary films aim not only to tell a compelling story but also to engage the public, inform the society and raise awareness about social justice issues, among other goals. Based on these significant effects, we developed a classification schema, related codebook for corpus annotation, and prediction model for detecting multiple types of impact of documentaries on individuals, such as change versus reaffirmation of cognition, emotions, and behavior.

Our data annotation and analysis showed that information products can change the individuals' perception about social justice problems, be associated with changes in people's attitudes toward societal problems, and move people to act. i.e.: donation to charities. We also found that types and magnitude of impact are closely correlated with the types and foci of the films. These are significant outcomes for sponsoring organizations such as foundations, focused movements, as well as filmmakers as it shows the potential impact of documentary films and highlights the importance of assessing impact beyond binary methods as well as developing new strategies for achieving certain types of impact based on prior projects. In addition, the proposed codebook and findings can advance the research in review mining by enabling the extraction of new levels of information from reviews- different types of impact- and goes beyond the previously conducted researches, such as ratings prediction and polarity extraction. To find and predict different types of impact we performed supervised learning by using three sets of features: linguistic, lexical, and psychological and achieved 81% accuracy (F1) using a Random Forest classifier, and 73% using SVM. (This abstract is accepted for ACM, CSCW '17)

TOWARDS AUTOMATIC DATA EXTRACTION FROM CLINICAL RESEARCH REPORTS: A CASE STUDY OF A SYSTEMATIC REVIEW OF ORAL PAIN RELIEF

Linh Hoang, Jodi Schneider

In healthcare, it takes a long time for new treatments to move from clinical studies into practice: perhaps an average of 17 years [Balas et al., 2000]. Systematic review is a critical step in this research translation process because it determines what is known. To do this, a systematic review analyzes all available evidence on a particular question through a series of steps, including data extraction. The current best practice for data extraction is for two people to independently identify and extract data from each research paper. Because the data extraction step is almost always performed manually, it is very time-consuming [Tsafnat et al., 2014] yet methodological errors may cause problems with the review's conclusions [Lundh et al., 2009]. Our long-term goal is to help reviewers synthesize the literature quickly and accurately by developing a semi-automatic support system for data extraction. Towards this end, we are currently conducting an in-depth case study of a single systematic review, a Cochrane Review about oral pain relief. Through manual annotation and a content analysis of the six studies synthesized by this Cochrane Review, we will develop hypotheses about which clinical data elements can be automatically extracted. We will also develop an annotated corpus which will enable us to propose methods for automatically supporting human reviewers in data extraction. Eventually, we plan to design a semi-automated support system, and to test the two hypotheses (1) that it can reduce the time and human labor required to conduct a review and (2) that it can maintain or increase the quality of the resulting review.

"WHAT IS YOUR EVIDENCE?" A STUDY OF CONTROVERSIAL TOPICS ON SOCIAL MEDIA

Aseel Addawood, Masooda Bashir

In recent years, social media has revolutionized how people communicate and share information. One function of social media, besides connecting with friends, is sharing opinions with others. Micro blogging sites, like Twitter, have often provided an online forum for social activism. When users debate about controversial topics on social media, they typically share different types of evidence to support their claims. Classifying these types of evidence can provide an estimate for how adequately the arguments have been supported. We first introduce a manually built gold standard dataset of 3000 tweets related to the recent FBI and Apple encryption debate. We develop a framework for automatically classifying six evidence types typically used on Twitter to discuss the debate. Our findings show that a Support Vector Machine (SVM) classifier trained with n-gram and additional features is capable of capturing the different forms of representing evidence on Twitter, and exhibits significant improvements over the unigram baseline, achieving a F1 macroaveraged of 82.8%.

TALKING BACK TO THE PUBLIC LIBRARY: MEASURING AND CONCEPTUALIZING THE LITTLE FREE LIBRARY MOVEMENT

Kate Williams, James Whitacre, Noah Oluwafemi Samuel, Elizabeth C. Hartman, Alexandra M. Budz, Katherine A. Cawley, Salem A. Gebil, Kelly Y. Greeling, Victoria E. Henry, Brittany S. Jones, Jenna Kim, Amanda M. Luna, Emily C. Purcell, and Paul A. Wheelhouse

In just one decade, North Americans and others have taken their own initiative to create and maintain an estimated 36,000-plus little free libraries right in their front yards. In the U.S., these “take a book, leave a book” boxes harken back 100 years to when the new library profession organized their libraries to deliver books to factories, shops, community centers, construction sites, and even railyard frequented by hoboes. Little free libraries also resemble in some ways the people’s libraries that emerged on the 2011-2012 Occupy sites as that movement challenged the 1% on behalf of the 99%. Protestors and police saw these as guerrilla libraries; police and legal action ensued, rallying many librarians. Using GIS and US Census data, we analyze the US little free libraries in relation to residential patterns to see who they are most likely serving. We study seven more closely and find out what is circulating. Judging by our alumni, librarians love these little book exchanges. Measured against the library science canon from Ranganathan to Wiegand, what does this renaissance of libraries an reading say to the 16,500 government-funded public libraries across the United States? What can the global profession learn from this surging movement?

APP AUTHORS: THE LAUNCH YEAR

Deborah Stevenson, Rachel M. Magee, DeAnza Williams

App Authors: Closing the App Gap II is an IMLS-funded project to develop a curriculum for teaching app creation to 9-11-year-olds in school and public libraries. Such a curriculum provides an avenue for youth, especially underserved youth who may have little prior interest in STEM-based learning or opportunity for coding, a chance to explore a popular and appealing format. This was the first year of our three-year project, and the poster will document the background principles we drew on and the development and implementation of the curriculum we taught at our two partner sites, one school library and one public library, this year.

EXPLORING J-DISC: SOME PRELIMINARY ANALYSES

Yun Hao, Kahyun Choi, J. Stephen Downie

This poster presents our recent work of conducting illustrative analyses of J-DISC data. JDISC is a specialized digital library for information about jazz recording sessions. Because of the rich structured and searchable metadata included in J-DISC, it has the potential for supporting a wide range of studies on jazz, especially the musicological work of those interested in the social network aspects of jazz creation and production. In this work, we conducted both informetric and network analyses of the entire J-DISC data which comprises data on 2,711 unique recording sessions associated with 3,744 distinct artists including such influential jazz figures as Dizzy Gillespie, Don Byas, Charlie Parker, John Coltrane and Kenny Dorham, etc. Our analyses show that around 60% of the recording sessions included in J-DISC were recorded in New York City, Englewood Cliffs (NJ), Los Angeles (CA) and Paris during the year of 1923 to 2011. Furthermore, our analyses of the J-DISC data show the top venues captured in the J-DISC data include Rudy Van Gelder Studio, Birdland and Reeves Sound Studios. By presenting these exemplar analyses, we aim to better illustrate the kinds of uses that musicologists could make of this collection as well as to show other potential uses of the data. We proposed that other potential uses of the J-DISC data include building ground truth for artist similarity and other MIR research, enabling new visualizations or developing interfaces for novel jazz music streaming services, etc. We also proposed that when associated with jazz audio, J-DISC data can be used to create a comprehensive digital library and could be the foundation of new avenues of research of

computational musicology. It is also remarked that although the data itself is limited in completeness, in the future J- DISC could still contribute to enhancing accessibility of jazz resources by becoming part of the Linked Open Data network to link to other jazz resources that are available online.

INCLUDING MARGINALIZED VOICES: BUILDING A COLLECTION OF DALIT RESOURCES

Nisha Mody, Nicole A. Cooke

The Hindu caste system positions Dalits, also known as untouchables, at the bottom. This inequality manifests itself into continual injustices within this community in South Asia. Dalit resources are essential because marginalized voices can be heard only if they are included. However, building such a collection is complex to develop and discover and analyze due to South Asian vernacular language differences and semantic vocabulary affecting discoverability. Focusing within India, this poster explores the process of discovering Dalit-specific resources (books, serials, and audiovisual resources), analyzing the makeup of the collection, developing a research guide, and making collection development recommendations for this critical subset within the South Asian Studies collection at the University of Illinois.

SHOPPING FOR SOURCES: AN EVERYDAY INFORMATION BEHAVIOR EXPLORATION OF GROCERY SHOPPERS' INFORMATION SOURCES

Melissa Ocepek

The field of everyday information behavior addresses how individuals interact with information in their daily lives. Previous research in the field has largely ignored the banal and quotidian portions of everyday life, such as grocery shopping, which represents a gap the current project fills. Through two empirical studies using qualitative methods, the present work presents the prevalence and variety of information sources used by grocery shoppers. Findings indicate that grocery shoppers rely on close human sources, domain specific sources, and on surprisingly few online sources throughout the process of their grocery shopping. The findings demonstrate the information-richness of grocery shopping and suggest that other everyday spaces may be fruitful areas for information behavior research.

MASSIVE DATA, INDIVIDUAL LEARNERS: CHALLENGES FOR DEVELOPING HOLISTIC VIEWS OF MOOC PARTICIPANTS

Paige Cunningham

In 2012 the University of Illinois and Coursera first partnered to offer MOOC courses. Between November 1, 2012 and March 31, 2015 there had been over 550,000 active enrollments in the university's Coursera courses, spread across 15 individual course titles (University of Illinois, 2016). As of September 1, 2016, the total number of live course titles has risen to 58, the majority of the additions being part of Coursera-based specializations and graduate degree programs. Early classes had finite start and end dates, but courses are now offered on an open-ended on-demand platform.

While MOOC courses generate large amounts of data about online learners, the data produced is not necessarily meaningful. Grade data is relatively easy to extract, but provides only a superficial understanding of learner engagement. More meaningful glimpses into learner engagement can be acquired from in-depth lecture-watching data and clickstream data. However, such data sources are not currently available for on-demand courses, making accurate comparisons between session-based and on-demand courses problematic. Additionally, in an on-demand environment it can be difficult to determine whether learners are finished with a course if they have not met the completion standards. A significant percentage of learners report meeting their own goals without meeting the official completion milestone, thus ceasing active participation, but may remain phantom members of an on-demand course indefinitely. Finally, previously both free and paid learners could complete all course activities, allowing for collection of all learners' data. A shift towards only permitting students to

complete graded assessments after paying for a certificate means that evaluation of their levels of engagement will be based on sparse lecture watching and forum participation data. Integrating different data sources makes defining an overall picture of learner engagement more difficult, though the process can provide granular pictures of individual students' participation in individual courses.

UNDERSTANDING THE NEEDS OF SCHOLARS IN A CONTEMPORARY PUBLISHING ENVIRONMENT

Katrina Fenlon, Maria Bonn, Harriett Green, Christopher R. Maden, Megan F. Senseney, Aaron McCullough

The advent of digital scholarship, together with the increasing momentum of open access for research, have thrown scholarly publishing into flux. In the shifting landscape of scholarly communication, libraries are taking on new roles in order to publish scholarship in innovative forms: by building institutional and data repositories, for example, or collaborating with faculty to develop multimedia monographs or open access journals. The Publishing Without Walls project at the University of Illinois is developing a service model for university libraries to support scholar-driven, openly accessible, scalable, and sustainable publishing practices. To this end, we are conducting a multimodal study of scholars' needs, objectives, and practices in this new age of publishing. This poster presents preliminary results of a large-scale survey of scholars, which aims to shed light on what and how scholars want to publish, when and why they choose to publish digitally, and how they understand the success of their digital publications. This survey, in tandem with the results of a series of interviews and focus groups, is actively informing the development of two digital monograph series at the University of Illinois University Library, along with a model for their development that may be useful to other libraries involved with publishing new modes of scholarship.

INVESTIGATING THE RELATIONSHIP BETWEEN GLOBAL MEASLES INCIDENCE AND SOCIO-DETERMINANTS OF HEALTH

Maria Bohri, Ian Brooks

Many common infectious diseases can be prevented by vaccines. Among them measles is a one of the most easily transmissible diseases that still has a high prevalence rate throughout the world. In order to determine fully the relationship between the measles coverage rate for a country and the measles incidence rate for a given year, this research focused on building a model to understand what health or development indicators, GDP per capita, healthcare availability, political stability ratings, etc., affect the incidence rate for measles globally.

TOWARDS A PROGRESSIVE MODEL FOR METACOGNITIVE SKILLS AND MAKERSPACE LEARNING

Jeff Ginger, Maya Israel, Lisa Bievenue, Rebecca Teasdale

Faculty and staff from the University of Illinois-Urbana Champaign (UIUC) College of Education, the Illinois Informatics Institute, and the Champaign Urbana Community Fab Lab have been actively collecting data in order to study a series of metacognitive strategies that support learning and engagement for struggling middle school students during makerspace experiences. The makerspace movement has gained recognition and momentum, which has resulted in many schools, libraries and places of informal learning integrating makerspace technologies and practices into their programming (Angevine & Weisgrau, 2015; Schneider, 2015, Abram 2013, Barniskis 2014). Given the paucity of studies to inform practitioners about what pedagogical supports help struggling learners engage in these makerspace experiences, this is an essential area of study. We would like to present our metacognitive skills data analysis model, preliminary findings as well as practical information for educators interested in fostering effective learning in makerspace settings.

TUMBLR AND GENDER PRONOUNS

Julia Burns Petrella

Tumblr is a popular microblogging website and mobile app which allows users to post, share, and remix a variety of media, including text, images, and animated GIFs. Users are not required to post any identity information on their blogs, though some choose to offer this type of information, and I began noticing that many users opt to provide their preferred gender pronouns. My research interest focuses on the use of Tumblr to communicate about and express gender identity online. Recent research has found that “for transgender, transsexual, genderqueer, and gender nonconforming people, emergent media technologies offer new outlets for self-representation” (Fink & Miller 611). By performing a content analysis of 200 individual posts and the corresponding blogs, I explore the context and content of communication about gender pronouns on Tumblr. Early analysis shows a diverse representation of gender identities, and the content of the blog posts shows topics including gender-related announcements, misgendering, and negative backlash toward those who post about nonbinary gender identities. Future directions for this research include continuing to analyze the wealth of data for additional meaning, as well as beginning to look at the data across various demographic factors. It would be greatly illuminating to combine this research with in-depth interviews with Tumblr users who blog about gender pronouns to further understand the contexts of the posts and to ask about whether the affordances of Tumblr, such as anonymity online, make this specific online community an acceptable and attractive space for expressing and communicating about gender identity.

GAMIFICATION IN CITIZEN SCIENCE: A CASE STUDY FROM GALAXY ZOO

Lo Lee, Linda C. Smith

While citizen science, also known as crowd science, is a relatively new term just defined by the dictionary, its practice dates from 1883. Citizen science denotes an information transfer and collaboration process between scientists and the general public. In a citizen science project, scientists ask for the public’s help to accomplish a research task under their supervision, and there is no need for the public to have solid background in the scientific domain to participate in the research. Currently, citizen science is widely seen in diverse disciplines, including astronomy, oceanography, ornithology, and biology. It is a method that a number of scientists rely on to efficiently conduct research. One of the prominent citizen science examples is an astronomy project named Galaxy Zoo. Galaxy Zoo is still an ongoing project in which scientists require volunteers’ assistance to classify galaxy images from the Sloan Digital Sky Survey. They use an online platform to display galaxy images, and volunteers learn how to categorize each galaxy by its characteristics. Whereas Galaxy Zoo has created significant results since it was first launched in 2007, the task design remains tedious, boring, and has no attraction to engage users to stay longer in the project. To address this issue, we use gamification as an approach to redesign the task in Galaxy Zoo. Regarding the power of storytelling and fantasy, in this research we investigate the feasibility of implementing narratives in the design of citizen science games. There are two prototypes of redesigned Galaxy Zoo systems: one is to incorporate a story adapted from *The Little Prince*, and the other is to incorporate an original story. The goal of this research is to examine whether the integration of a familiar story is able to enhance users’ motivation more when compared to a newly developed narrative.

YOUNG RESEARCHERS

Rachel M. Magee, Margaret Buck

The Young Researchers project is a multi-stage research project working with Illinois teens to collaboratively develop research and informal learning materials. We are examining how teens interact with science, technology, engineering, and mathematics (STEM) topics as well as their everyday life

technology use. In the current stage of the project, we are working with a group of teens from a small community in Illinois. These teens, or Young Researchers, are learning about how to conduct research by actively participating in study design, participant recruitment, data collection, and analysis and reporting of research findings. The Young Researchers will be visiting UIUC in the spring to share what they learn and to see how students here engage in research activities. In the future we'll be working with additional sites around Illinois. Conducting this participatory research with youth has several outcomes. We're learning more about teen perspectives on STEM experience and technology use. We're also investigating how to use computing technology to support informal learning, and developing best practices for involving teens in research and learning. The Young Researchers are building literacy skills, becoming researchers, and contributing to scholarship while exploring their potential to pursue research and STEM careers. Working with Young Researchers around the state, we will iteratively develop and assess a learning module to support youth in acquiring STEM and research skills. This will be publicly available online, allowing teachers, librarians, parents, and youth themselves to stimulate engagement with research and STEM learning.

NEXT-GENERATION INNOVATION APPLICATIONS FOR URBANA-CHAMPAIGN SMART GIGABIT COMMUNITY

Martin Wolske, Chieh-Li "Julian" Chin

Presidential initiatives like "Smart Cities" and federal frameworks like "Smart and Connected Communities" are helping to focus attention on the potential for new highspeed Internet networks to magnify the work of citizens, non-profits, governments, and private industry. The Center for Digital Inclusion has collaborated with US Ignite, a 501(c)(3) nonprofit organization with original inspiration from the White House Office of Science and Technology Policy and the National Science Foundation, to build a living smart gigabit applications testbed in Urbana-Champaign that serves as one of the 15 Smart Gigabit Communities in the nation. The goal of the project is to foster the development and deployment of next-generation innovative applications that leverage the highspeed network such as Urbana-Champaign Big Broadband (UC2B) and the rich human and technology resources in our campus. We aim to identify researchers and developers from the University of Illinois and the Urbana-Champaign community to develop applications focused on six areas of national priority, including: Education and Workforce Development, Energy, Health, Public Safety, Transportation, and Advanced Manufacturing. These applications will provide transformative benefits to our community and share with other users on the Smart Communities Network. The poster will cover information about the Smart Gigabit Communities and UC2B that the Center for Digital Inclusion has been working on, and engage audience to actively participate in the development of next-generation innovative applications.

ISSUE-FOCUSED DOCUMENTARIES VERSUS OTHER FILMS: RATING AND TYPE PREDICTION BASED ON USER-AUTHORED REVIEWS

Ming Jiang, Jana Diesner

User-authored reviews can be a valuable resource that offers a window into micro-level engagement of people with information products such as books and films. Leveraging reviews for media impact assessment is an insufficiently understood task. Based on a corpus of film reviews that we collected from Amazon, we extend classic review mining research by developing an understanding of the perception of documentaries in comparison to feature films. We also identify unique text patterns of reviews between these two types of films as well as between low versus high ratings. For the performances of classification, our results show that ratings of non-documentary (feature film) reviews can be predicted with higher accuracy (73.67%, F1 score) than ratings of documentary reviews (68.05%). We can also predict whether a review was written for a documentary or non-documentary film with an accuracy of 71.32%. Our results suggest that people have different foci when reflecting on different types of films. For text patterns, we find that in contrast to reviews of feature films, comments on documentaries are: 1) shorter but composed of longer sentences, 2) less emotional but containing

more negative terms, and 3) lexically more concise with more verbs than nouns and adjectives. Compared to low-rated reviews, comments with a high rating are shorter, more emotional, and contain more positive than negative sentiment. Overall, this work contributes to advancing our understanding of the impact of different types of information products on individual information consumers. This work has been published for ACM Hypertext and Social Media (HT'16).

ACQUIRING AND REPRESENTING DRUG-DRUG INTERACTION KNOWLEDGE AS CLAIMS AND EVIDENCE

Jodi Schneider, Richard D. Boyce

Potential drug-drug interactions (PDDIs) are a significant source of preventable drug-related harm. Poor quality evidence on PDDIs, combined with prescribers' general lack of PDDI knowledge, results in thousands of preventable medication errors each year. One contributing factor is that PDDI knowledge lacks a standard computable format. To address this, we are researching efficient strategies for acquiring and representing PDDI knowledge, focusing on assertions and their supporting evidence. We are acquiring knowledge from several sources. First, we have transformed 410 assertions and 519 evidence items from prior work. Second, we are examining FDA-approved drug labels, and so far annotators have identified 609 evidence items relating to pharmacokinetic PDDIs from 27 FDA-approved drug labels. Third, annotators have found 230 assertions of drug-drug interactions in 158 non-regulatory documents, including full text research articles. We are building a two-layer evidence representation, with both generic and domain-specific layers. The generic layer reuses the Micropublications Ontology to annotate assertions and their supporting data, methods, and materials. For the domain-specific component we are building DIDEO—the Drug-drug Interaction and Drug-drug Interaction Evidence Ontology. DIDEO adds specific knowledge, such as the study types required to establish a given type of PDDI. The current version of DIDEO has 385 subclass axioms, and reuses formalized knowledge items, including from the Drug Ontology, Chemical Entities of Biological Interest, the Ontology of Biomedical Investigations, and the Gene Ontology.

DEVELOPING A CODING TEMPLATE FOR ANALYZING COLLEGE-LEVEL WORKS OF LITERATURE IN HIGH SCHOOL

Pompilia Burcica, Rachel M. Magee

This project will work with high school students to develop a curriculum that promotes understanding complex literature through coding tasks. Difficult novels can defy structure, lack a meaningful plot, lack a beginning or end, and appear as directionless or gibberish. With coding considered a new literacy, combining hard novels and coding appear to be a daunting match. This project aims to support teaching hard novels using a simple, stratified language, in an open-source environment, and through a design recipe that speaks to teachers' work in the classroom and their vocabulary. Efforts to understand literature are sometimes focused on narrative structure and how events in a logical sequence lead to a cohesive plot. However, in his declarative theory of composition, Hemingway claims that "none of the significant things are going to have any literary sign marking them...you have to figure them out by yourself." (p. 51). The coding activities will focus thus, not on language, structure, or characters, but on an anchor that many secondary teachers use: physical places. Place has many merits: it is inextricably tied to identity, to travel and tourism, local traditions and gender particularities. Starting small, with a stylistically sparse and short novel, *The Sun Also Rises*, this project will develop a package of coding activities that use physical places as the main anchor to grasp a novel. Two programming applications will be used: the Keyword in Context (KwiC) application, which is analogous to the context of the keyword, and Text Segmentation, which will sort location. By connecting to this complex text through active programming, students will benefit from this empowering technology by practicing the mechanisms of programming, and by making predictions about the settings through the eyes of the characters, or illustrating geospatially scenes according to the sentiment measurements of the keywords in context. The corollary of their semester-long work on this novel is the creation of a walk or travel handbook based on these mining activities.

STARTED WITH A HURRICANE: INCORPORATION OF LAW SCHOOL LIBRARIES SERVICES AND PROGRAMMING WITH LAW SCHOOL PRO BONO SERVICE AND PUBLIC INTEREST LAW TO SUPPORT NON-PROFIT COMMUNITIES

S.K. (Kayleigh) Van Poolen

Katrina and the subsequent levee system failures hit New Orleans the weekend I began law school. The scenes were horrific, and many of us reeled from the overwhelming injustice from the lack of information to the underlying reasons for such profound loss of life. I went from struggling to write legal briefs one weekend to working with thousands of other law students to rebuild New Orleans and the surrounding communities over several years. Consistently, the challenges of rebuilding New Orleans involved information—those needing information and those attempting to provide information. One of the challenges involved educating hundreds of lawyers and law students on how to navigate the civil code system of Louisiana. Building on those New Orleans' experiences, as well as other academic, volunteer and professional experiences, I seek to explore the questions of how is information generated, used and distributed in non-profit organizations. Of particular interest are organizations with an emphasis on social justice or social impact research. Additionally, I want to explore how a law school, specifically an academic law library could provide information, programming or research assistance. The research consists of three small projects, all of which create a foundation for further academic study. First, interviews non-profit organizations will establish a foundation for the development of an ongoing assessment of information needs. Second, data from the ABA (American Bar Association) published law school data on pro bono service programming (PBSP) and public interest law (PIL) will build a composite index of how much PBSP and PIL occur within a particular law school's curriculum. Third, a primer will outline how a law library can develop programming and services to support the PBSP and PIL for their law school community. By combining these three aspects, the aim is to create a more integrated conversation of how the law school community can support the community in which it resides.

TRENDS IN CENTURIES OF WORDS: PROGRESS ON THE HATHITRUST+BOOKWORM PROJECT

Peter Organisciak, J. Stephen Downie, Jacob Jett

The digitization of scanned book collections has resulted in new ways to inquire about trends in history, culture, and language. However, when presented with text at the scale of millions of books, the size makes it conceptually and technically difficult to appreciate these underlying patterns. The HathiTrust+Bookworm (HT+BW) project is addressing this difficulty through visualization, developing a suite of tools intended to nurture inductive exploration of large collections. Though scholars can use this project's visualization tools for their own collections, HT+BW is providing an implementation of one of the most generally useful ones: the HathiTrust Digital Library, a consortial collection of published materials that holds nearly 15 million scanned volumes. We will demonstrate the newest improvements to the HT+BW project and discuss its value to scholars.

RADICAL READING STUDIES: EVALUATING THE POSITION OF READER-FOCUSED CURRICULUM IN INFORMATION SCIENCE EDUCATION

Mikki Smith, Elaine Martaus

In 1997, Wayne A. Wiegand's article "Out of Sight, Out of Mind" outlined the relevance of reading research in Library and Information Science education; in 1999, Eliza Dresang's Radical Change explored the kind of disciplinary ground that Wiegand advocated, focusing on the content of information objects and their relationship to readers. Nearly 20 years later, these two works continue to provide an intersection from which to consider the position of reading research within the broader professional discourse of LIS and in relation to youth services studies in particular. Drawing on Wiegand's argument

that curricula codify professional values, we endeavor to evaluate the current significance of reading research to LIS education, especially as educators negotiate the relationship between print, digital, and youth cultures. The first phase of this study involves content analysis of course titles and descriptions of current offerings from the top LIS programs in the most recent U.S. News & World Report rankings across three areas: Best Library and Information Studies Schools, School Library Media, and Services for Children and Youth. This poster presents the results of the initial stages of this research, providing a visual representation of course offerings related to literacy, readers, and reading and the contexts in which those subjects are studied, and offers preliminary analysis. Possibilities for further research are mentioned, including an expansion to include course syllabi, a comparison with course information from the late 1990s, and a content analysis of titles and abstracts from LIS journals.

DIGGING, REACHING, AND LEARNING: AN UPDATE ON THE FIRST YEAR OF THE HATHITRUST RESEARCH CENTER'S LIBRARIAN TRAINING PROGRAM

Harriett Green, J. Stephen Downie, Eleanor Dickson, Ruohua Han

Modern academic librarianship demands reconfigured skillsets and expertise to meet the rapidly evolving needs of today's students, faculty, and researchers as they increasingly pursue digital and data-driven scholarship. These developments speak to an urgent need to provide training for academic librarians to support digital scholarship. Seeking to address this need, the University of Illinois in partnership with Indiana University, Lafayette College, Northwestern University, and the University of North Carolina at Chapel Hill has launched a 3-year, train-the-trainer project funded by the Institute for Museum and Library Services. The project focuses on building librarians' digital scholarship skillset using the HathiTrust Research Center (HTRC) as an example for pursuing text analysis research, and aims to develop a curriculum that will be shared as an open educational resource at the end of the grant. By creating opportunities for librarians to learn concepts and techniques in computational text analysis, this initiative seeks to reinforce the role of the library as a space to encounter "big data" tools and methodologies. This presentation will report on current developments of the project, including the design and implementation of a pilot curriculum taught in spring and fall of 2016, as well as lessons learned in instructional design for LIS practitioners. It will also speak to what is ahead in year two of the project, during which a partner-vetted and improved HTRC curriculum will be shared more broadly through additional workshops.

RESPONSIBLE CONDUCT OF RESEARCH IN HUMAN CENTERED DATA SCIENCE: COMPLYING WITH NORMS AND REGULATIONS

Chieh-Li "Julian" Chin, Jana Diesner

Researchers frequently collect, use, and analyze publicly available data from social networking platforms, online production communities, and customer review websites, among other sources, as part of their research activities. Technically, it can be feasible and straightforward to access and obtain public data, while considering the ethics, norms, and regulations applicable to these data requires additional awareness, knowledge and skills. This is partially due to the fact that multiple types of rules may apply, including institutional and sectoral norms and regulations, laws and regulations related to privacy as well as security, intellectual property including copyright, terms of service, and personal ethics and values. We provide an overview on these types of regulations, and also clarify on common misassumptions between free as in "free speech" (i.e. freedom from restriction, "libre") versus free as in free beer (i.e. freedom from cost, "gratis"). We outline several approaches to responsibly conducting research with public available data, and solutions to enhancing the expertise of researchers on implement data regulations, especially for the case of human centered data science.

ONLINE INFORMATION SEEKING FOR CONTROVERSIAL TOPICS

Aseel Addawood, Masooda Bashir

Over centuries, people were getting most of their information from newspapers, television, and word of mouth. In this digital age, in the contrary, people are getting most of their information from different online sources as blogs, online newspapers, and social media. In this study, We would like to investigate online users information seeking behaviors regarding controversial topics. Controversial issues do not always have a true basis of truth. For example, the community is still debating about the MMR vaccines and if they cause autism or not. In this study, we are conducting a pilot survey where we want to understand first what issues they consider as controversial. After that, we want to know what type of information sources users uses to learn more about any controversial topic. The types of information available at different sources vary, so does the quality. Also, we want to understand what factors affect people's selection of these online information sources. The purpose of the survey is to explore what affect users selection of information sources for controversial topics.

TWO AND A HALF DIMENSIONAL PRINTING: USING 3D PRINTING TO MAKE RARE MATERIALS ACCESSIBLE

Jon Sweitzer-Lamme, Rachel M. Magee

This project is designed to bring rare objects with cultural value out from behind displays and into the hands of diverse audiences. Using 3D scanning technology, this work aims to scan a collection of wood blocks carved during the 1560s, currently housed in UIUC Rare Book and Manuscript Library (RBML). Now extremely rare, the wood blocks were originally used in central Europe for the printing of an herbal—a compendium of medically useful plants. These illustrations are considered to be some of the best illustrations of plants of their times. Printing from these blocks is fraught with preservation and other risks. Now in RBML, they are not in a condition to allow for regular handling, and they are only appreciated aesthetically. However, they were and remain fundamentally objects of craft, and are designed to be used. The full artistry in these items comes with the print created from them, rather than the wood blocks alone. With the support of RBML, the Preservation department at the UIUC library, and the iSchool, the project will create a digital model of these 3-dimensional items, and then print copies of them on 3D printers. These surrogates can be used as the original wood blocks were intended to be used. Once the 3D-printed surrogates are created, the project hopes to enable students to print from these several hundred year old blocks, works of some of the greatest wood block carvers of their era, in order to connect viscerally and physically with the materials, rather than simply visually. While preservation of the objects themselves is of course vital, it is also important to preserve the methods and processes through which those objects gained cultural value— preservation both through use and of the use of the objects, with the added benefit of not threatening in any way the objects themselves.

MOBILIZING TAXONOMIC DATA: THE IMPORTANCE OF USER-CENTERED DESIGN

Andrea K. Thomer, Michael B. Twidale, Jinlong Guo, Roger A. Burks, Matthew J. Yoder

Through the NSF-Funded "Transforming Taxonomic Interfaces" project, we are working to understand the user interaction and experience needs of taxonomists – the biologists that organize, describe, classify and name life on earth. With an estimated 8.7 million (rapidly disappearing) species on our planet, this is no small task, and these are no small data. Circumscribing and mobilizing this semantically rich data is critically important for modern biology, and thus, a key challenge for modern academic publishing.

Robust user interfaces are fundamental to streamlining the pipeline from data collection to publication of open data. In this poster, we show user interface prototypes that take advantage of underlying

semantic data (e.g. ontologies like "PATO") to enhance the user experience (UX), minimize the labor of applying formal ontologies, and increase the semantic richness of the user's end product, taxonomic descriptions. User-centered design can greatly improve the potential for successful interfaces, it follows that it is therefore an integral part of publishing open data.

EXPLORING THE DYNAMICS OF INTERDISCIPLINARY COLLABORATION USING SOCIAL NETWORK ANALYSIS: THE EFFECTS OF FORMAL AND INFORMAL NETWORK STRUCTURES

Ly Dihn, Barley William

There is reason to suspect that formal and informal social networks both play important roles in fostering interdisciplinarity within scientific organizations. Formal structures are often incentivized, providing the resources necessary to sustain relationships. Informal structures, however, are more apt for the exchange of knowledge and information. Using social network analysis, we examine the role of formal and informal working relations in the development of interdisciplinarity within a scientific organization in the midst of implementing interventions to foster greater collaboration. We propose the following research question: What structural features (formal and informal) of a social network are best related with an ego's propensity to have interdisciplinary relationships? Network data were obtained from a multiple name-generator network survey taken by 178 members of a large interdisciplinary scientific organization. Respondents were able to list up to 40 names of colleagues with whom they engaged in work relationships such as advice seeking and relying on to perform tasks. Participants then rated the perceived similarity between each ego-alter's primary areas of expertise and reported their level of familiarity with each alter's primary area of expertise. We used these measures to produce a novel metric capturing the presence of "valuable interdisciplinary relationships" (i.e. those with alters who have distant expertise, but in an area where the ego has a sufficient level of familiarity to facilitate communication). We found clear differences in the formal and informal network structures, in which key players differ in each network. Egocentric analyses of key players in the informal network also reveal important members of the organization who are influential and can promote ideas of interdisciplinarity efficiently. We found that the informal network (advice-receiving) to be denser, thus smaller diameter, than the formal network, suggesting that informal relationships are able to provide the knowledge resources necessary to strengthen interdisciplinary collaboration. Continuing egocentric analyses will explore how each individual's structural position within formal and informal network relates with their propensity to have as well as produce "valuable" interdisciplinary ties.

THE ONION ROUTER: UNDERSTANDING A PRIVACY ENHANCING TECHNOLOGY COMMUNITY

Hsiao-Ying Huang, Masooda Bashir

Internet technologies have made mass surveillance prevalent and much easier to carry out, while at the same time making personal privacy more difficult to protect. The ubiquity of personal data processing has raised public awareness about the infringement of information privacy. To protect users' information privacy, several initiatives have been developed to provide Privacy Enhancing Technologies (PETs). One of the most well-known PETs is the Onion Router (Tor) network, which provides users with online anonymity. The Tor network is supported by a group of volunteers who contribute their resources to sustain the availability and quality of the service. However, Tor Volunteers may find themselves in a tough spot at times because Tor network is often monitored by law enforcement, which makes this PET community different from any other open-source initiatives. To explore this volunteer community's motivation for providing their services despite the risks, we conducted an online survey. Our study results reveal that one of the main motivations for these volunteers is to advocate and provide privacy for online users. In addition, Tor-relay operators report on their views about anonymous networks, the challenges they face, and how their belief in providing

an opportunity for everyone to access information without interference or censorship is a key component of their volunteer participation.

SIMULATING SOCIAL SYSTEMS AT SCALE

Les Gasser, Santiago Nunez-Corralles

There is tremendous interest in computational modeling, simulation, and analysis of social, socio-technical, and socio-environmental systems, in many different application areas including basic science, government/policy, and commerce/industry. Social simulation has proven useful for basic scientific research in social, behavioral, organizational, cognitive, linguistic, evolutionary, and biological sciences. Large-scale simulations provide tremendous educational opportunities for students and faculty to explore and visualize alternative theories and models of complex social systems, and to study how they change. Most existing successful work in projects such as those cited above is primitive on the scales of simulated sizes of social aggregates, ability to integrate and execute multiple modeling perspectives, and dynamic complexity. There exists no widely accepted/available/ integrated super-scale capability. Beyond simply not having appropriate computational infrastructures, there are significant gaps in the basic science and technology of multiscale modeling, high-speed simulation, data integration and management, visualization and the representation of social systems for the current set of interesting problems, and this has been the case for many years. This project's motivating vision is to position UIUC as a major center for effective, efficient, accessible, large-scale simulations of social systems. Long term, the project will have impacts locally (providing state-of-the-art infrastructure for expanded computational research at UIUC; partnerships and collaborative opportunities; drawing more academic disciplines into computational research); and broadly (new theories of scalable simulation; micro-macro linkages; substantive theories of social systems in context; new applications to policy and “predictive analytics”). This project is tackling the early stage: creating a prototype infrastructure and a set of proof-of-concept experiments that will uncover fundamental barriers to the vision above. Along with NCSA, we are partnering with researchers at the Sandia National Labs, the Network Sciences and Simulation Lab at Virginia Tech and with the Institut de recherche pour le développement au Vietnam (IRD) in Hanoi and Paris who are supplying us with data and models.

LINKING SCHOLARS AND SEMANTICS: DEVELOPING SCHOLAR SUPPORTIVE DATA STRUCTURES FOR DIGITAL DUNHUÁNG

Jacob Jett, J. Stephen Downie

Dunhuáng Academy is digitizing the contents of the Mòg#o Caves, a network of 492 Buddhist temples and grottos built over a 1000-year span. Open questions remain regarding how to assess the caves' contents and provide scholarly access to it. This poster envisions one possible way that existing ontologies could be used to accomplish this.