TABLE OF CONTENTS

Participants

Alteri, Suzan (University of Florida)

Beveridge, Aaron (University of Florida)

Clapp, Melissa (University of Florida)

Dobrin, Sidney I. (University of Florida)

Freeman, Richard (University of Florida)

Gitzendanner, Matthew (University of Florida)

Hart-Davidson, William (Michigan State University)

Kidd, Kenneth (University of Florida)

Martin, Cathlena (University of Montevallo)

Morey, Sean (Clemson University)

Rice, Jeff (University of Kentucky)

Taylor, Laurie N. (University of Florida)

Van Horn, Nicholas (Ohio State University)

Abstract and Statements of Innovation and Humanities Significance

Abstract
The MassMine project team representing participants from the Department of English, George A. Smathers Libraries (Libraries), and Research Computing at the University of Florida (UF) requests $60,000 to finish the version 1.0 release, establish a robust training program, and promote the MassMine open source software. MassMine enables researchers to collect their own social media data archives and supports data mining, thus providing free access to "big data" for academic inquiry. MassMine further supports researchers in creating and defining methods and measures for analyzing cultural and localized trends, and developing humanities research questions and data mining practices. The primary aims of this project are to: 1) refine the MassMine tools to support collection, acquisition, and use of available social media and web data; and, 2) develop a training program with online resources for supporting the broad use of MassMine by humanities researchers, regardless of experience.

Statement of Innovation
Humanities researchers currently lack sufficient access to social media data, tools for data mining, and tools for processing data for analysis. MassMine is open source software in development to address these concerns specifically by humanists for the needs of humanists by providing a set of easy to use tools for creating social media data archives, querying and mining the archives, and revealing the processes and technologies for enabling generation of new methods and new questions.

Statement of Humanities Significance
MassMine's version 1.0 release will enable new approaches to small and big data for humanists by creating access to data with tools for data mining, processing, and analysis. This project will result in a powerful data tool for humanists with a simple GUI interface. Using an iterative development process in collaboration with humanities scholars, project results include training resources and tool documentation for MassMine for increasing capacity for data-intensive research in the humanities.

Enhancing the Humanities through Innovation

The MassMine project team representing participants from the Department of English, George A. Smathers Libraries (Libraries), and Research Computing at the University of Florida (UF) requests $60,000 to finish the version 1.0 release, develop a robust training program, and promote the MassMine open source software. MassMine enables researchers to collect their own social media data archives and supports data mining, thus providing free access to "big data" for academic inquiry. MassMine further supports researchers in creating and defining methods and measures for analyzing cultural and localized trends, and developing humanities research questions and data mining practices. The primary aims of this project are to: 1) refine the MassMine tools to support collection, acquisition, and use of available social media and web data; and, 2) develop a training program and corresponding online resources for supporting the broad use of MassMine by humanities researchers, regardless of experience.

The MassMine project is a Level II start-up grant proposal to develop MassMine for broader scale implementation to support humanities research needs for social media data collection, mining, and analysis. The MassMine developers have recognized and are responding to the importance of access for humanities data research. Humanities researchers currently lack sufficient access to social media data, further entrenching a "digital divide."[1] Humanities researchers must innovate more accessible tools for data mining, processing, and analysis. As Adam J. Banks explains, technological access is a socio-cultural concern. Banks defines technological access as material, functional, meaningful, and transformational. Material or physical access becomes gatekeeper because access to it is a prerequisite to the basic knowledge that is required to utilize technology.[2] Because of licensing limitations and advanced technical skill demands, humanists have been restricted from meaningful and transformative research using social media and web data. MassMine's initial and ongoing development seeks to address specific concerns of access and use within humanities research practices. Currently, MassMine is a console application, accessed using a command line interface (Apple, Linux, Windows) with which users input a basic configuration file that controls the data collection and data processing functionality. Users can run the MassMine console on standalone computers for individual research or on cloud-style servers. MassMine is operating successfully on UF Research Computing's servers, with MassMine accessible by UF researchers and collaborators at other institutions. MassMine version 0.1.0 code is available as open source on GitHub.

The MassMine Startup project will provide a set of easy-to-use tools and a training program for humanists to create social media data archives, query and mine the archives, and engage with processes and technologies for generating new methods and questions.[3] The MassMine Startup project seeks funding for a software programmer, cloud server hosting, and training program design, to: 1) develop a GUI; 2) build the Export & Processing Module; and, 3) implement a full training program for humanities researchers using MassMine to conduct data research. The MassMine GUI will utilize the same underlying console engine, ensuring parallel capabilities for console or GUI versions. Currently, MassMine uses data frames to store information and supports data collection from the Twitter and Google APIs as well as sets of user-supplied URLs—Facebook and Wikipedia APIs will be added as data sources by January 2015. MassMine currently stores raw social media data and web data that has useful additional information attached (e.g., timestamps, geolocation data). Extraneous data is also attached (e.g., HTML, markup, punctuation, irrelevant URLs, non-specific attached data) and this unrelated data can impede optimal analysis. The planned Export & Processing Module will add support for storing data in SQL and MongoDB database formats, exporting data to additional formats (CSV, TXT, and XLSX), and functionality for data curation (e.g., reviewing, cleaning, and subsetting data).[4] Data curation is a critical

---

[1] Boyd, D. & Crawford, K. "Critical Questions for Big Data." Information, Communication, and Society, 2012.
[2] Banks, A. *Race, Rhetoric, and Technology: Searching for Higher Ground*. Urbana, IL: NCTE, 2006.
[3] Friedlander, A. Asking Questions and Building a Research Agenda for Digital Scholarship, Washington, D.C.: Council on Library and Information Resources, 2009.
[4] Ogburn, J. "The Imperative for Data Curation." portal: Libraries and the Academy, 10(2), 2010.

part of the data research process and can consume the majority of time on any given research project.[5] With the technical enhancements, MassMine will be a comprehensive tool for collecting, processing, and exporting data to enable greater access for humanists in developing data research questions, undertaking data research, collecting data, analyzing data, and informing broader interdisciplinary data-intensive methodologies initiated by the humanities.

Social media postings are significant resources for humanities scholarship, comprised of content text of postings with valuable information attached—geolocation information and access dates support analysis of real world locations where texts have connected and moved, including tracking circulation. Social media postings often encompass forms beyond textual data, including videos, static and animated images, and memes that combine text and image elements. Humanistic modes of inquiry can be productively brought to bear on these materials, and should rely on textual practices and methods to inform the implementation of humanities data-intensive research.

Despite the value of social media postings, humanities research opportunities are currently limited because social media postings are often controlled by user licensing or service agreements, which restrict access. Humanities researchers face three connected problems when attempting to study social media and web data: 1) high cost of access to data-resellers who package and charge for licensed data from social media postings; 2) purchased data access is not for raw data; and, 3) purchased data is most often pre-filtered, already-visualized, and made available through limited browser tools focused on marketing and brand management needs. These problems can be avoided by accessing data through APIs (application programing interfaces), but accessing and collecting data via APIs, especially for systematic data processing and exporting, requires coding knowledge. Data obtained through APIs generally remains under licensing or user agreement restrictions that limit display, sharing, and certain types of data usage. Various tools support parts of this data research process; however, tools with integrated support for data collection, query development, and data mining processes are unavailable. Whereas the sciences must consider data practices in terms of the sharing and provenance of data sets that scientists have already acquired,[6] the humanities require a more comprehensive approach.[7] The MassMine Startup project will support researchers as they create their own data archives in a manner that complies with permissible uses of APIs. MassMine is designed to leverage existing APIs for large-frame support. Combined with the development of a training program, researchers will have access to data research tutorials, including training on conforming to acceptable uses of data providers as well as ethical considerations regarding privacy practices for social media and web research.

MassMine's Project Team uses open and collaborative development models, which recognize humanities scholars as core users who are integral to ongoing design and expansion. The Project Team uses a grounded approach for development, where the Project Team engages users to identify needs, conducts iterative development following the needs and concerns expressed by humanities researchers. Because humanists have had such limited access to data and data tools, training is an essential part of the proposed project. The project will include a comprehensive training program that begins with creating a Scholars Group of humanities scholars who will initially use the console application, with Project Team support as they develop research questions, collect data, analyze data, and inform broader data research practices. The Scholars Group will serve as a core cohort of humanities researchers who will respond to feature and GUI development through their feedback during and following training sessions with the Project Team. In the process of supporting MassMine's development, the Scholars Group will pursue their individual research with the complementary goal of producing publishable data analyses and visualizations. Following the release of the GUI version, the training program will expand by offering open sessions focused on different interests and needs, including teaching data research in the classroom, training for humanists about data research, and focused training for researchers in Children's Literature.

---

[5] Steve Lohr. "For Big-Data Scientists, 'Janitor Work' is Key Hurdle to Insights." *New York Times*: Aug. 17, 2014.
[6] Szalay, A. & Gray, J. "The World-Wide Telescope." *Science*, Sept. 2001.
[7] American Council of Learned Societies. "Our Cultural Commonwealth: The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences." NY: ACLS, 2006.

Research about Children's Literature and popular culture provide abundant examples of research questions related to books, films, toys, games, and related user communities and activities. Many of these have widespread community responses to authors and works. The varied training sessions will be opportunities for conducting outreach and gathering user feedback to inform MassMine's development and training offerings. The GUI release, added functionality, and training program will position MassMine to provide a scalable foundation for humanities scholars interested in employing social media and web data for research.

MassMine 1.0 will support many scholarly approaches (*see examples in appendix*) including Children's Literature and transmedia approaches. For example, Mary Roca is interested in studying how Mattel's products and interactive materials offer narrative content and directions for play, while also functioning as scripts for consumers using social media. She explains that Mattel attempts to control its brand while mining its fans for new products and content. By using MassMine 1.0 Roca will be able to study how Mattel employs a franchise management style to promote its narratives, by investigating *Monster High* and *Ever After High* for how Mattel's social media activity works in combination with the related consumer products provides data on how branded fiction promotes a consumption-based American girlhood. For another example, researchers studying particular countries and events will be able to use MassMine as a tool to collect and analyze events associated with particular locations. For example, Petrine Archer and Claudia Hucke created the online exhibition *About Face: Revisiting Jamaica's First Exhibition in Europe* to celebrate Jamaica's 50th Anniversary of Independence by revisiting *Face of Jamaica*, the country's first post-independence exhibition to tour Europe. *Face of Jamaica* toured Europe in 1963-1964 yet was never viewed in Jamaica. *About Face* reconsidered the original exhibition by re-presenting its art and related materials online. Using MassMine, scholars will be able to collect social media texts on current reception of artists and artworks featured in this nation-defining exhibit to inform research on particular artists, culture, and nationalism.

In addition to supporting the research contained in the scholarly contributions, MassMine version 1.0 will enable data research in the classroom with minimal time and technology requirements. Committed to building a strong foundation of data literacy in the humanities, MassMine will benefit teaching by enabling humanities courses to include data research, without programming or coding skills, by providing an easy to use interface. Rather than spending too much time collecting and archiving data, classes will be able to focus on framing research questions and analyses. MassMine is designed to operate long-term for large data collection projects, but it is also effective for short-term projects within a single semester—thus introducing students in the humanities to data-intensive methodologies.

Environmental Scan

The Digital Research Tools (DiRT) Directory contains many entries on tools for data collection and analysis. However, available tools do not adequately respond to the comprehensive need for tools that assist in data collection, archiving, querying, and analyzing for general humanities scholars. Programming or coding skills are often mandatory, along with funds for purchasing data, and many tools are designed to support data research without considering the limited technical support provided to humanities scholars. Tools like TAGS: Twitter Archiving Google Spreadsheet and programming guides like the Programming Historian support users with technical skills for data collection and research. Commercial data collection and analysis providers (e.g., Radian6 by SalesForce, GNIP by Twitter, Topsy by Apple, SumAll) offer free trial periods followed by expensive payment plans, with certain limits and data presets that focus on business analytics and marketing research, which cannot be used to study many of the diverse humanities researchers' questions. Building from the humanities' long history of textual analysis, recent innovative work includes text analysis by data mining large research archives. Data mining tools for already collected data or pre-existing archives (e.g., WordSeer, MALLET, Hathi Trust Research Center) often allow users to upload large corpora for indexing, text parsing, topic modeling, information retrieval, and machine learning. Tools for specific needs include tools that support necessary data curation processing (e.g., Open Refine), analysis and visualization of network structures (nodes, edges, connections) (e.g., NodeXL), and certain levels of data access and analysis and for specific data sources

(e.g., Webometric Analyst, Mozdeh for Twitter). Resources also include trainings as with the Digital Humanities Data Curation Institutes (2013-2014). MassMine's version 1.0 release will mark a significant contribution by providing improved access multiple data sources, by eliminating the need for advanced technical skills in data collection and processing, and by providing a robust training program specifically developed to support humanities research.

History and Duration of the Project

In response to data access problems, development on MassMine began in response to needs unmet by available tools. Development began in late 2013 to support research about complex circulation networks[8] by investigating the concept of "hyper-circulation"[9] using Twitter to study the relationship of trends to respective locations and content. The ubiquitous R language was chosen as the coding environment for its well-supported programming package for accessing the Twitter API for Twitter data on trends to approach theoretical problems posed by theories of hyper-circulation. Since the Twitter API provides limited historical data, more data was needed for circulation studies, informing requirements for MassMine. MassMine was coded to access the API at the maximum allowable bandwidth, archive new data as it was available, and continue to collect systematically over long periods of time to allow scholars to build large research data archives with minimal technical demands. MassMine's current 0.1.0 release supports the Twitter API, Google API, and aggregate sets of web page URLs provided by the user. The console application does not require programming knowledge for users to collect data. The application is controlled with a configuration file that is basic text so users can edit with any text editor. To use the MassMine console application, users open the command line console, run a simple command to start the R programming language, and then submit the command (MassMine) to start the software.[10] The console interface allows users to save their API usernames and passwords for quick restarting of the software, and gives the option to edit simple text-based configuration files that direct the software for the kinds of data to collect and duration. MassMine's 0.10 supports installation on both local computers and hosted servers (e.g., UF Research Computing) with the same functionality available for both.

MassMine was designed to operate properly with minimal system and bandwidth demands for accessing, pulling, archiving, and processing data. MassMine 0.10 runs efficiently on older computers with minimal requirements for local users and ensures greater ease for support by central service providers. On servers, MassMine can run as a single or multiple parallel instances with MassMine hosting by UF Research Computing starting in 2014. This project grant period (5/1/2015-4/30/2016) will build on existing support. UF Research Computing's support for UF researchers extends to their collaborators and will support the training program for this project. MassMine was also designed for lightweight and small storage demands even with big data research. Initial collection of trend data from across the US encompassed more than 1.14 million lines of data when pulling data at the maximum bandwidth allowed by Twitter's API, which resulted in a small data set (slightly over 120MBs). Because of inherent redundancy of social media and web application data, once compressed the data was less than 6MBs.

Work Plan

Specific tasks to be completed during the grant period, as detailed in the appendix, are: 1) develop a GUI that creates and edits the configuration file; 2) build the Export & Processing Module for storing data in different formats; and, 3) create a training program with in-person and online trainings; providing resources for: installing and using MassMine, developing questions for data research, and broader resources for humanities researchers in doing data research; 4) test and review features and functions of MassMine with the Scholars Group of humanities researchers; and, 5) release MassMine 1.0.

The Project Team will meet monthly to review progress on goals and project timeline milestones. Work plan feasibility is evident from planned activities and assigned responsibilities for Project Team

---

[8] Taylor, M. *The Moment of Complexity: Emerging Network Culture.* Chicago: U of Chicago P, 2001.
[9] Dobrin, S. *Postcomposition.* Carbondale: SIU Press, 2011.
[10] See appendix with guide to running MassMine with textual descriptions and screen shots.

members for all work elements. The Project Team has explicit and reasonable goals for the life of the grant, team members with appropriate skills and successful collaborative project experiences, and technical and institutional support to achieve project goals. The Project Team's diverse members bring unique skills and perspectives along with support and stakeholder commitments from their institutional areas. The proposed project results and products will be:

1) Graphical User Interface (GUI), complementing the console interface, with both interfaces served by the same engine, and with parallel functionality accessible through both

2) Export & Processing Module: additional data storage options (MongoDB, SQL) and module for processing (data curation with MassMine storing complete raw data which includes useful and extraneous data attached) and exporting data (CSV, TXT, XLSX) for import into various software as needed for the research goals (e.g., IBM's SPSS, MS Excel, WordSeer, MALLET)

3) Full Training Program: Planned training sessions include in-person trainings at UF and Clemson and online webinars, as well as guides, documentation, and resources for enabling data research (e.g., how to build data collections, process and export the data for analysis). All training sessions will include discussion of critical and ethical considerations for data research with socio-technical concerns related to protection of human subjects for data privacy and IRBs. Planned training sessions will include:

- Installing MassMine on Servers: Training for infrastructure providers on MassMine server installation for individuals and groups of researchers will include technical aspects and discussion on central service provider needs to support MassMine and data research
- Installing and Using MassMine for Humanities Data Research:
  Session One: Installing and using MassMine on local computers
  Session Two: Using MassMine hosted by UF Research Computing
  Session Three: Developing research questions, project scope, and goals for using MassMine (individually and collaboratively); training will include software and methodological assistance, discussion of data acquisition strategies for statistical needs, and intellectual goals
- MassMine for Humanities Data Research: Session without installation; will cover using MassMine on UF Research Computing servers, developing research questions, project scope, and goals for using MassMine individually, review of software and methodological concerns including data acquisition strategies for statistical needs and intellectual goals, and other supports
- MassMine for Teaching Data Research in the Humanities: Will build from prior session, providing an overview on and considerations for teaching with MassMine in the classroom
- MassMine for Data Research for Children's Literature: Version of training for Data Research in the Humanities will focus on Children's Literature for gathering feedback on supporting a specific area for focused research needs and as these needs inform larger concerns
- Advanced Data Training Session(s): To be planned in consultation with Scholars Group. Possible topics: methods for inspecting and querying collected data, tools and procedures for exploratory data analysis, hypothesis-driven descriptive and inferential statistical investigations

Staff

An Advisory Board comprised of Humanities Scholars from several institutions will contribute expert guidance on all MassMine project activities.

Contributed Cost Share: Project Participants from the University of Florida

Sidney I. Dobrin, PhD, Project Director, Research Foundation Professor of English, UF (.20 FTE cost share, totaling $24,037). Project Role: Dobrin will guide the overall project, support communication among Scholars Group and Advisory Board members, develop partnerships with other universities for central service support, and collaborate on outreach to specific research communities in the humanities.

Laurie N. Taylor, PhD, Project Co-Director, Digital Scholarship Librarian, UF (.10 FTE cost share, totaling $9,034). Project Role: Taylor will guide the project and focus on the training program for trainings and materials.

Matthew Gitzendanner, PhD, Project Co-Director, Biological Data Scientist, Research Computing Training Coordinator, UF (.06 FTE cost share, totaling $5,825). Project Role: Gitzendanner will guide the project, support training for central service providers and the Scholars Group using MassMine hosted by UF Research Computing, and provide expertise on developing software for research.

Suzan Alteri, MLIS, Curator of the Baldwin Library of Historical Children's Literature, UF (.05 FTE cost share, totaling $3,368). Project Role: Alteri will liaise with Children's Literature scholars for supporting MassMine training, research question development, reference consultations, and collaboration with other institutions.

Melissa Clapp, MLIS, Humanities Librarian & Digital Humanities Library Group (DHLG) Scholars Studio Coordinator, UF (.05 FTE cost share, totaling $4,013). Project Role: Clapp will liaise with the DHLG to support MassMine training activities, reference consultations, and collaborations with humanities librarians at other institutions and for ongoing humanities data research support.

Richard Freeman, PhD, Anthropology Subject Specialist Librarian & Digital Humanities Library Group (DHLG) Member (.05 FTE cost share, totaling $3,687). Project Role: Freeman will collaborate with the other project team members to provide and coordinate the Digital Humanities Library Group's support for training and outreach activities, including liaising support with scholars in the social sciences and humanities.

NEH Funding Request: Project Participants

Aaron Beveridge, PhD Student & MassMine Developer, UF ($12.12/hour for 809 hours, totaling $10,189). Project Role: Beveridge, as the MassMine Project Assistant, will coordinate communication, schedule trainings, gather feedback, and serve as the Scholars Group's primary contact for developing research questions, using MassMine on UF Research Computing's servers, and testing the Export & Processing Module.

Nicholas Van Horn, PhD Student & MassMine Developer, Ohio State University (1 FTE grant funded, totaling $33,303). Project Role: Van Horn, as the MassMine project programmer, will be responsible for the Export & Processing Module, GUI, and adding any required data sources, features, or functions informed by the Scholars Group, and Advisory Board. Van Horn will also be responsible for managing version control, debugging, and updates through the GitHub system.

Final Products and Dissemination

Final products include releasing all code as open source on GitHub for the MassMine GUI and Export & Processing Module, as well as the full training program and outreach resources. Dissemination will include dissemination through the training program and by project participants. Dissemination will also include press releases on the tool and research projects using MassMine, email announcements to scholarly lists, trainings at THATCamps in 2015, Florida Digital Humanities Consortium events, and others. The Project Team will seek for MassMine to be included in courses at UF (e.g., Data Literacy Common Core course, all UF undergraduates), Clemson University, and other institutions. Possible funding sources for subsequent phases include NEH's Digital Humanities Implementation Grants, The Social Media Research Foundation, and others. MassMine code is open source, so any researcher or developer can access the code and submit revisions. Members of the Project Team plan to continue contributing code for research and teaching needs.

NATIONAL ENDOWMENT FOR THE Humanities

Budget Form

OMB No 3136-0134
Expires 7/31/2015

Applicant Institution: *University of Florida*
Project Director: *Sidney Dobrin*
Project Grant Period: *05/01/2015 through 04/30/2016*

click for Budget Instructions

	Computational Details/Notes	(notes)	Year 1	(notes)	Year 2	(notes)	Year 3	Project Total
			05/01/2015-04/30/2016		01/01/20__-12/31/20__		01/01/20__-12/31/20__	
1. Salaries & Wages								
(1) Nicholas Van Horn; Postdoctoral Associate	Temporary Software Developer	100%	$29,446	%		%		$29,446
(1) Aaron Beveridge; OPS-Library Outreach/Training	$12.12/hr X 31 hrs X 26.1 pay periods	100%	$9,807	%		%		$9,807
		%		%		%		$0
		%		%		%		$0
		%		%		%		$0
		%		%		%		$0
2. Fringe Benefits								
13.10%	$29,446 base		$3,857					$3,857
3.90%	$9,807 base		$383					$383
3. Consultant Fees								
								$0
4. Travel								
								$0
								$0
5. Supplies & Materials								
								$0
6. Services								

UF Research Computing	Matching Program		$3,200					$3,200
7. Other Costs								
								$0
8. Total Direct Costs	**Per Year**		**$46,693**		**$0**		**$0**	**$46,693**
9. Total Indirect Costs								
DHHS on 06/28/13	**28.5% Per Year**		$13,307		$0		$0	**$13,307**
10. Total Project Costs	(Direct and Indirect costs for entire project)							**$60,000**

11. Project Funding	**a. Requested from NEH**	Outright:	$60,000
		Federal Matching Funds:	$0
		TOTAL REQUESTED FROM NEH:	**$60,000**
	b. Cost Sharing	Applicant's Contributions:	$64,203
		Third-Party Contributions:	$0
		Project Income:	$0
		Other Federal Agencies:	$0
		TOTAL COST SHARING:	**$64,203**
12. Total Project Funding			**$124,203**

Budget Narrative

Salary & Wages plus Fringe (UF) – NEH Request ($46,693)
The project team plans to hire the two original MassMine creators: Nicholas Van Horn, Data Scientist and MassMine Developer, as a Postdoctoral Software Developer (1 FTE totals $33,303) to develop the Export & Processing Module, graphical user interface (GUI), and other enhancements and supports as identified by the Scholars Group, Advisory Board members, and other scholars in the humanities who will be providing feedback; and, Aaron Beveridge, PhD Student in English, as MassMine Project Assistant (809 hours at $12.12/hour, totals $10,189) to coordinate communication, schedule trainings, and serve as the Scholars Group's primary contact.

Services
UF Research Computing matching program ($3,200).

Salary & Wages plus Fringe (UF) – Contributed Cost Share ($49,963)
Cost share will be provided by Department of English key participants as follows: UF Research Foundation Professor, Sidney Dobrin (Project Director) (.20 FTE totals $24,037) will lead and guide the overall project, and pursue new partnerships and collaborations with universities and research communities in the humanities to promote and disseminate MassMine.

Cost share will be provided by Smathers Libraries key participants as follows: Digital Scholarship Librarian, Laurie Taylor (Project Co-Director) (.10 FTE totals $9,034) will collaboratively guide the overall project, focusing on implementing the full training program with all collateral resource development. Melissa Clapp (.05 FTE totals $4,013) will provide and coordinate the Digital Humanities Library Group's support for training and outreach activities. Suzan Alteri (.05 FTE totals $3,368) will provide and coordinate support for training and outreach activities with Children's Literature scholars. Richard Freeman (.05 FTE totals $3,687) will collaborate with the other project team members to provide and coordinate the Digital Humanities Library Group's support for training and outreach activities, including liaising support with scholars in the social sciences and humanities.

Cost share will be provided by Research Computing key participants as follows: Data Scientist & Research Computing Training Coordinator, Matthew Gitzendanner (Co-PI) (.06 FTE cost share, totals $5,825) will support users for MassMine hosted by UF Research Computing.

Indirect Costs (UF) – NEH Request ($13,307)
NEH funding is requested for IDC rate of 28.5% for UF as follows: $46,693 in base direct costs for the single project year.

Indirect Costs (UF) – Contributed Cost Share ($14,240)
This represents IDC for UF's contributed cost share of $49,963.

ORIGINAL

COLLEGES AND UNIVERSITIES RATE AGREEMENT

EIN: 59-6002052

ORGANIZATION:
University of Florida
Finance & Accounting Division
111 Tigert Hall - PO Box 113200
Gainesville, FL 32611-3200

DATE:06/28/2013

FILING REF.: The preceding agreement was dated 07/12/2012

The rates approved in this agreement are for use on grants, contracts and other agreements with the Federal Government, subject to the conditions in Section III.

SECTION I: INDIRECT COST RATES

RATE TYPES: FIXED FINAL PROV. (PROVISIONAL) PRED. (PREDETERMINED)

	EFFECTIVE PERIOD				
TYPE	FROM	TO	RATE(%)	LOCATION	APPLICABLE TO
FINAL	07/01/2010	06/30/2011	46.50	On-Campus	Organized Research
PRED.	07/01/2011	06/30/2012	46.50	On-Campus	Organized Research
PRED.	07/01/2012	06/30/2014	49.00	On-Campus	Organized Research
PRED.	07/01/2014	06/30/2015	50.00	On-Campus	Organized Research
FINAL	07/01/2010	06/30/2011	26.00	Off-Campus	Organized Research
PRED.	07/01/2011	06/30/2015	26.00	Off-Campus	Organized Research
FINAL	07/01/2010	06/30/2011	46.40	On-Campus	AREC (A)
PRED.	07/01/2011	06/30/2012	46.40	On-Campus	AREC (A)
PRED.	07/01/2012	06/30/2015	41.00	On-Campus	AREC (A)
FINAL	07/01/2010	06/30/2011	25.00	Off-Campus	AREC (A)
PRED.	07/01/2011	06/30/2015	25.00	Off-Campus	AREC (A)
FINAL	07/01/2010	06/30/2011	33.60	On-Campus	Other Spons Activity
PRED.	07/01/2011	06/30/2012	33.60	On-Campus	Other Spons Activity

ORGANIZATION: University of Florida
AGREEMENT DATE: 6/28/2013

TYPE	FROM	TO	RATE(%)	LOCATION	APPLICABLE TO
PRED.	07/01/2012	06/30/2015	28.50	On-Campus	Other Spons Activity
FINAL	07/01/2010	06/30/2011	26.00	Off-Campus	Other Spons Activity
PRED.	07/01/2011	06/30/2012	26.00	Off-Campus	Other Spons Activity
PRED.	07/01/2012	06/30/2015	25.00	Off-Campus	Other Spons Activity
FINAL	07/01/2010	06/30/2011	50.00	On-Campus	Instruction
PRED.	07/01/2011	06/30/2015	50.00	On-Campus	Instruction
FINAL	07/01/2010	06/30/2011	26.00	Off-Campus	Instruction
PRED.	07/01/2011	06/30/2015	26.00	Off-Campus	Instruction
PROV.	07/01/2015	Until Amended			Use same rates and conditions as those cited for fiscal year ending June 30, 2015.

*BASE

Modified total direct costs, consisting of all salaries and wages, fringe benefits, materials, supplies, services, travel and subgrants and subcontracts up to the first $25,000 of each subgrant or subcontract (regardless of the period covered by the subgrant or subcontract). Modified total direct costs shall exclude equipment, capital expenditures, charges for patient care, student tuition remission, rental costs of off-site facilities, scholarships, and fellowships as well as the portion of each subgrant and subcontract in excess of $25,000.

(A) Agriculture Research and Education Center and Florida Medical Entomology Lab within the Institute of Food and Agriculture Science.

ORGANIZATION: University of Florida

AGREEMENT DATE: 6/28/2013

SECTION I: FRINGE BENEFIT RATES**

TYPE	FROM	TO	RATE(%)	LOCATION	APPLICABLE TO
FIXED	7/1/2013	6/30/2014	17.10	All	Clinical Faculty
FIXED	7/1/2013	6/30/2014	26.30	All	Faculty
FIXED	7/1/2013	6/30/2014	33.30	All	TEAMS Exempt
FIXED	7/1/2013	6/30/2014	45.50	All	TEAMS Hourly
FIXED	7/1/2013	6/30/2014	18.10	All	Housestaff / Clinic Post Docs
FIXED	7/1/2013	6/30/2014	7.20	All	Grad Asst / Post Docs
FIXED	7/1/2013	6/30/2014	4.60	All	OPS / Temp Faculty
FIXED	7/1/2013	6/30/2014	1.60	All	Student OPS / FWSP
PROV.	7/1/2014	Until amended			Use same rates and conditions as those cited for fiscal year ending June 30, 2014.

** DESCRIPTION OF FRINGE BENEFITS RATE BASE:

Salaries and wages.

ORGANIZATION: University of Florida

AGREEMENT DATE: 6/28/2013

SECTION II: SPECIAL REMARKS

TREATMENT OF FRINGE BENEFITS:

The fringe benefits are charged using the rate(s) listed in the Fringe Benefits Section of this Agreement. The fringe benefits included in the rate(s) are listed below.

TREATMENT OF PAID ABSENCES

Vacation, holiday, sick leave pay and other paid absences are included in salaries and wages and are claimed on grants, contracts and other agreements as part of the normal cost for salaries and wages. Separate claims are not made for the cost of these paid absences.

OFF-CAMPUS DEFINITION: For all activities performed in facilities not owned by the institution and to which rent is directly allocated to the project(s), the off-campus rate will apply. Actual costs will be apportioned between on-campus and off-campus components. Each portion will bear the appropriate rate.

ORGANIZATION: University of Florida

AGREEMENT DATE: 6/28/2013

Fringe Benfits include: FICA, State Unemployment, Workers' Compensation, Retirement, Life Insurance, Health Insurance, Leave Cash Outs, Sick Leave Pool Payments, Clinical Disability Insurance and Parental Leave Program.

On or prior to June 30, 2011, equipment means an article of nonexpendable tangible personal property having a useful life of more than one year, and an acquisition cost of $1,000 or more per unit. Effective July 1, 2011, the defined acquisition cost is $5,000.

The rates contained in this Agreement reflect the combined cost of the University of Florida and The University of Florida Research Foundation, Inc., and will apply to grants and contracts awarded to the Foundation.

APPLICATION OF INDIRECT COST RATES TO DOD CONTRACTS/SUBCONTRACTS:

In accordance with DFARS 2231.303, no limitation (unless waived by the institution) may be placed on the reimbursement of otherwise allowable indirect cost rates incurred by an institution of higher education under a DOD contract awarded on or after November 30, 1993, unless the same limitation is applied uniformly to all other organizations performing similar work. It has been determined by the department of Defense that such limitation is not being uniformly applied. Accordingly, the following rates do not reflect the application of the 26% limitation on administrative indirect costs imposed by OMB Circular A-21.

TYPE	Effective Period	Rate(%)	Locations	Applicable To
FINAL	07/01/10-06/30/11	48.5%	On-Campus	Orgn Research
PRED.	07/01/11-06/30/12	48.5%	On-Campus	Orgn Research
PRED.	07/01/12-06/30/14	51.0%	On-Campus	Orgn Research
PRED.	07/01/14-06/30/15	52.0%	On-Campus	Orgn Research
FINAL	07/01/10-06/30/11	28.0%	Off-Campus	Orgn Research
PRED.	07/01/11-06/30/15	28.0%	Off-Campus	Orgn Research
PROV.	07/01/15-Until Amended	Use same rates and conditions as those cited for fiscal year ended June 30, 2015.		

NOTE: This agreement updates the Fringe Benefits Rates section only.

ORGANIZATION: University of Florida

AGREEMENT DATE: 6/28/2013

SECTION III: GENERAL

A. LIMITATIONS:

The rates in this Agreement are subject to any statutory or administrative limitations and apply to a given grant, contract or other agreement only to the extent that funds are available. Acceptance of the rates is subject to the following conditions: (1) Only costs incurred by the organization were included in its facilities and administrative cost pools as finally accepted: such costs are legal obligations of the organization and are allowable under the governing cost principles; (2) The same costs that have been treated as facilities and administrative costs are not claimed as direct costs; (3) Similar types of costs have been accorded consistent accounting treatment; and (4) The information provided by the organization which was used to establish the rates is not later found to be materially incomplete or inaccurate by the Federal Government. In such situations the rate(s) would be subject to renegotiation at the discretion of the Federal Government.

B. ACCOUNTING CHANGES:

This Agreement is based on the accounting system purported by the organization to be in effect during the Agreement period. Changes to the method of accounting for costs which affect the amount of reimbursement resulting from the use of this Agreement require prior approval of the authorized representative of the cognizant agency. Such changes include, but are not limited to, changes in the charging of a particular type of cost from facilities and administrative to direct. Failure to obtain approval may result in cost disallowances.

C. FIXED RATES:

If a fixed rate is in this Agreement, it is based on an estimate of the costs for the period covered by the rate. When the actual costs for this period are determined, an adjustment will be made to a rate of a future year(s) to compensate for the difference between the costs used to establish the fixed rate and actual costs.

D. USE BY OTHER FEDERAL AGENCIES:

The rates in this Agreement were approved in accordance with the authority in Office of Management and Budget Circular A-21, and should be applied to grants, contracts and other agreements covered by this Circular, subject to any limitations in A above. The organization may provide copies of the Agreement to other Federal Agencies to give them early notification of the Agreement.

E. OTHER:

If any Federal contract, grant or other agreement is reimbursing facilities and administrative costs by a means other than the approved rate(s) in this Agreement, the organization should (1) credit such costs to the affected programs, and (2) apply the approved rate(s) to the appropriate base to identify the proper amount of facilities and administrative costs allocable to these programs.

BY THE INSTITUTION:

University of Florida

(INSTITUTION)

Michael V. McKee

(SIGNATURE)

Michael V. McKee

(NAME)

Assistant Vice President and University Controller

(TITLE)

July 2, 2013

(DATE)

ON BEHALF OF THE FEDERAL GOVERNMENT:

DEPARTMENT OF HEALTH AND HUMAN SERVICES

(AGENCY)

Darryl W. Mayes

(SIGNATURE)

Darryl W. Mayes

(NAME)

Deputy Director, Division of Cost Allocation

(TITLE)

6/28/2013

(DATE) 0301

HHS REPRESENTATIVE: Steven Zuraf

Telephone: (301) 492-4855

Biographies

Advisory Board

An Advisory Board will contribute, at no cost to the project, expert guidance on the MassMine project activities including software, functionality, training program, documentation, and related products. The Advisory Board will be comprised of Humanities Scholars from several institutions who are investigating data research supports in the humanities. Advisors include: Kenneth Kidd, PhD, English and Center for Children's Literature & Culture, UF; Cathena Martin, PhD, Game Studies & Design, University of Montevallo; and Sean Morey, PhD, English, Clemson University. **Project Role:** Advisors will evaluate all technical features, GUI design, and the training program in terms of how all elements together support humanists in developing research questions and practices for data research. The Advisory Board's role is to: 1) provide guidance in framing humanities research questions and practices needing data research support; 2) provide expert perspectives about system functionality, interface design, and the MassMine training program; 3) recommend and select Scholars Group participants (Advisory Board members may also elect to serve); and, 4) promote MassMine to other scholars in their research communities.

NEH Funding Request: Project Participants

Nicholas Van Horn, PhD Student & MassMine Developer, Ohio State University, is the co-creator of MassMine along with Aaron Beveridge. He is an active researcher in the interdisciplinary field of computational cognitive neuroscience. Work in this area represents the convergence of advances in a number of related fields, including neuroscience, psychology and psychophysics, computer vision, artificial intelligence, mathematical modeling, as well computer science more broadly construed. His work on visual perception, learning, and memory has emphasized the collaborative nature of the field, resulting in many multi-author peer-reviewed publications in high-impact journals and talks/posters at top conferences. He has won multiple research awards for his work, and his mathematical and computer science expertise has enabled him to design and program many strictly controlled experimental protocols, including a project using the functional magnetic resonance imaging (fMRI) scanner at OSU to study patterns of activity in the brain during analogical reasoning. Further, his focus on computational modeling led to the development of several computer models of human memory, as well as a large-scale investigation of visual object recognition that compared performance of a biologically-inspired computer algorithm to the results of human performance from five behavioral studies. The work required many thousands of lines of custom software and was deployed on the Linux cluster at the Ohio Supercomputer Center, the results of which led to a published journal article and talk at the Vision Sciences Society, the premier vision conference in North America. In parallel with this research, he actively writes and maintains open source software related to writing and productivity, statistical analysis, and other core software functionality such as tools for automated text processing. This work in part led to the development of the current functionality of MassMine, for which he remains the lead software developer.

Aaron Beveridge, PhD Student, MassMine Developer, English Department, UF, is the co-creator of MassMine along with Nicholas Van Horn. His research investigates the intersection of data science and humanities research paradigms--focusing on the importance of tool creation and software development as they motivate the ongoing expansion of available research methodologies in the digital humanities. With an emphasis on writing studies and circulation studies, his current project tests theories of hyper-circulation as they attempt to explain the delivery and recirculation of digital media within complex networks. He presented his work with MassMine at the largest international conference for writing studies, rhetoric and composition, the Conference on College Composition and Communication (CCCC 2014), and he has been accepted to present an update of the software along with new data analysis at the same conference in 2015. He is currently working with the George A. Smathers libraries at UF to develop trainings for the Scott Nygren Scholars Studio (a Digital Humanities lab) to teach Arduino

microcontroller programming/prototyping, and an additional set of trainings that teach text mining and natural language processing.

Contributed Cost Share: Project Participants from the University of Florida

Sidney I. Dobrin, PhD, Project Director, Research Foundation Professor of English currently serves as Graduate Coordinator for the Department and for ten years served as Director of Writing Programs in the English Department. Dobrin is the Founding Director of Trace Innovations Initiative, an online hub for research in media ecology, technology, and writing. Dobrin has published seventeen books about writing, technology, ecology, and media. He continues to publish in these research areas and anticipates the release of three new books this year with others to follow. His 2011 book *Postcomposition* received the W. Ross Winterowd Award for best book published in composition theory. Dobrin is frequently an invited, keynote, and plenary speaker at conferences and universities, both internationally and domestically.

Laurie N. Taylor, PhD, Project Co-Director, Digital Scholarship Librarian, conducts research to create scholarly cyberinfrastructure through data/digital curation, digital scholarship, while developing socio-technical supports (people, policies, technologies, communities) to create, sustain, and integrate digital scholarship and data curation across communities, and fostering an environment of radical collaboration made possible in the digital age or the age of Big Data. She works heavily with the Digital Library of the Caribbean (dLOC, serving as technical director for this international collaborative), UF Digital Collections, UF Research Computing, and the SobekCM Open Source software community. She has been the principal investigator, co-PI, and investigator on many grants, including co-principal investigator on the ARL PD Bank, a digital scholarship project to centrally collect academic library job position descriptions for immediate and long-term analysis, and planning to meet needs related to future changes in academic libraries in the digital age. Her teaching and training spans undergraduate Digital Humanities courses, graduate writing courses, and workshops on digital technologies. She has published refereed articles on data curation, digital scholarship, collaborative international digital libraries, library and information science, digital media, open access, and literature; and she co-edited a collection on digital representations of history and memory, *Playing the Past: Video Games, History, and Memory.*

Matthew Gitzendanner, PhD, Biological Data Scientist, Research Computing Training Coordinator, coordinates the Research Computing training program, provides expert support as a Bioinformatics Specialist for Research Computing users, conducts research as a research faculty member in the Biology Department, teaches computational biology courses for undergraduate and graduate students, and develops software for scholarly research. His research spans a broad array of topics generally related to evolutionary genomics, with topics ranging from population and conservation genetics to genomics and bioinformatics.

Suzan A. Alteri, MLIS, Curator of the Baldwin Library of Historical Children's Literature, UF, conducts research on the materiality of the book, special collections in the classroom, and research on historical religious tracts. She works with the newly created Baldwin Library Scholars Council to determine grant proposals, and publication opportunities for both graduate students and researchers working with books from the Baldwin Library. She was the principal investigator on the grant Forging a Collaborative Structure for Sustaining Scholarly Access to the Baldwin Library and is a project team member on the 'Developing Librarian': Digital Humanities Pilot Training Program. She has presented on "The Little Golden Books," "Digital Curation," "Introduction to the Baldwin Library Digital Collection," "Digital Collections and Foundations," and on "Digital Collections and Scholarship." In addition, Suzan has curated these exhibits: *Bigger, Better, Best: the Panama Canal in Children's Literature, When Phantasie Takes Flight: the Art & Imagination of Arthur Rackham,* and *Grimm Changes* on the work of the Brothers Grimm over time. Her most recent publication, "The Classroom as Salon: a Collaborative

Project on Daniel Defoe's *Robinson Crusoe*" appeared in *Digital Defoe: Studies in Defoe & His Contemporaries*. She regularly liaises with the Department of English on campus, which includes the Children's Center for Literature and Culture.

Melissa Clapp, MLIS, Humanities Librarian & Digital Humanities Library Group (DHLG) Scholars Studio Coordinator, is the Instruction & Outreach Librarian for the Humanities & Social Sciences. She joined the faculty of UF in 2007. Her research interests include digital humanities, research practices of students, and learning methods. She holds a Master of Information Studies degree from Florida State University and MA in English from Northern Illinois University.

Richard Freeman, PhD, Anthropology Subject Specialist Librarian & Digital Humanities Library Group (DHLG) Member, is presently working on two digital projects with UF faculty members. One project is working with a body of historical photographs of the construction of the transcontinental railroad. The second is creating new visual content for the digital collection entitled: "Vodou Archive" housed within the Digital Library of the Caribbean (dLOC) in the UF Digital Collections. Freeman also worked as an archivist at the National Gallery of Art in Washington D.C. and as an assistant professor of anthropology. He has made numerous presentations at conferences and has several publications on the culture of Argentine politics and visual anthropology. He is currently working on a paper about the digital photographs project and a chapter for an edited volume on Haitian Vodou ceremonies. He is also an active member of the DHLG and is presenting on building support for the digital humanities in libraries at the 2014 conference for the Florida Chapter of the Association of College & Research Libraries (FACRL).

Data Management Plan

MassMine is being developed specifically to support data research needs in the humanities. This includes the ability to access and engage with all levels of tools and data research. Open source code is essential to support external review for reproducible research, support ongoing open development to support data research in the humanities, and enable and foster collaboration among humanists for data research. In developing MassMine, one of the Project Team goals is for MassMine to exemplify open and collaborative approaches for software development and training in the process of improving access to data research. The same overall alignment will be used in making all technical decisions including those related to the GUI for MassMine. Like MassMine's other components, the GUI will be based on open standards and compliant code to support use on any operating system with an open standards compliant web browser (Windows, Apple, Linux). MassMine code is already publicly available through GitHub and will be released to GitHub on an ongoing basis. MassMine is released under the GPL license as open source for download by anyone. Using GitHub others can also "fork" a copy of the code. Forking is the term for creating a new version of the software where developers can continue development on a separate trajectory to submit new changes and additions to the software. Versioning and debugging will be controlled through GitHub's update/submissions system, and new changes will be developed and released through that same system under the supervision of the MassMine team at UF.

For success, all materials for the project need to be shared openly and as widely as possible. The investigators commit to openly sharing all data in a timely manner. The proposed MassMine project focuses on software development and training which do not involve any private or otherwise restricted data, and do not involve any data that would present a risk to disclosure. The team does not anticipate any privacy issues, ethical issues, or intellectual property issues. Because MassMine enables other research projects, for the research data collections created by MassMine which could potentially have privacy and other concerns, the project training program will explicitly include data privacy and IRB approvals as supporting resources for data research.

In addition to MassMine code on GitHub through regular releases, each major release version of the code also will be archived to the broadly accessible IR@UF. Project documentation, tutorials, and training materials will be hosted in the IR@UF. Materials will include documentation, project examples, sample data sets, guides on additional resource articles and related open source analysis software, etc.

The Smathers Libraries at UF commit to archiving and making materials accessible on an ongoing basis and at project end. This is in keeping with normal practices of the Libraries' commitment to open and expedient dissemination of grant products and grant materials (e.g., "Unearthing St. Augustine" grant materials) to support research needs and to assist in building a culture of grantsmanship. UF dedicates staff time to digital preservation and access from the Digital Production Services staff, IR@UF Manager, Digital Development & Web Services Team, Digital Librarian, and others.

The project will generate a variety of data materials, with the majority being code, training resources, and documentation. Specific forms include: whitepaper, planning materials, reports, webinar videos, training materials, and meeting notes. Programming for MassMine was developed in and will continue to use the R language as the underlying technology for MassMine. R is an open source language, as are all of the development environments for coding in R, so technical resources for the programming and software development of MassMine are freely available and well supported. Documentation will be embedded in source code, in separate ASCII files (e.g., plain ASCII, Asciidoc, HTML, XML) and/or in formatted files (e.g., PDF, DOCX, PPTX). Training and support materials will be stored in standard formats (e.g, HTML, PDF, AVI, PPTX, etc.). Researcher datasets and accompanying files will be made available in their original and normalized formats (brief list of selected, recommended formats).

The Project Team will use GitHub for sharing code and code documentation, with all data openly available for anyone through GitHub. For permanent and findable support, all grant data materials will be openly accessible and preserved in the IR@UF, powered by the SobekCM software, which provides metadata for all materials (at the item, group, and aggregation levels), permanent identifiers and URLs, multiple file formats and digital object packages (preservation and access copies), and more. All materials for this project will be openly accessible and will be made available as soon as possible, with the supporting metadata for findability and usability, with all project data made available at minimum twice each year and the majority of the project data made available in regular releases each week or more frequently.

The Libraries are committed to long-term digital preservation of all materials in the UF Digital Collections (UFDC), including the IR@UF, and in UF-supported collaborative projects as with the Digital Library of the Caribbean (dLOC). Redundant digital archives, adherence to proven standards, and rigorous quality control methods protect digital objects. Through UFDC, the Libraries provide a comprehensive approach to digital preservation, including technical support, reference services for both online and offline archived files, and support services by providing training and consultation for digitization standards and long-term digital preservation. The Libraries support locally created digital resources as powered by and hosted with the SobekCM Open Source Repository Software, including the UFDC which contains over 381,000 digital objects with over 30 million files (as of February 2014). The Libraries create METS/MODS metadata for all materials. Citation information for each digital object also is automatically transformed by the SobekCM software into MARCXML and Dublin Core. These records are widely distributed through library networks and through search engine optimization to ensure broad public access to all online materials.

In practice consistent for all digital projects and materials supported by the Libraries, redundant copies are maintained for all online and offline files. The digital archive is maintained as the Florida Digital Archive (FDA) which was completed in 2005 and is available at no cost to Florida's public university libraries. The software programmed to support the FDA is modeled on the widely accepted Open Archival Information System. It is a dark archive and supports the preservation functions of format normalization, mass format migration and migration on request. As items are processed into the UFDC for public access, a command in the METS header directs a copy of the files to the FDA. The process of forwarding original files to the FDA is the key component in UF's plan to store, maintain and protect electronic data for the long term. If items are not directed to load for public access, they do not load online and are instead loaded directly to the FDA (*more information*).

UF UNIVERSITY of FLORIDA

College of Liberal Arts & Sciences
Department of English

4008 Turlington Hall
PO Box 117310
Gainesville, FL 32611
352-392-6650
352-392-0860 Fax

September 1, 2014

Dr. Sidney I. Dobrin
Department of English
CAMPUS

Dear Sid,

I am writing to confirm my commitment and participation as a member of the Advisory Board Team for your proposed project "MassMine: Collecting and Archiving Big Data for Social Media Humanities Researchers." This is an exciting project with great potential significance for interdisciplinary research, especially as the humanities embrace empirical and quantitative methods of research and knowledge production. I am a scholar of children's literature, with particular interests in canon and field construction and histories of the children's archive, so this project is of particular interest to me. This semester, for example, I am teaching a graduate seminar on the children's literature archive, drawing on our preeminent Baldwin Library of Historical Children's Literature. In that class we are reading Moretti's *Graphs, Maps, and Trees*, an exploration of quantitative research for the humanities, and students will be conducting various experiments in data mining and analysis, albeit on a smaller scale. The opportunities for research when it comes to current children's literature and children's media are also exciting, especially since so far the conversation about children's media has been dominated by researchers outside the humanities. Children's literature scholars are just now beginning to turn to social research methods, and the MassMine project could greatly enhance the collective sense of possibilities. Childhood studies, moreover, is on the rise as an interdisciplinary field, and the MassMine project can both draw from and extend that field's range and import. My sense is that the MassMine project has the potential to transform not merely how we conduct our work but also our understanding of what that work actually is, or could be. I look forward to working more closely with you as this project develops.

Sincerely,

Kenneth Kidd
Professor and Chair
kbkidd@ufl.edu

UNIVERSITY of MONTEVALLO

Game Studies and Design
Hill House, Station 6501
Montevallo, AL 35115
T. 205.665.6501
Cmartin16@montevallo.edu

September 9, 2014

Dear Dr. Sid Dobrin,

I am writing to confirm my commitment and participation as a member of the Advisory Board Team for your proposed project on "MassMine: Collecting and Archiving Big Data for Social Media Humanities Researchers." As a tenure-track assistant professor of Game Studies and Design, my teaching and research areas include a variety of game categories, including board, card, video, and tabletop role-playing games. My gaming emphasis comes out of a larger study of children's literature and culture. Additionally, I am also the Director of the Honors Program at the University of Montevallo and support digital humanities projects with my Honors faculty, and have worked with our QEP librarian to incorporate information literacy into Honors courses.

I support MassMine for my research and for the good of my colleagues. MassMine will help provide faculty with an easy interface with which to do social media data mining. At a small, public liberal arts university such as mine, with no computer science program, we need access to training and user-friendly platforms that aid our online research. This need is true across COPLAC institutions and beyond. My current research is on tabletop role-playing games, and I could use MassMine to assess the sociological impact of these types of games within the United States and determine how the narrative revolving around these types of games have shifted since the late 70s in the public perception. But this software can support a large variety of game focused projects, such as has already been demonstrated with Kyle Bohunicky's Game Studies and Cultural Preservation project.

Sincerely,

Cathlena Martin, PhD
Assistant Professor
Coordinator of Game Studies and Design (GSD)
Director of the Honors Program

CLEMSON UNIVERSITY

DEPARTMENT OF ENGLISH

Clemson University
816 Strode Tower
Clemson, SC
29634

P 864-656-3193
F 864-656-1345

September 1, 2014

4008 Turlington Hall
P.O. Box 117310
Gainesville, FL 32611-7310

Dear Professor Dobrin:

I am writing to confirm my commitment and participation for your proposed project on "MassMine: Collecting and Archiving Big Data for Social Media Humanities Researchers." I am currently a member of the Advisory Board.

Broadly, my general areas of expertise include Digital Media, Digital Humanities, Environmental Humanities, Technology Studies, and Writing Studies. This project is important to my research as it provides a cutting-edge and robust platform that will allow me to use new digital tools toward data-mining methodologies, helping me to investigate many of the questions I am currently exploring, especially as related to how the intersection of social media, digital writing, and emerging technologies forges new identities of nature and environment.

My own research aside, the open-source and collaborative nature of the MassMine platform will provide many other scholars engaged in Digital Humanities with a new tool that will help them perform their own research. In my view, MassMine is highly adaptable and can be used for many social media, big data, and digital humanities projects.

As a scholar working in these areas, I find your work with MassMine impressive, important, and necessary given the current expansion and interest in the digital humanities and related fields (which I argue includes social media, big data, and digital archival research). As part of my commitment to this project, I will dedicate time and expertise to the project to help it succeed.

Thank you very much for the opportunity to serve on the advisory board, and please do not hesitate to contact me if you have any further questions.

Sincerely,

Sean Morey
Assistant Professor

www.clemson.edu/caah/english

September 6, 2014

Dr. Sidney Dobrin
Research Foundation Professor
Department of English, University of Florida
PO Box 117310
Gainesville, FL 32611-7310

Dr. Dobrin,

I am writing to express my commitment to accept the postdoctoral position as described in your proposed project on "MassMine: Collecting and Archiving Big Data for Social Media Humanities Researchers" should it be accepted by the National Endowment for the Humanities. I will eagerly satisfy all responsibilities as they are outlined in the proposal budget and narrative, and will gladly offer my time and effort to further advance the project.

My work in computational modeling of cognition, with its inherent technological and data-intensive difficulties, has familiarized me with the benefits and challenges of large data sets. Together with my research and teaching in the social sciences, I am intimately aware of the need of not only systems for managing the burgeoning world of large and complex data sources, but also for infrastructure for facilitating the training and collaboration of researchers. Once realized, I strongly feel that the current proposal will address these concerns as they apply to scholars in the humanities.

Finally, as co-creator of the MassMine software package, I am highly invested in its success and ensuring that it meets and exceeds the description provided in your project outline. I am quite confident in the personnel you have recruited, and I believe the project is worthy of consideration by the NEH.

Sincerely,

Nicholas Van Horn
Adjunct Professor
Department of Psychology, Capital University
nvanhorn@capital.edu

MICHIGAN STATE
UNIVERSITY

Dear NEH-ODH Review Committee Members ,

I write to express my strongest support for the *MassMine Implementation: Collecting and Archiving Big Data for Social Media Humanities Researchers* project proposed by the team at the University of Florida. As a humanist scholar who relies on social media sources for data, I can attest to the difficulty cited by the MassMine team in accessing, archiving, and querying social media posts. As humanists, we often want full text of posts as well as metadata for analysis, requiring either tedious collection of information by hand or cost-prohibitive access to proprietary datastores. With a resource like MassMine, access to multiple sources of social media postings and associated metadata will be less expensive in terms of time and/or money to achieve and much, much easier to store than it is today.

Let me offer a detailed example of how I might use a tool like MassMine in the service of one of our ongoing projects here at Michigan State University's Writing in Digital Environments and MATRIX research centers. The project gathers posts from social media users that mention words and phrases linked with foodborne illness, a controlled list used by state and local health department workers around the country.

A support vector machine-based classifier then analyzes the natural language as well as post metadata – location, date, etc. – to assign each post a score. We then use the score to weigh the available evidence indicating where there may be outbreaks of foodborne illness that bear further investigation by local health officials. The reasoning done by our system is probabilistic, meant to complement other means of disease surveillance and response while decreasing the incidence of severe illness from outbreaks that are detected too late. We call it a "*stasis* engine" because it weighs evidence using the classical categories from rhetorical *stasis* theory.

A major hurdle in the development of our work on this project has been access to the places where people share updates with information about their foodborne illness symptoms. What is "oversharing" to many is data for us, whether it is in a time/space window known to correspond to an outbreak (e.g. in archived data) for testing or in "live stream" data corresponding to a particular geographical region for live or real-time alerts. We have had to gather our data for training and testing the system by hand, and it has taken many hours and many thousands of dollars in salary to do so. A tool like MassMine would make work like ours much, much easier!

Beyond my own personal interest in a system like MassMine, I see tremendous value in it for historians, anthropologists, and researchers doing work in cultural studies and popular culture just to name a few distinct areas. Consider the value of using a tool like MassMine to preserve the *zeitgeist*, reflected in the activity of various social media platforms, associated with a particular cultural event or phenomenon. Recently, for instance, both social media and traditional media outlets have reported on the ways that different social platforms such as Twitter and Facebook reflect very different contemporaneous portraits of American culture. In the late Summer of 2014, at the same time that the unrest and controversy in Ferguson, MO was trending on Twitter, folks were dumping ice water on their heads in support of the ALS Association over on

College of
Arts & Letters
Associate Dean for
Graduate Education

479 W. Circle Drive
Linton Hall
East Lansing, MI 48824

Telephone: (517) 353-6720
Fax: (517) 355-0159

hartdav2@msu.edu

Facebook. Using a tool like MassMine, researchers could collect and compare social trends like these at a scale and at a level of detail that is currently not feasible.

I have confidence that the MassMine team can make excellent use of the resources provided by the NEH to take a tool currently useful for a handful of researchers at the University of Florida and develop an open-source tool for humanities researchers around the country and the world. The project plan they propose is solid and achievable, as is their plan for dissemination, testing and training. Moreover, MassMine is the sort of project that is useful enough to a broad enough group of tech-savvy researchers that it will likely attract a large developer community. This is important whenever a team proposes an open-source project, because the biggest risk is that there will be insufficient development effort and know-how available among the core user community to take the project forward once it has been given an initial boost. I do not see this problem with MassMine.

I could not be more excited to see MassMine move foreward. I think it has tremendous potential to serve many researchers in ways that are not even fully obvious until the tool finds its way into the hands of creative humanist scholars. I wholeheartedly support the project, and I've let Dr. Dobrin and the others on the team know that I'm happy to be an early test user, an early adopter, and an advisor on the project if they find my assistance at all useful. Having built both commercial and open-source systems myself, I believe this project is as safe a bet at the proposal stage as any I have seen.

Sincerely,

William Hart-Davidson, Ph.D.
Associate Professor, Writing Rhetoric & American Cultures
Associate Dean for Graduate Studies, College of Arts & Letters
Co-Director, Writing in Digital Environments Research Center @ MATRIX

UNIVERSITY OF KENTUCKY

Division of Writing, Rhetoric, and Digital Media
1353 Patterson Office Tower
Lexington, KY 40506-0027
859 257-7002
fax 859 323-1072
wrd.as.uky.edu

August 15, 2014

Colleagues:

I am writing in support of the MassMine collaborative project at the University of Florida.

Over the last several years, Humanities scholarship has demonstrated increasing interest in the generic appellation "big data." Rather than limit scholarship to textual analysis or textual interpretation (in a text or in a cultural moment) as has often been the case, scholarship now explores the ways patterns, trends, ideas, beliefs, and so on can be identified by building significant data sets out of the traditional objects of scholarly focus. Big data has come to mean all of the information available for scholarly pursuit and the difficulty encountered in isolating such data and using it for research means. Data, we have come to realize, needs to be mined.

This work has proven to be essential to Humanities scholarship as scholars begin to understand that in order for analysis to be effective, it depends on large sets of information, some of which cannot be fully accessed without the available tools of data collection (i.e. software). National scholars such as Stephen Ramsey and Lev Manovich have repeatedly demonstrated the need to situate Humanities scholarship in terms of data and data mining. Whether through locating trends over bodies of work or periods of time or whether via visualizing trends or concepts, both of these kinds of data mining projects are redefining Humanities scholarship in quantitative ways. In turn, academic work is rethinking how search works – from popular, online portals such as Google to more nuanced algorithms and collections that challenge the ways we assemble information for various purposes. At the heart of all contemporary scholarship is search and how the data discovered in search is used.

The MassMine project situates itself as another important contributor to this kind of academic work and promises to be a leader in the field. In addition to tapping into traditional sources of information, MassMine promises to utilize open APIs in order to incorporate less obvious sources of information, such as Twitter, Facebook, and everyday social media platforms. By doing so, MassMine promises to generate a more complete network of intellectual and popular interactions that can greatly aid in various areas of scholarly research.

As this grant proposal makes clear, however, the collection of data is often difficult. Not only does appropriate software need to be designed for specific purposes, training must take place as well. As an open source platform without coding requirements, MassMIne is designed to

ease the process as well as provide the necessary training for students and faculty to effectively use the software. MassMine taps into the open nature of APIs in order to utilize the Web and many social media platforms' emphasis on large scale integration and development. In that sense, it is accessible financially (it is free and open source) and pedagogically (it will be accompanied by appropriate training and documentation).

We might consider this proposal in contrast to traditional university Humanities research tools like simple library search, which often scans subscription portals one object at a time. A basic library search is not a data driven search since it only returns exact or near exact matches that correspond to the university's paid holdings. That is, the typical university database is limited in how it returns responses or allows a researcher access to different patterns of information. We might then consider a data driven search, such as that proposed by MassMine, that would explore a number of texts simultaneously, within a library's collection and elsewhere, identifying potential relationships, including relationships not yet visible or obvious. For instance, contemporary approaches to search will often reveal one dominant narrative regarding an issue (whether that issue is political, social, or economic) that attempts to speak for the moment in its entirety. A broader, data driven search that is not limited in scope or textual analysis might, instead, turn up a number of narratives or patterns that challenge dominant thinking on the subject or allow for new insights the researcher could not obtain via a typical library search.

With this last point, this is where I see MassMine playing a significant role in our current period of data and digital search. MassMine can alter the teaching and performance of university scholarship at multiple levels and for faculty and students. The more information we can access and work through, the better we become at understanding broader implications of phenomenon and their meanings. The more information we can access and work through, via an application like MassMine, the better we are to approach complex narratives and situations. If MassMine were available at the University of Florida, I would ask undergraduate students I work with to use it from the University of Kentucky, where I teach. I see its contribution as an incredibly valuable addition to the work we do in the Humanities. For too long, we have taught our students limited search skills via the very limited tools that they have access to. My own scholarship has attempted to theorize such shifting search strategies and the frustrations with pedagogies designed in opposition to digital search. With MassMine, teaching can greatly shift to include strategies appropriate to the digital age we work within.

Division of Writing, Rhetoric,
and Digital Media
1353 Patterson Office Tower
Lexington, KY 40506-0027
859 257-7002
fax 859 323-1072
wrd.as.uky.edu

Because MassMine will be housed at the University of Florida, it is important to note how it can serve not only the land grant mission of the university, but the broader mission of higher education: to allow continued access to information for intellectual and research purposes. Given many private efforts to limit information accessibility, MassMine will be extremely important to scholarship. As a public university project, MassMine will benefit researchers and students in a much broader way than other applications (many of which limit access via fees or other restrictions).

I strongly recommend that the NEH support this application. MassMine is exactly the kind of tool we need in contemporary research and scholarship.

Sincerely,

Jeff Rice
Martha B. Reynolds Professor in Writing, Rhetoric, and Digital Studies
University of Kentucky
j.rice@uky.edu

Appendix: Detailed Work Plan

Pre-Startup Phase: 2014

- April: MassMine 0.1.0 released on GitHub (Van Horn, Bevridge)
- April: Presentation on MassMine at THATCamp-Gainesville (Beveridge)
- Fall: Capital University, course on data science using MassMine (Van Horn)
- Nov. 13: Training on using MassMine on UF Research Computing Servers (Beveridge)
- Facebook and Wikipedia APIs added as data sources by January 2015 (Van Horn)

May-June 2015: Scaling for Many Users: Server Installation Scaled; Updates for More Backend Supports

- Confirm project charter with graduate student and postdoctoral researcher mentoring plans and user support processes for Scholars Group and all users (Project Team)
- Software: Share on ongoing basis on GitHub and with major releases to the IR@UF (Van Horn)
- Software: Implement storage support options for SQL and MongoDB; begin developing Export & Processing Module; initial development activities for adding the GUI (Van Horn)
- Collaborate with Advisory Board to identify potential Scholars Group members (Project Team)

July 2015

- Software: Develop specifications to add data cleaning and subsetting functionality (Van Horn)
- Develop initial training program materials, user testing and feedback processes, and documentation for iterative development for new features, functions, GUI display, and related socio-technical workflows and data practices (Project Team)
- Training Session on Installing MassMine on Servers: for infrastructural service providers (e.g., central computing units, research groups/units/departments) for MassMine server installations supporting individuals and groups of researchers. Training to include: technical aspects, discussion and planning for supporting socio-technical processes for data privacy needs and integrating workflows with IRBs, and feedback from central service providers on their needs to support MassMine and data research (Beveridge, Dobrin, Gitzendanner)

August 2015

- Confirm Scholars Group (5-7 humanities scholars from UF and Clemson); ensure access to MassMine hosted on UF Research Computing (Project Team)
- Schedule training sessions for in-person and online webinars in consultation with Scholars Group and specific additional user groups; begin promoting training sessions (Project Team & DHLG)
- Software: Release beta Export & Processing Module with data curation functionality (e.g., remove unnecessary HTML, markup, irrelevant URLs, spurious symbols or punctuation, non-specific data attached to API access) and with export to CSV and XLSX (Van Horn, Beveridge)
- Update documentation and training materials for installing MassMine, using Export & Processing Module, and submitting bugs and feature requests through GitHub as tailored to various user groups including those new to GitHub (Project Team)

September 2015: Training Sessions with Scholars Group (Project Team, DHLG)

- Trainings with humanities scholars for their data research projects, feedback on MassMine functions and resources, and establishing overall collaborative development processes
- Session One (in person, webinar): Installing and using MassMine on local computers
- Session Two (in person, webinar): Using MassMine on UF Research Computing systems
- Session Three (in person at UF and Clemson): Developing research questions, project scope, and goals for using MassMine (individually and collaboratively) with software and methodological assistance, discussion of data acquisition strategies for statistical needs, intellectual goals, and socio-technical concerns with data privacy, IRBs, and other supports (session feedback to create project examples for future online trainings)
- Liaise with Scholars Group (users on UF Research Computing or local computers) for initial data collection for projects; assistance as-needed; documenting discussions to inform development

October 2015: Collaborative Development and Iterative Testing (Project Team)

- *Scholars Group data collection ongoing with training and support for project duration*
- Liaise with Scholars Group on inspecting data output including checking for inconsistencies/abnormalities, supporting technical needs for computing descriptive statistics on initial data as an added check, and gathering feedback on user experience
- Liaise with Scholars Group for iterative testing processes on technical functionality and instructions on preliminary data inspection with examples in common analysis software (Excel, SPSS, R) regarding the analysis and interpretation of data; as well as instructional support on data research methods including identifying measurable variables, approaches to "existing data" and correlation designs, and core techniques of descriptive statistics

November 2015: Ongoing Collaborative Development and Iterative Testing (Project Team)

- Software: Finish software development updates identified by Scholars Group; release updated Export & Processing Module (Van Horn)
- *Scholars Group data collections will now include large data tables of information, well suited to and needing methods for grouping and narrowing of search for variables of interest*
- Liaise with Scholars Group for analysis and data visualization software needs with the Export & Processing Module

December 2015: Iterative Testing and Development

- Software: Update the Export & Processing Module for functionality and tool design based on feedback from Scholars Group (Van Horn)

January 2016: GUI Alpha Release; Ongoing Collaborative Development & Iterative Testing

- Software: Release MassMine GUI alpha version; finish updated Export & Processing Module; complete any remaining software updates for GUI standalone installations and as identified by Scholars Group data collection (Van Horn)
- Liaise with Scholars Group to support installing GUI software version; gathering their feedback on the GUI version (Project Team)

February 2016: Finalize GUI Release, Documentation and Resources, Advanced Data Trainings

- Create documentation for MassMine GUI interface (Project Team, DHLG)
- Scholars Group reviews documentation, provides feedback, suggestions for resources related to their research questions and areas (Project Team, DHLG)
- Advanced Data Training Sessions: Liaise with Scholars Group on needed trainings and conduct training sessions, possible topics may include methods for inspecting and querying collected data, tools and procedures for exploratory data analysis, hypothesis-driven descriptive and inferential statistical investigations (Project Team, DHLG)

March 2016: Training Sessions for New Users and Scholars Group (in person and online)

- MassMine for Humanities Data Research: Session on using MassMine; covers developing research questions, project scope, goals for using MassMine individually, review of software and methodological concerns including data acquisition strategies for statistical needs and intellectual goals, socio-technical concerns with data privacy, and IRB (Project Team, DHLG)
- MassMine for Teaching Humanities Data Research: Session builds from MassMine for Humanities Data Research, covers examples, opportunities, and considerations for teaching and using MassMine in the classroom (Beveridge, Dobrin, Van Horn)
- MassMine for Children's Literature Data Research: Session tailored to support Children's Literature researchers and session will inform how to best support other specific research areas for how specific needs shape shared concerns (Project Team, DHLG)

April-May 2016: Promotion & Outreach for MassMine GUI Version

- Finalize documentation and materials (Project Team, DHLG)
- Press release for official MassMine 1.0; promoting MassMine at THATCamps in 2015 and other promotional activities (Project Team, DHLG)

Appendix: Selected References and Resources

MassMine and Related Resources

- MassMine Open Source Software Code on GitHub: https://github.com/n3mo/massmine
- Training Workshop Handout for Using MassMine on UF's Research Computing Cloud Server for upcoming training on Nov. 13, 2014, http://ufdc.ufl.edu/AA00025501/
- MassMine Project Assistant, Position Description: http://ufdc.ufl.edu/AA00022054/00011/pdf
- Case Studies of Using MassMine for Data Collection, Curation, & Analysis in Humanities Research, http://ufdc.ufl.edu/AA00025511/00001/pdf
- Beveridge, Aaron. Conference Presentation. "Writing Studies and Data Science in the 4th Paradigm." Indianapolis, IN: CCCC, March 19-22, 2014.
- Beveridge, Aaron. Presentation and Data Training. "Humanities Software Development: Data Mining and Writing Studies." Gainesville, FL: THATCamp-Gainesville, April 24, 2014: http://gainesville2014.thatcamp.org/2014/04/02/humanities-software-development-data-mining-and-writing-studies/
- UF Research Computing Resources: http://www.hpc.ufl.edu/resources/
    - To support scholars in fully testing MassMine within the full research process, Project Team members will utilize and integrate trainings on related needs (e.g., trainings on GIS, SPSS, data management, etc.); examples from UF: http://cms.uflib.ufl.edu/datamgmt/trainingongoing

Humanities Data Research

- Jockers, Matthew L. *Text Analyis with R for Students of Literature.* Springer, 2014; with textbook materials available for download from author website: http://www.matthewjockers.net/text-analysis-with-r-for-students-of-literature/
- Graham, Shawn, Ian Milligan, and Scott Weingart. *Big Digital History: Exploring Big Data through a Historian's Macroscope.* Imperial College Press, 2014 and online: http://www.themacroscope.org/
- *The Programming Historian,* resources on coding and development for data collection and processing, http://programminghistorian.org/
- Flanders, Julia and Trevor Muñoz. "An Introduction to Humanities Data Curation." *DH Curation Guide: a Community Resource Guide to Data Curation in the Digital Humanities,* http://guide.dhcuration.org/intro/
- Willford, Christa and Charles Henry. *One Culture: Computationally Intensive Research in the Humanities and Social Sciences, A Report on the Experiences of First Respondents to the Digging into Data Challenge,* Washington, DC: Council on Library and Information Resources, 2012: http://www.clir.org/pubs/reports/pub151/pub151.pdf

Appendix: Humanities Research Questions Needing MassMine's Data Collection, Curation, and Analysis Functionalities

Examples collected following MassMine's version 0.1.0 release in summer 2014.

Dislocating the Hip: Accounting for taste through spreadable media

Shannon Butts, English Department, University of Florida

Over the past few decades, a rhizomatic approach to scholarship has expanded the ways we map information, embracing multiplicity alongside an archeology of knowledge. Yet within contemporary media studies, terms like 'going viral' still invoke biological metaphors that muddle the power relations between producers and consumers. Who really crafts what we view as 'hip' and how can we attribute emerging trends? According to Henry Jenkins, the viral emphasis on replication and transmission fails to consider the networked reality of everyday communication – like a childhood game of *Telephone*, ideas change through sharing. Jenkins' work with 'spreadable media' engages the viral, but also acknowledges the participatory aspect of media circulation – a process that often transforms, repurposes, or distorts information as it passes through diverse communities. The initial spread of 'Fall's Hottest Fashions' might begin within the pages of Vogue, but 'go viral' as a street style bricolage posted to Pinterest or trending on Twitter. Building on Jenkins' concept of spreadable media, this project works to map the circulation of cool and the hype of hip within the world of popular fashion, dislocating traditional origin stories of producer and consumer. My research will use MassMine to examine print publications alongside user generated content to trace the evolving multiplicity of 'what's hot or not' in the Twittosphere. Tracking styles from initial appearance, through circulation, and variation, I aim to better understand how social media creates a platform for popularity and controls trends of information. Orange may be the new black, but what about next season?

Game Studies and Cultural Preservation: Mapping Archival Discourse

Kyle Bohunicky, English Department, University of Florida

In James Newman's *Best Before: Videogames, Supersession and Obsolescence*, Newman argues that digital games, as a medium, have proven remarkably durable, but despite cultural acceptance, these media demonstrate a troubling fragility. Questioning the putative "naturalness" of decay and obsolescence in the medium, Newman suggests that digital games are rapidly disappearing thanks to marketing, advertising, and journalistic discourse. The digital games industry, Newman suggests, has shown little interest in preservation and the project of game history and heritage. At a discursive level, a large part of this issue is that the digital games industry, developers, and players are entrenched in a language of computing that creates an illusory sense of archival activity. While battery saves, password systems, save states, memory cards, and save spots have helped players preserve their personal activities within games, the discourses that emerged from and around game memory are opaque and widely understood as consistent throughout the history of games. I suggest that closer attention to the discourses and conversations these storage technologies developed from and developed can give the medium a stronger understanding of its past and future. Thus, this project intends to nuance Newman's claim about illusory archival discourse by showing how different storage technologies affect the discourse about the heritage and future of both digital games and play. Such an investigation will need to span multiple communities including forums such as Neogaf, FAQ pages such as GameFAQs, and emulation sites such as Zophar's Domain and NESbox. Thus, Mass Mine's web mining feature that pulls data from specific URLs will be useful to show the relationship between digital storage and the medium's future/past that emerges in its discourse. Additionally, this work will enable me to detail the archival strategies that players themselves have developed to resist the industry's neglect of its history.

Teaching Data Research in the Humanities and Social Science Classroom

Nicholas M. Van Horn, Department of Psychology, The Ohio State University

A key component of training in the social sciences is the development of research skills, including an understanding of the relevant terminology, classification, methods, and trends in use by investigators.

Central to these concepts is the collection and analysis of theoretically-driven data. Students enrolled in a research methods course are taught to be informed consumers and producers of research. Commonly, to achieve this each student is mentored through the process of the scientific method by conducting a guided research project. Often, this involves identifying and taking a developed research plan from conception, to evaluation, to dissemination by means of a written paper and/or presentation. In the classroom setting, however, time and resources limit the scope of viable research hypotheses possible in such contexts. As a result, research questions are restricted to those answerable by simple, non-experimental approaches such as qualitative, survey, and correlational designs. Further, access is commonly limited with respect to populations and data sources of interest, and the introductory nature of the course virtually ensures that students lack the necessary skills to acquire and examine high-value data. These restrictions can weaken the intrinsic motivation and enthusiasm of students forced into compromised research projects due to the logistic constraints of the classroom. I plan to use MassMine as a pedagogical tool in the teaching of research methods in the social sciences in the fall of 2014 primarily as a means to overcome a subset of these problems. MassMine will serve two primary functions in my course. First, it will enable access to a new class of "existing data" research designs using non-trivial data sources (e.g., social networks) that are relevant and meaningful to my students. Second, data analysis and exposition are challenging skills to teach and to learn. Instructional examples are typically driven by inconsequential in-class surveys or existing "toy" data that tacitly separates the analysis of data from the research context that it rightfully belongs in. By contrast, MassMine will enable me to develop a quick research question live in class with student participation, collect data to test the class' hypothesis, and then immediately perform an analysis in real time. By situating the analysis within the context of development and acquisition, I believe students will connect with the quantitative and qualitative profile of the results in ways that theoretically-devoid examples do not.

Tracking Images Across Social Media

Laurie Gries, Department of English, University of Florida

Circulation Studies, the study of rhetoric and writing in motion (Gries 2013), is an emergent area of study within two disciplines: Communication and Rhetoric And Composition/Writing Studies. While much important work has already been done in this area in regard to textual circulation, when it comes to visual rhetoric, studies of circulation are limited by the lack of software to easily access, organize, and analyze social media data. As Shepard Fairey's now iconic Obama Hope image makes evident, thanks, in part to social media, images circulate, transform, and engage in divergent collective activities at viral rates. As described in *Iconographic Tracking: A Digital Research Method for Visual Rhetorics and Circulation Studies* (2013) as well as my forthcoming book *Still Life with Rhetoric: A New Materialist Approach for Visual Rhetorics* (2015), I have developed a digital research method called iconographic tracking to trace the circulation and transformation of viral images. Such method has proven effective in tracking the viral circulation of a single image. However, several problems currently exist that Mineware has potential to address.

First, iconographic tracking largely depends on manual research—a time intensive labor that is incapable of keeping up with a viral image in a digital age. In addition this method currently relies on multiple software programs to store, organize, analyze, and visualize data. In the past, I have relied on Zotero to store data, but due to glitches with this software, I have lost a significant amount of data. In addition, Zotero demands user input to capture website and image URLs—a time consuming process that software programs such as Mineware ought to be able to automate. Lastly, Zotero does not include analytical or visualization components; therefore, I have been forced to rely on manual coding methods and diverse visualization programs, each of which require their own training. For such time-related and labor reasons, iconographic tracking needs a reliable "one-stop" software program such as Mineware to support its research methods. I am currently working on writing a book-length rhetorical biography of the Obama Hope image—a research project that demands more research with iconographic tracking. I plan to use MassMine to support this research project. Specifically, I will rely on MassMine to track the

circulation and transformation of the Obama Hope image across social media; capture website and image URLs; and analyze and visualize research findings.

Transgender Representation in Social Media
RL Goldberg, English Department, University of Florida
Recent years have seen a dramatic increase in transgender representation in mainstream film and television, but this has not been mirrored in literary fiction, a surprising lacuna given the liberal range of play and potential that fiction presents. Instead of fiction, as Jay Prosser shows in *Second Skins: The Body Narratives of Transsexuality*, transgender representation in literature almost exclusively takes the form of memoir and autobiography. Certainly, there is no dearth of transgender autobiography, historical or contemporary. As Joanne Meyerowitz points out, transgender narratives proliferated from the early 1920s as doctors, predominantly in Germany, agreed to provide gender-affirming procedures to patients with crossgender identification. As North American media began publishing stories on sex change operations, readers with crossgender identification were able to not only imagine surgical intervention for themselves, but also found the language to express their desire for gender-affirming treatment. Yet even today, transgender writers predominantly write memoir rather than fiction. Though in recent years there has been relative interest in mainstream transgender literary fiction, examples are still limited. Instead, transgender authors seem to be producing fiction elsewhere—on their websites or blogs, self-publishing, or publishing with small presses. Increasingly, this is democratizing the publishing industry (for trans and cis writers alike) as writers who would not gain traction with mainstream publishing houses are finding niches and outlets for expression. For my project, I propose using MassMine to explore transgender fiction being produced on the margins, particularly, on social media. Through empirical analysis I will be able to understand current trends in transgender and queer fiction, and speculate on the future of transgender literature. Especially given how diffuse the transgender community is—international and diverse—such analytics are invaluable to understanding the production of fiction in the increasingly global context.

“no más combis”: tracing contemporary public transport reform in Lima, Peru
Jamie Lee Marks, Department of Anthropology, University of Florida
Over the past few decades, social scientists have increasingly discussed infrastructural systems, arguing that scholarly discussions be broadened to include the individuals that circulate through, imagine, and (re)constitute these systems. My dissertation research will provide an ethnographic portrait of large-scale transportation reform currently underway in Lima, Peru. Since 2010, the Metropolitan Lima Municipal Council (MML) has prioritized implementing a series of regulations and public works to create an integrated transport system. The next phase of the reform will restructure 50% of public transportation vehicles and routes in the city over the next few years, and will require the gradual removal of the majority of vehicles and routes that the most of Lima’s residents have been using since the 1990s. Despite the importance of existing vehicles, routes, and transit workers in the daily journeys of Lima residents until present, municipal transport reform campaigns explicitly reference existing buses and those who operate them as too chaotic, unsafe, unclean, and informal to be part of the city’s future. These campaigns urge competing narratives about Lima’s political, economic, and social history and future to the surface—charging conversations about mobility and transit and rendering social understandings of these phenomena available for critical analysis. Using a multi-sited ethnographic approach that combines digital and traditional participant observation, I will analyze how various social actors experience, remember, imagine, and narrate Lima’s public transit infrastructure. My fieldwork is based on (1) participant observation in the spaces of social and sensory encounter associated with emerging, existing, and disappearing transit infrastructures (2) semi-structured interviews with members of various publics involved in Lima’s mobile landscape—including transit workers, journalists, urban planners, NGO workers, and Lima residents of varied ages and socioeconomic classes and (3) discourse analysis of portrayals the transportation reform circulated in online news forums, as well as in Online Social Network (OSN) spaces such as Twitter and Facebook. These digital discursive spaces present a novel challenge to analysts interested in citizenship, how residents imagine infrastructural reform, and the relationships

between municipal campaigns and public understandings. MassMine will allow me to systematically organize posts and comments on Twitter and Facebook, rendering them manageable and intelligible for critical analysis. This is a critical aspect of my research project.

Mapping Premediation on Social Media
Jake Greene, English Department, University of Florida

The rise of networked media in the twenty first century inaugurated undeniable changes to how world events circulate within society, which, when viewed in light of the lingering cultural trauma of 9/11, illuminates the emergence of the medial phenomena that Richard Grusin refers to as "premediation." According to Grusin, premediation is the process through which a society anticipates the affective grounds of future trauma by mediating a variety of potential narrative paths. Although Grusin clearly states that premediation materializes in many cultural forms (film, literature, television, etc.), he is primarily interested in how the mainstream news media construct and disseminate narratives of premediation in response to environmental, economic, and political concerns.

For this research, I propose to combine Grusin's theoretical formulation of premediation with an empirical application of Mass Mine's data mining software in order to identify premediated news stories and trace their circulation on Facebook and Twitter. Specifically, I am interested in analyzing how users construct original text in their posts when linking to premediated news stories as a way of testing Grusin's claim that premediation functions as an "affective prophylactic" against future trauma.

From nightly reminders about the impending dangers of global climate change to recent speculations on the implications of conflict in the Middle East, the mainstream media is rife with premediated news. As a case study for tracking premediation on social media, I will follow the (re)circulation of news stories related to the recent increase in sinkholes in the state Florida. This story is useful in not only providing a limited scope through which to test this methodology, but it also connects to the more general category of ecological premediation.

Posthumanism
Melissa Bianchi, **English Department, University of Florida**

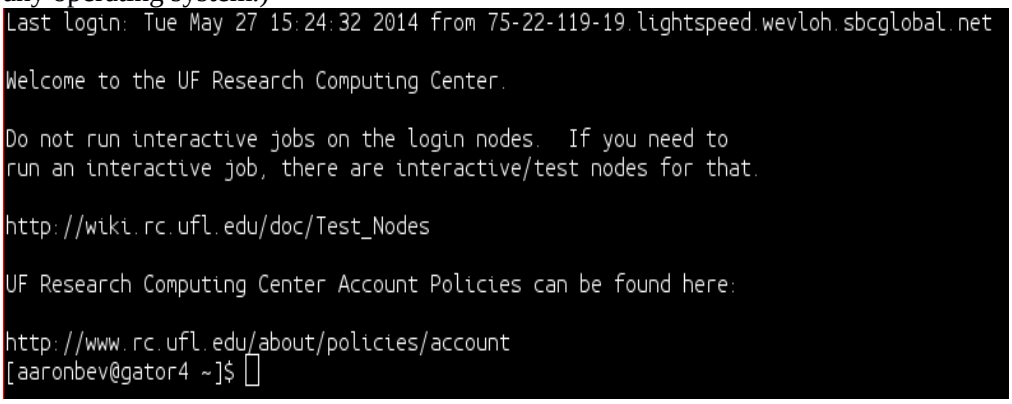
Contemporary scholars of human-animal studies have argued that ideologies perpetuating human social injustices are closely linked to those that justify the institutional exploitation of nonhuman animal species. For example, Cary Wolfe (*Animal Rites: American Culture, the Discourse of Species, and Posthumanist Theory*. Chicago: U of Chicago P, 2003) claims that: "as long as it is institutionally taken for granted that it is all right to systematically exploit and kill nonhuman animals simply because of their species, then the humanist discourse of species will always be available for use by some humans against other humans as well, to countenance violence against the social other of *whatever* species—or gender, or race, or class, or sexual difference" (8). Wolfe's argument forges a significant link between humans and other animals by suggesting that institutional exploitation and violence against both nonhuman species and certain human groups stems from a singular source: speciesism. He indicates that to combat the marginalization of any category of living beings requires that we attend to how our discourse reiterates presumptions of superiority over social others.

My research builds on Wolfe's work by tracing how a humanist discourse of species is employed and rallied against on one particular social media website: Twitter. Because Twitter is often used as a platform to raise awareness, advocate, and attack particular ideologies through hashtag movements, the website offers a means for organizing and tracking social movements that defend the rights of marginalized groups. This project will use Mass Mine to gather data from two recent and popular hashtag movements, #Blackfish and #Yesallwomen, that speak to cetacean and women's exploitation, respectively. I will compare the rhetoric and circulation of these hashtags to examine what similarities, differences, and links exist in their discourses. From this data, we may determine how discussions of cetacean exploitation on inform the ways we identify and define women as "social others" on Twitter, and suggest productive avenues for changing these discourses.

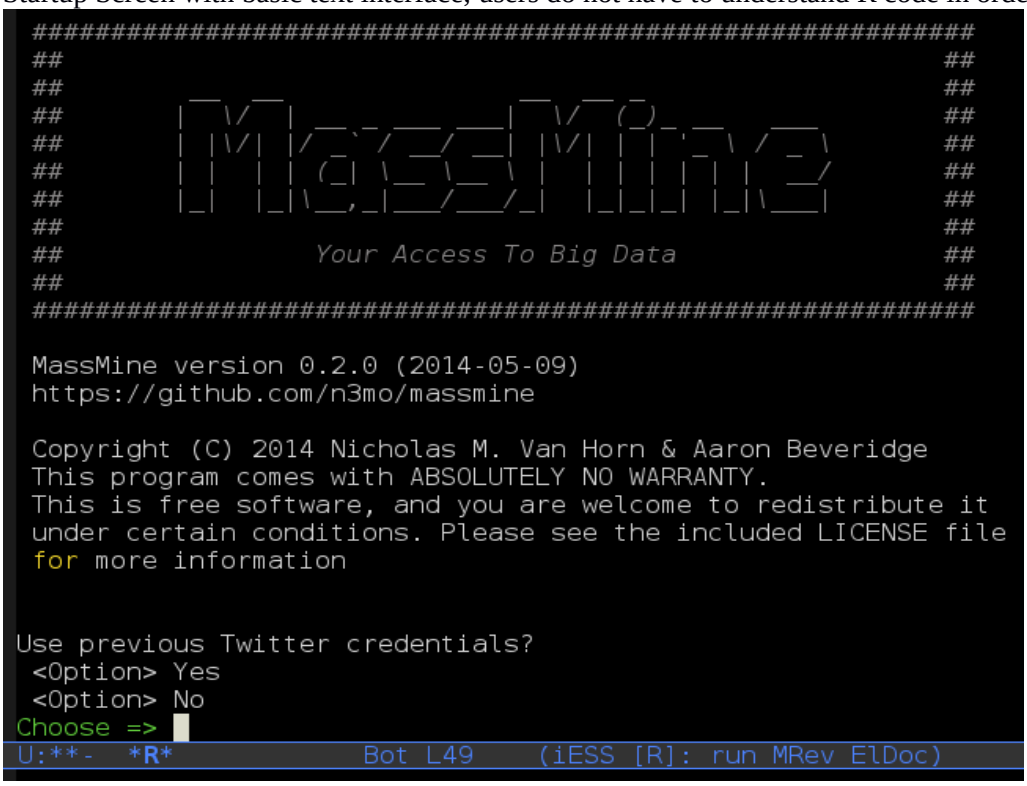
Appendix: Workshop Handout: Using MassMine on UF's Research Computing Cloud Server

Login and Startup

Remote login to the UF cloud server setup specifically for MassMine research, through a Linux console on a computer with the Ubuntu OS installed. (UF provides training on accessing cloud resources through any operating system.)

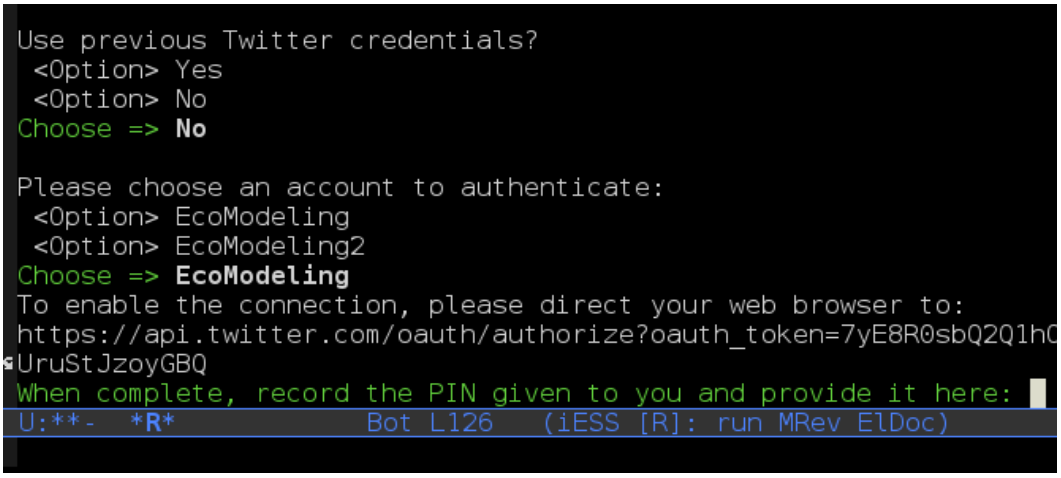


Startup Screen with basic text interface; users do not have to understand R code in order to collect data.

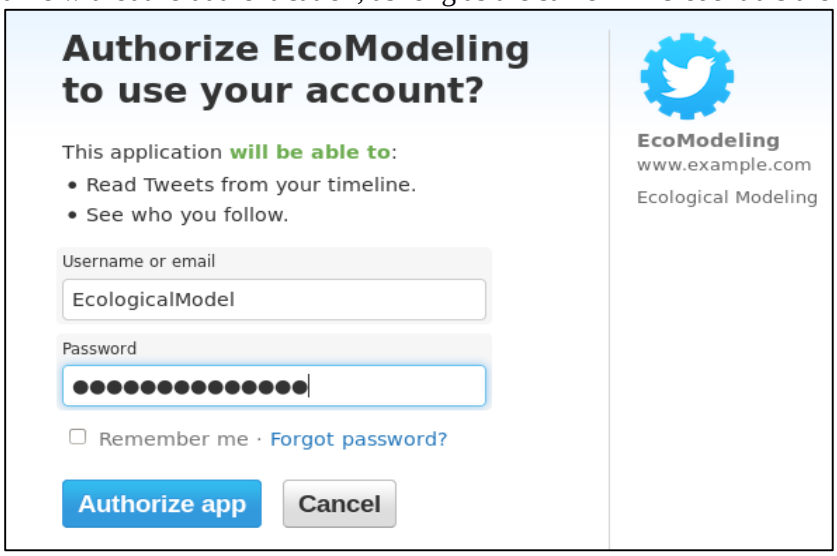


API Connection

MassMine checks the last configuration file and offers to re-authenticate API connection to restart a similar data collection.



Based on the configuration file information, MassMine automatically displays the API account for authentication. Authentication only needs to happen once; after that, users can run the software each new time without re-authentication, as long as the same API credentials are used.

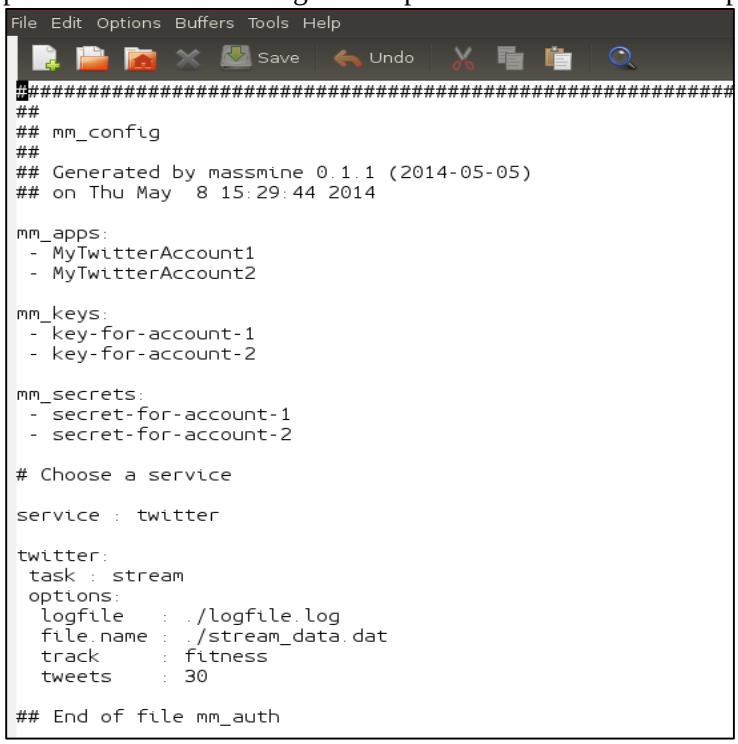


The API provides an authorization PIN, which syncs MassMine with the user's API access once the PIN is entered.

Configuration File

MassMine collects data based on the directions provided in the configuration file. The screenshot below shows the configuration file opened in a simple text editor. The configuration file is machine readable, and editable in any basic text editor on any operating system.

Users can save and re-process multiple configuration files to run different kinds of data collection activities the console application of MassMine. Templates of various basic configuration files are in process for use in trainings and experimentation for console application users.



Success Screen

MassMine responds to let the user know when a collection finished successfully without error, with all data automatically saved to the cloud space provided by UF's Research Computing.