

3. DESCRIPTION OF THE PROJECT AND ITS SIGNIFICANCE

The Processing Foundation is partnering with the UCLA Center for Digital Humanities (UCLA DH) and the Interface Design programme of the University of Applied Sciences Potsdam / Fachhochschule Potsdam (FHP), to create experimental visualizations of the data in news coverage from television and social media that take up principles of visual perspective as a way to facet, search, and analyze big data in ways that will extend current conventions in data visualization. The project will also create the custom programming tools needed to develop these visualizations. The project aims to explore, visualize, and make accessible various viewpoints in narrative and visual constructions of reality that different news sources create for their audiences. This will allow users of the visualizations to experience the significance of world events from multiple perspectives, literally and graphically, through creation of approaches that introduce graphic conventions from pictorial representation into data visualizations. The project seeks funding in the amount of \$350,000 over two years, to be divided as follows: forty percent for FHP to develop the visualizations, the interface design, and custom software tools; thirty percent for UCLA DH to develop and annotate the data, narrative elements, and characterizations; and thirty percent for the Processing Foundation to develop the core software utilized for the visualizations.

The project will utilize UCLA's NewsScape International Library of Television News, a large archive of broadcast television news from more than 40 channels worldwide, supplemented by a large corpus of social media data. Curated by the UCLA Library and directed by Professor Francis Steen of the Program for Digital Humanities and the Department of Communication Studies, NewsScape (tvnews.library.ucla.edu) began as an analog collection of television news during the Watergate hearings, and has since grown to a collection that contains more than 250,000 news shows, textually and visually indexed by around three billion timestamped words and several hundred million images. In 2010, the project was awarded a four-year grant in the NSF Cyber-enabled Discovery and Innovation Program, which led to the development and implementation of a computer vision and natural language processing search engine for the database. The proposed visualization platform will be open-sourced and publicly accessible online.

Processing, an open-source software and development environment, was first developed in 2001 to facilitate the convergence of visual thinking with computer programming. Operating within a visual context, Processing has been leveraged by a growing community of visual designers, scientists, researchers, and scholars, significantly advancing the understanding of large datasets through interactive information visualizations.

The Interface Design program and the Interaction Design Lab (IDL) of the Fachhochschule Potsdam (University of Applied Sciences) address aesthetic, social and cultural developments in the area of digital technologies. The Interaction Design Lab (IDL), founded in 2004, is the research lab of the Interface Design program. Over the last several years, the Interface Design programme of the Fachhochschule Potsdam, under the direction of Professor Boris Müller, has produced seven comprehensive data visualization projects of digital humanities research, six of which have used Processing, and all of which have been made publicly accessible online.

Our proposed partnership seeks to connect the UCLA NewsScape Archive with the innovative visualization expertise of FHP, by way of the Processing software, in order to produce visualizations of the enormous datasets contained within the archive. These visualizations will be made publicly available online, enabling the NewsScape database to be parsed, researched, and accessed in momentous new ways. Significantly, the software tools developed for the project will be enfolded into Processing's open-source libraries, facilitating future toolkits for digital humanities research.

4. NARRATIVE

Significance

EXPERIMENTAL VISUALIZATIONS—

NewsScape’s vast archive of data presents an opportunity to try several experimental approaches to visualization. Data visualizations are typically used as a basic method of analysis. In classic data visualization parlance, they allow patterns in data to be shown in a highly efficient and effective manner, compressing substantial amounts of quantitative information into a graphic display that aids analysis and comprehension: we can see patterns when metrics are expressed as graphics that are difficult, even impossible, to grasp in numerical or tabular form. In working with an archive the size of NewsScape, this conventional approach is useful, but not necessarily experimental. It would allow us to “see” patterns in the data, but not show us much about the assumptions on which visualizations work or to expand the language of visualization techniques. We would like to attempt both of these things in engaging with Processing.

EXAMINING THE ASSUMPTIONS IN VISUALIZATIONS—

Because all visualizations are metrics expressed as graphics, the analysis of the models on which parameterization, the initial assignment of value to data, is performed is crucial to exposing assumptions about their production. The relationship between data and display is not fixed, but mutable. A data set can be presented in any number of standard formats. But the nature of the data, the assumptions about the model, and the ways the production of knowledge work in data-display relations is another matter. How do we show the intellectual model on which the data were extracted from the phenomena? Can we use displays to do this? We think we can, and must, since the typical approach to information visualization is exemplified by emphasis on clarity and transparency, terms that suggest that the display is the data, or equivalent to it. The phrase, “form follows data,” which is the mantra of Edward Tufte, characterizes and drives this approach. But data, which are the result of parameterization use to extract quantitative information from phenomena (in our case, from the complex phenomena of news footage, with all of its many dimensions, nuances, intertextual and cultural play of references at the visual, verbal, and media levels), are already a construct. We suggest that creating visualizations of the model of data extraction, parameterization, and display protocols that refract rather than present data would be a proof of concept of the fundamental epistemological challenges of visualization. To do this requires making multiple, and iterative, lifecycles of the process of parameterization (data extraction), modelling of graphical conventions, and analysis of visualization/display, to show how these each expose different features, qualities, issues, assumptions in our understanding of the materials in the NewsScape archive. Why news? Because nothing in our culture is more complex in its curious production of a discourse of apparency and transparency (the sense that what is made is what is and that it is fully legible in the media) than the news. Our goal is to expose the intellectual models of newsmaking, through the process of the analysis of data display of this elaborate archive, and to create tools that allow users and viewers to see the constructed of news discourse through a set of critical, analytic lenses. What would these techniques look like?

EXPANDING THE LANGUAGE OF VISUALIZATION TECHNIQUES—

(a. Point of view in data)

Data visualizations use a graphic language of Cartesian coordinates, usually in planar view, sometimes in three-dimension formats, and sometimes with temporal dimensions (either

a slider or an animation that is time-based). What these visualizations leave out are the point of view systems that organize pictorial images through the use of perspective. Perspective is a system that locates the construction of an image in a point of view. Perspectival systems are not all the same—the system developed as European Renaissance perspective is located within a single viewers field of vision, while Byzantine perspective inverted this to locate all views within the gaze of an omniscient, all-seeing God. Cinema techniques and photographic approaches always have an implied “author” of any image, and theories of the gaze, of point of view as gendered and inscribed within conventions of cultures and their practices, are elaborately developed in film theory and media studies. These theoretical perspectives have not been applied to information visualizations, and the supposed neutrality and objectivity of information graphics has been reinforced as a result. We propose developing information visualizations that take the sentiment analysis and other rhetorical features of the NewsScape archive, and show how point of view is built into news production. We intend to do this in part through use of natural language processing tools that allow fine-grained analysis of the NewsScape transcription of the closed caption feed (for instance, analysis of auxillary verbs such as *should*, *could*, *would*, *can*, *will*, *might*) in their declarative and conditional forms). Creating visualizations that weight the sentiment differently across stations, media outlets, and news rhetoric would make a major contribution to the language of visualizations, allowing the “speaker” of a story to be identified as a structuring element of the news.

A news presentation is an act of storytelling: raw events on the ground must be turned into narratives that orient the reader to causes and consequences, meaning and significance, importance and urgency, action and observation. The media create their audiences, using a wide range of communicative techniques to convey causal reasoning and a moral perspective. We propose to develop tools for visualizing the divergent realities created by the media, so that we can begin to understand, in a visual, semantic, and accessible way, the following research questions:

- What does the world look like from the point of view of MSNBC, or the BBC, or FOX News?
- What is brought to the foreground and given salience?
- What is selectively backgrounded or passed over?
- Which causal networks are recruited to explain events, and how do these open up and close down opportunities for action and intervention?
- How are values and priorities conveyed and embedded into the structure of the narrative?

Rather than treating data simply as fact, we aspire to create visualizations that allow us to parse massive datasets into meaningful, approachable, and interactive analytic tools, allowing researchers to build subjective viewpoints and platforms, as well as large-scale, objective views of the whole. These tools will allow users to respond to specific research questions, like focusing on certain TV channels’ ideology and political profiles, or finding qualitative differences between corporate media and social media, to name but a few. By building visualizations of massive datasets, we can allow users to step into the shoes of a succession of audiences, adopting shifting perspectives and illuminating the broader process of meaning-making in the news. The impact on cultural fields is significant: narratives, hypotheses, and conclusions can be generated by being able to analytically parse the data, in ways that are specifically oriented toward interrogating and understanding the world-views that are constructed by the television news media.

(b. Comparative models of causal reasoning)

This second dimension of the visualizations will take the perspectival approach, but extend it into a comparative one. The specific topic to be used as a case study in the research is a proposition that news cycles follow a particular cycle of explanation, in which successive phases are characterized by different views of events and different language modes, is a research issue informing the NewsScape project. Creating a visualization that allows these marked distinctions to be seen and tracked would allow users to watch the unfolding of news analysis along particular rhetorical lines. Building on the point of view system described above, it would allow comparison of different voices and arguments within the news landscape.

(c. Symbolic vs. literal analysis)

Most data visualization takes literal information and quantifies it, but the analysis of language and image must take symbolic dimensions into account when we are dealing with the complexities of human culture and communication. We propose to create visualizations that are based on the analysis of the symbolic dimensions of news language, not only literal ones, and that push the domain of visualization beyond mere counting to speculative analysis. To do this, we will examine the ways in which news cycles and social media reports of unfolding events accrete extra value to certain terms or concepts, expand temporal moments, or transform a term or phrase into a “meme” in the natural language stream.

Each of these three approaches has been part of discussion among the participants at UCLA DH for the last two years, who have created a set of very preliminary sketches and analytic approaches for their realization. It would be very exciting to be able to take these theoretical ideas into an experimental phase in which we could create and test models of these experimental graphics. We believe these are of value to the NewsScape project, but also, that the findings would be of value to the field of data visualization and what we would call “interpretative graphics” as an extension of the field of information graphics.

Our analysis will make use of the latest techniques in computer vision and natural language processing, building on the work that has already been completed under a preceding grant from the National Science Foundation’s Cyberenabled Discovery and Innovation program. This will allow us to include metadata on facial expressions, trait judgments, clothing, gender, age, and ethnicity derived from the visual stream, and correlate them with conceptual frames, named entities, and syntactic structures derived from Stanford NLP tools, Pattern from CLiPS in Antwerp, and Semafor from the ARK team at CMU. The infrastructure of three billion words of annotated caption text, in timestamped registration with the sound and video, is already in place, allowing us to focus directly on innovations in perspectivized interactive multimodal visualizations.

DATA VISUALIZATION TOOLS FOR DIGITAL HUMANITIES—

The digitization of cultural artifacts has become a standard for museums, libraries, archives, and other cultural institutions. Manuscripts, drawings, paintings, prints, and sculptures, as well as more recent mediums like photography, film, video, and TV programmes, are becoming available through the web. The internet — as a medium itself — not only helps to archive these and make them available to the public, but it also creates new kinds of artifacts. Databases like Wikipedia and the Internet Movie Database (IMDb), and services like Twitter, Instagram, and Facebook, can be understood as new contemporary cultural artifacts, which require new consideration toward methodologies of archiving, preserving, interpreting, and analyzing.

Both the digital archives — that provide access to previous analog material — as well as

the new contemporary cultural artifacts, are enriched with layers of metadata. It is thus possible to access the artifacts themselves — but also show the relationship between them, creating a complex web of relationships and properties. The most powerful answer to creating an understanding of these complexities and to finding insights in large collections are explorable interfaces and data visualization.

Data visualization concerns the use of interactive computer graphics to get insight into large amounts of abstract data, such as multivariate, hierarchical, and network data. Conventional methods to depict data, such as bar-charts, pie charts, and line-charts, are useful to show aggregate information on a high level, but fail to depict large and complex datasets in detail.

Specific datasets need specific visualizations. In order to meet the research questions and the specific demands of the NewsScape Database, the partners will develop a range of individual visualizations and interactive interfaces that will allow the digital humanities to explore the data, generate new insights, and create new findings. Furthermore, the tools will allow the partners to make visualizations publicly available. The NewsScape archive can be freely explored by the public. Users can create and publish data views, enabling the public to participate in the debate.

PARTNERS—

The Processing Foundation is a leader in developing software and tools for non-technical audiences, the Interaction Design Lab at the University of Applied Sciences Potsdam is a leader in information visualization, and the UCLA Center for Digital Humanities, as one of the leading research centers in the field, represents the core interests of the humanities. Each partner will respond to the other, creating a vital and innovative interdisciplinary approach that addresses disciplinary questions and creates unique solutions. The collaboration between these three institutions is a distinctive opportunity to develop and champion innovation, new tools and methods, and create models for research projects within the digital humanities.

The collaboration between these three institutions is an ambitious and innovative approach toward digital humanities research, facilitating new pathways for the visualization of the large datasets produced by such research. The development of specific toolkits in parallel with the implementation of visualizations works to enhance our questions around how local, national, and international realities get constructed through news media by both creating visual representations of the data, and by establishing new technology that directly responds to the needs of the field. The ultimate goal of this unique partnership is to bring together three areas of knowledge in service of the digital humanities, and, specifically, we aim to produce visualizations of the NewsScape Archive that will set new standards within the fields of both information visualization and digital humanities research.

The partnership will demonstrate Processing's central role to the digital humanities, particularly its use in visualization. Processing is already a highly successful programming environment that allows non-programmers and programmers alike to create their own software-based visual projects. This project will extend these possibilities in the direction of data visualization techniques for the digital humanities, building specific software libraries that respond directly to the project and to the field at large. These specific tools will subsequently be generalized, and made available to the public as part of the Processing software.

Processing is a unique tool for research collaborations in the digital humanities, because it is designed as a first programming language, with a focus on teaching non-programmers the basic ideas and skills of programming. At the same time, it is a powerful programming language that scales to sophisticated professional use. Because of this dualism, it encourages collaboration

between non-programmers and more technical collaborators. Ideas and code can be worked on and moved around to different members of the team in a fluid manner. Because Processing is a programming language with a focus on visual production, it's also ideal for collaborations between visualization experts and humanities researchers. In contrast to d3, a popular language for creating visualizations on the web, Processing is more general and powerful for working with massive data sets and graphics. In contrast to the language R, Processing has less of a focus on statistics and more of a focus on graphics. Processing and R are compatible when used in series; R is used to prepare a data set and then Processing is used to visualize it.

History, scope and duration

The NewsScape Archive's functionality has been extensively expanded within the last four years, after being funded by an NSF Cyberenabled Discovery and Innovation grant. Going into the proposed project, the Archive has in place extensive textual and visual annotations. These include identification-level annotations such as named entity recognition and parts of speech from Natural Language Processing tools, and person, scene, and landmark detection from Computer Vision tools. Characterization-level annotations include sentiment scores and frame selection using automatic semantic role labeling (ASRL) with FrameNet, and social trait parameters (judgments of trustworthiness, competence, etc.) from Computer Vision. Automated story segmentation and multivariate topic clustering is also implemented for the whole NewsScape collection.

The first stage of our project will be Identification of Topics, done collaboratively between all three partners, where we identify parameters for proposed datasets to be visualized. One example topic, which we propose to begin with, is the news coverage of movements with a large popular participation, including: Tahrir Square, conflicts in Ukraine and Palestine, and the Green Revolution in Iran. Within these movements, we have outlined an initial set of dimensions through which to begin organizing the data: geographical, different kinds of media (democratic/social media vs. elite/corporate media), and the contrast between mass social media documentation (Twitter, Reddit, RSS feeds, etc.) and corporate media coverage on television. From these, we will leverage our textual and visual data mining parameters and begin to establish the various visualization criteria and plans. Within this stage, we will also define the necessary technical components of the environment, as well as the scope of the framework for technical platforms. This will require significant collaboration between the Processing Foundation and FHP, in order to begin mapping out the requirements to build the necessary technical tools. In tandem with this data structuring, we will create a set of sketches for the visualization of topics from different points of view within the media landscape. Taking the data for the topics, we will create a set of "weighted" metrics with which to test the construction of coordinates that include point of view. These will be paper and digital prototypes that are not connected to the data sets, but are constructed with the data sets in mind. Similar prototypes will be made for the comparative models and the feedback from social media into the NewsScape data. These will be storyboarded to show how the data visualizations can be built from the datasets.

The second stage is the Development of Custom Visualization Tools. Because the media data created for this collaboration through NewsScape is unique and can lead to new visualization techniques, new software tools will need to be envisioned and created as a precursor to building the visualizations. The Processing programming language is a general coding language with a focus on creating graphics. More specific tools are built from its core functionality. For

example, previous visualization libraries such as the Unfolding library for cartography (written at the FHP) are built with Processing. Processing is capable of creating all standard visualization techniques such as scatter plots and force-directed graphics, but it's not limited to these standard representations. For this project, we will consider and imagine new visualization techniques and will follow through with realizing tools to enable the most promising. This stage will require the collaborative efforts of the Processing Foundation, the FHP and input from UCLA DH. At this stage we expect to create some digital models of perspectively constructed data visualizations that can be experienced/redrawn from specific points of view. These will be created in Processing, but not linked directly to the data until the models of point of view, comparative modelling, and feedback loops have been developed and tested for legibility. One of the challenges to be addressed in this phase will be user testing to see how much instruction for use is required for these experimental conventions in display.

The third stage is the Development of Data Visualization Prototypes. This encompasses the design and implementation of prototypes for the NewsScape database, with a key deliverable. During the core development of these prototypes, we propose to have local work stations at UCLA and Potsdam, demonstrating the work in progress. At this stage, we will be developing general software libraries within Processing that respond directly to the needs of the custom datasets. The prototypes will connect front end display and back end data. The various paper and story-boarded test cases will be mocked up in simple formats, with room for detailed modifications (number of variables, filtered browsing of data and on-the-fly generation of results from faceted search) to be added once the visual models are working. Implementation of these more detailed versions will happen iteratively, with additional features of the dataset extracted and given specific graphical expression in the visualizations.

The fourth stage is Evaluation of Prototypes, based on criteria realized by all three partners and an advisory board. The advisory board brings together scholars affiliated with the Center for Digital Humanities at UCLA.

The fifth and final stage is Documentation and Dissemination, which includes the technical documentation of the framework; a publicly available online visualization showcase with a live connection to the NewsScape corpus; a visualization exhibit for public access at the Libraries at UCLA and and FHP; downloadable visualization tools for local deployment and customization to other datasets, along with the full code libraries, tools, and technical documentation; along with presentations at conferences and publication of the research results in scholarly journals.

In year one, the focus will be on the first three stages of identification and development, where all three partners will work in close dialogue together. Based on the custom datasets developed in stage one, the Processing Foundation and team at FHP will begin work toward building the necessary software tools. Paper prototypes and storyboarded visualizations will be made without connection to the data set, but made with specific data features and formats as part of the tests (for instance, first, second, or third person mode of address, or sentiment analysis, or topic repetition and lifecycle, etc.). Assignment of specific graphic variables to each element of the visualization, and tests of perspectival, comparative, and feedback models will be made. Year one culminates in the beginning of stage three: Development of Data Visualization Prototypes.

In year two, the prototype development will continue, accomplishing the workstation deliverables, made accessible in Potsdam and at UCLA. Evaluation will take place in concert with the partners and the advisory board. Linking the visualization prototypes to the data and creating a faceted search/query tool to link the visualization to specific features of the data on the fly will be crucial. Analysis of the results through publication and critical analysis will also be

part of the final phase of the project. Have we succeeded in providing insight in the NewsScape Archive, and in the process, have we created some effective, new, experimental visualization approaches?

Beyond the proposed two years of this grant period, the project will fold into its partners' existing projects. The visualizations will become a part of the public-facing components of the NewsScape Archive at the UCLA Library, and the core Processing language code will remain an integral part of the Processing code base.

Methodology and standards

The groups' collaborative workflow will be regulated by established methods and standards. The data to be visualized is exchanged in a standard json format, ensuring that the software developed will be readily deployable in other contexts. Coordination between the project participants will be achieved through teleconferences, collaborative co-presence tools, email exchanges, and mutual site visits. Bi-weekly team meetings using large-screen Polycom teleconference rooms at UCLA and FHP in Potsdam will support community building, creative idea generation, and prototype feedback.

In the first stage, the entire team will identify the potential in the existing NewsScape datasets and work creatively together to sketch ideas. The most promising ideas move on to the second stage, where they are functionally analyzed. This analysis informs the development of new code infrastructure to support the emerging visualization ideas, as well as new forms of corpus annotation. From here, a cycle of creating and evaluating prototypes will lead to the publishable results for the project, a combination of demonstration projects, conference presentations, and scholarly publications.

The code created during the process will be version controlled to store and archive of the entire endeavor. All partners will work with open file formats and open source code standards. Newly written code will be opened under appropriate licenses. Code will be developed on the website GitHub, where it will remain publicly accessible.

Intellectual Property Rights and copyright

The Processing software code and libraries written for the NewsScape visualizations will be open source under the GNU Public License (version 2) as is all of Processing software. The copyright for the Processing Software is owned by the Processing Foundation.

Environmental Scan

Our proposal follows a lineage of cutting-edge data visualization projects within the field of the digital humanities. An abridged list below:

History Flow (hint.fm/projects/historyflow)

A visualization project by Fernanda Viégas (PhD, MIT) and Martin Wattenberg (PhD, UC Berkeley), who lead Google's "Big Picture" visualization research group in Cambridge, Massachusetts. From the website: "In 2003, we decided to investigate the dynamics behind editing in Wikipedia. History flow is the method we invented to make sense of the data we collected." This project goes behind the text in Wikipedia into the social dynamics of

collaborative editing as revealed through the log data.

Pantheon: Mapping Historical Cultural Production (pantheon.media.mit.edu)

The Pantheon project, undertaken at MIT's Media Lab, is "an effort for the first time to generate a global comprehensive map of famous connections," with world maps, treemaps, matrices, and scatterplots. People can be sorted by profession or country of origin, and you can create lists of celebrities during specific time frames. It offers some remarkable insights into cultural values throughout history.

Selfiecity (selfiecity.net)

Directed by Dr. Lev Manovich, this project investigates the phenomenon of selfies in five international cities, presenting findings about the demographics of people taking selfies, their poses and expressions, through interactive media visualizations that allow users to navigate the whole set of 3200 photos.

Phototrails (phototrails.net)

From the website: "Phototrails is a research project that uses experimental media visualization techniques for exploring visual patterns, dynamics, and structures of planetary-scaled, user-generated shared photos. Using a sample of 2.3 million Instagram photos from 13 cities around the world, we show how temporal changes in number of shared photos, their locations, and visual characteristics can uncover social, cultural and political insights about people's activity around the world."

Specifically, the work undertaken by the Interface Design Lab at Fachhochschule Potsdam provides a foundation for demonstrating Processing's use as a key visualization software, and will guide us in our collaboration.

A standout example project in the area of visualizing digital humanities at the FH Potsdam is *MACE / mæve* (see Appendix B). The EU project *MACE* connects databases on architecture and provides a community of students, teachers, and researchers with interactive visualizations for the exploration of architectural projects. It integrates vast amounts of content from diverse repositories created in several large previous projects as well as from existing architectural design communities.

In 2008, *MACE* had the opportunity to develop an interactive exhibition for the 11th International Architecture Exhibition of the Venice Biennale. The resulting installation *mæve* (*MACE-Everyville*) by Boris Müller, Till Nagel, et al. provides visual and tangible access to the social and intellectual networks behind architectural projects.

mæve connects the entries of the Everyville student competition and puts them into the larger context of *MACE* content and metadata. By placing physical project cards on an interactive surface, users can explore the presented projects, embedded in an organic network of associated projects, people and media. The interaction with the cards is not limited to the one person, entire groups and teams can explore the content together.

Other projects undertaken at FHP include:

Unfolding, a PhD thesis project by Till Nagel. *Unfolding* is a library to create interactive maps and geovisualizations in Processing. The aim of the toolkit is to enable interaction designers, visualization experts, and others to easily create interactive geographic maps.

lit, a studio project by Jan-Erik Stange and Sebastian Meier. "lit" enables users to explore books by the locations contained in them on a multitouch table. It consists of a search interface that enables you to type in keywords like in a standard library catalogue interface. It provides the

user with the sequence of locations mentioned in the book. By tapping an icon on the textframe, the location points on the map are numbered according to their position in the book and connected with lines.

Culturegraphy, an MA thesis project by Kim Albrecht. Culturegraphy investigates cultural information exchange over time also known as ‘memes’. These networks can provide new insights into the rich interconnections of cultural development. (This project was built with D3.js.)

Splendor, by Matthias Löwe and Sebastian Schwinkendorf. Splendor is a visualization of gps tagged images and its relation to sights of any city in the world. The visualization creates a map of photographic paths in the urban landscape.

Sousveillance, a BA thesis project by Florian Schulz. Sousveillance explores how people react when their own photos are extracted from the internet and returned as prints to their neighborhood. The project looks for publicly available photos from Instagram that were obviously taken at the user’s home and brings them back to his or her environment.

Understanding Shakespeare: Towards a Visual Form for Dramatic Texts and Language, done as a BA thesis at FHP, by Stephan Thiel and available online at www.understanding-shakespeare.com. Built mainly with Processing, the goals of “Understanding Shakespeare” are to introduce a new form of reading drama, to help understand Shakespeare’s works in new and insightful ways, and to address our changed habits of consuming narrative works and knowledge through the capabilities of information visualization.

Work plan (see also Appendix A)

In year one, the focus will be on the first three stages of identification and development, where all three partners will work in close dialogue together. Based on the custom datasets developed in stage one, the Processing Foundation and team at FHP will begin work toward building the necessary software tools. Year one culminates in the beginning of stage three: Development of Data Visualization Prototypes.

In year two, the prototype development will continue, accomplishing the workstation deliverables, made accessible in Potsdam and at UCLA. Evaluation will take place in concert with the partners and the advisory board.

Nature of collaboration and staff

NAME	TITLE	INSTITUTION	TIME PERCENTAGE
Boris Müller	Principal Investigator	FHP	10%
Marian Dörk	Principal Investigator	FHP	5%
TBD	Researcher	FHP	hourly
TBD	Researcher	FHP	hourly
TBD	Student Assistant	FHP	hourly
Ben Fry	Lead software developer	PF	10%
Casey Reas	Lead software developer	PF	10%
Daniel Shiffman	Lead software developer	PF	10%
TBD	Senior software developer	PF	hourly
TBD	Software developer	PF	hourly
Francis Steen	Senior researcher	UCLA	10%
Johanna Drucker	Senior researcher	UCLA	5%

Wing Kai Chan	Senior software developer	UCLA	25%
TBD	Graduate Research Asst.	UCLA	50%
TBD	Program Analyst IV	UCLA	hourly

The partnership will be structured as follows: Professor Boris Müller of Fachhochschule Potsdam (FHP), as Co-Principal Investigator of the project, with Co-PI Marian Dörk, will lead of team of recent graduates in the development and implementation of the visualization prototypes. Professor Müller has over a decade of experience of leading research teams in the development of visualizations of large datasets.

This FHP team will interface with Processing Foundation software developers in the U.S. to coordinate specific functionalities to be incorporated into a Processing library that is oriented primarily toward the project, in response to technical parameters established by the partnership.

At UCLA, under the direction of Johanna Drucker and Francis Steen, researchers will work to extend the existing textual, visual, and aural annotations and create a suite of custom data sets based on the inquiries of the project’s research questions.

Final product and dissemination

The project’s proposed results, in keeping with the initial mission of both the NewsScape Archive and Processing Foundation to facilitate publicly accessible information, will be made available online at:

NewsScape—newsscape.library.ucla.edu
 Processing—processing.org
 FHP—interface.fh-potsdam.de

Specific visualization sets will be published under project domain names, as well as linked to the sites of the institutional partners. The technical products will be enfolded into a Processing library, with the relevant documentation published on Github.

The research outcomes from UCLA will be presented in scholarly publications and at international conferences where work at the intersection of communication studies, the humanities, discourse analysis, and data visualization are part of the ongoing professional conversations. This includes the annual conference on Digital Humanities, DataVis meetings in Los Angeles, workshops and institutes in the network of Digital Humanities activities, publication venues such as the *Journal of Visualization* (Springer), and online publication venues such as the *Digital Humanities Quarterly*. In addition, results will be disseminated by the Red Hen Lab, a distributed laboratory and consortium for research on multimodal communication, directed jointly by Francis Steen and Mark Turner.