

# **Epidemiology of Information Report to DiD Project Directors Meeting**

August 30, 2013

- **Overview of goals and objectives of DiD project**

This project develops data mining methods to understand the relationship between the spread of disease and the spread of information. The project uses a case study of reporting about the 1918 Spanish Flu epidemics in digitized American and Canadian newspapers. Data mining methods using historical newspapers can suggest new ways to interpret information flow and disease transmission in contemporary and future epidemics.

- **Challenges & lessons learned from international collaboration**

Our project has three components of international collaboration, all with Canadian partners: co-PI Gunther Eysenbach of the University of Toronto; advisory board members Ann Herring, Graeme Hirst, and Jian Pei; and the Peel's Prairie Provinces collection hosted by the University of Alberta.

- **Digital humanities, social sciences and computational based research methods in the context of big data projects**

Humanities and social science approaches include close reading of selected newspaper articles, historical assessments of source materials, and humanities interpretations of the spread of disease and the reporting provided by newspapers. Computational based research includes topic modeling, segmentation, tone classification, visualization, and reticulation analysis. This project has explored ways to connect these methods in ways that acknowledge yet also enhance the research questions and findings that are most relevant to scholars in each of these fields. In this way, the multi-disciplinary team contributes to better understanding of the common question of how information about disease can be analyzed.

- **Indicators of success**

At each stage, the project has been shaped by the distinct disciplinary approaches of the project directors, including computer, history, medical humanities, and information / library sciences. The design of research questions, the development of methods, the testing of results, and the

analysis of outcomes have emerged in a constant dialogue between and within these disciplinary perspectives. In addition, our methods and findings have been extensively reviewed at three points by the advisory board members, whose fields include medical anthropology, computational linguistics, computer science, and epidemiology. These reviews have directly shaped subsequent work on the project.

- **Measuring impact**

Our methods and findings will be presented at a research symposium on October 17, 2013 in the Washington DC region, with experts in data science, medical history, and information analytics. This symposium provides an opportunity to disseminate the results of the project to a broader audience, thus connecting big data analysis to digital humanities.

- **Knowledge dissemination mechanism and tools**

A research report based on the presentation to the Shared Horizons conference and aimed at a computer science / data analytics audience, was published in *Computer*, a journal of the IEEE. A research report intended for an audience of historians has been accepted for publication in *Perspectives on History*, published by the American Historical Association. Further publications based on the presentations to the symposium will be aimed at scholars and general audiences in medical rhetoric, epidemiology, and information analytics.

- **Importance of libraries, archives & data repositories**

All of our newspapers come from library collections, including the Library of Congress and the University of Alberta. In addition to the librarian who is a co-PI on this project, we have worked extensively with other librarians at Virginia Tech on data management and presentation.

- **Capacity building and training**

This project has depended extensively on the work of two Virginia Tech Graduate Research Assistants, Kathleen Kerr from the Writing and Rhetoric program in the Department of English and Samah Gad from the Discovery Analytics Center in the Department of Computer Science. Undergraduate students have participated in the Vaccination Research Group coordinated by Bernice Hausman and in a seminar on the Spanish Flu taught by Tom Ewing.