

---

## Smart Data Approaches to Exploring Independent Datasets across Disciplines, Media, and Perspectives for Research in the Humanities

**Marcia Lei Zeng**

mzeng@kent.edu

Kent State University, United States of America

**James Lee**

james.jlee@uc.edu

University of Cincinnati, United States of America

---

Research in the humanities has embraced the data-driven environment where advanced digital technologies have created the possibility of novel and hybrid methodologies. While the world is amazed by the many “V”s of Big Data (volume, velocity, variety, variability, veracity), the “V”alue of such data relies on the ability to achieve big insights from such data at any scale, great or small (Kobielus 2016). Smart Data can be considered trusted, contextual, relevant, cognitive, predictive, and consumable. In humanities research, Smart Data emphasizes the organizing and integrating processes from unstructured data to structured and semi-structured data, to make the big data smarter (Kobielus 2016, Schöch 2013).

In this panel, the interdisciplinary research teams from two universities will share their research findings and products as well as their experiences of using smart data approaches. As is the case with many other projects, the teams have to face the inherent challenges of converting heritage materials that were not machine-processable into digital datasets. Furthermore, the teams (again, like many others) have used common applications of available digital technologies and tools, such as GIS mapping, text encoding, fact mining, database construction, and visualization. In addition, the projects have employed sophisticated computer logics, the Linked Data models, network theory, and temporal-spatial data analytics, among others. However, the unique value of these projects lies in the exploration of independent datasets across disciplines, media, and perspectives.

Some of such datasets were built from unstructured data and media, while some other datasets have been existed isolated because the values are hidden in the silos. By drawing on many types of data simultaneously and interactively in an unprecedented manner, the research findings and established resources help reveal the unknown-unknowns and interpret significant values.

To begin the panel, Dr. Marcia Zeng (Professor of information science) and Dr. Hongshan Li (Professor of history) will report on the project “Digital Humanities Research with Smart Big Data — A Network Framework of Innovation History” using the case of the Liquid Crystal Institute (LCI) at Kent State University (KSU), the birthplace of liquid crystal displays. Through the work of its faculty and alumni, LCI has had a significant impact on the way the world sees things –on our smartphones, tablets, and computer screens (Bos, 2015). By nature, innovations and inventions demand collaboration across various kinds of networks. More and more collaborative efforts, instead of individual hero-inventors, resulted in the innovations in the last two centuries and, even more apparently, today. This project has focused on a community of scientists in one large institution instead of individual scientists. The project intended to use comprehensive data from cross disciplines and perspectives to discover meaningful patterns in the history of innovation. The presentation will discuss various research methodologies applied to different types of data used by a research team consisting of more than 10 faculty members and research assistants from the disciplines of information science, history, geography, physics, visual communication design, and mass communication. The presenters will share the integrated research findings regarding the sophisticated relationships and networks of contributed factors and impacts over LCI’s 50-year history that complement traditional study of the history of science and technology. The presentation also aims to share our lessons and roadmaps of taking smart data approaches with the intention of helping more researchers to overcome the challenges in researching the innovation history in the digital age.

The second presentation will be given by Dr. James Lee and Arlene Johnson, co-directors of the Digital Scholarship Center at the University of Cincinnati (UC), along with members of their team. They will describe the team’s research entitled “Linked Reading,” which uses sophisticated machine logics to allow researchers to directly query, analyze, and visualize or sonify data from multiple independent

datasets, including the University of Cincinnati's Elliston Poetry Archive. Since 2010, UC Libraries and the Department of English & Comparative Literature have collaborated on The Elliston Project, an audio archive of over 700 recordings of poetry or poetry-related content. The recordings span seven decades and include over 450 poets, including Wendell Berry, Robert Frost, Allen Ginsberg, Louise Glück and a host of others. Alan Liu, a pioneer of digital humanities in literary studies, considers the Elliston project to be a "world-class poetry audio archive," which has the potential to "alter the dominant understandings of a 'digital archive' developed for textual materials." Linked Reading allows one to examine scholarly questions on the question of poetry from multiple angles at once by pivoting laterally between multiple audio and text datasets. A unique opportunity for researchers and educators lies in constellations of micro and nano datasets that have been inadequately studied or even ignored. This approach, gathering smaller datasets of creative materials into a linked network, allows the team to leverage the local strengths in the humanities and creative arts (represented in stellar fashion by Elliston) to facilitate heretofore impossible research projects. Consider, for instance, poetic tone. It's a basic tenet of poetry instruction that the poem on the page is but a score to be performed; and yet, poetry scholarship is in virtually every instance a study of the printed poem. The project illustrates a potential of reshaping this well-studied topic by 'reading' the sonic features of poetic tone in massive numbers of poetry recordings across many linked smaller archives. As such, the aims of this research are: 1, to transform the techniques of linked data into an analytical and interpretive method, and 2, to adapt well-established machine learning techniques honed on text datasets for the analysis of large archives of born-audio creative works.

## Bibliography

**Bos, P.** (2015). Impact of our graduates on the industry. In: Morgan, S. et al. (eds.) 50 Years of Innovation, pp.34-35. Kent, OH.: Kent State University.

**Kobielus, J.** (2016). The Evolution of Big Data to Smart Data. Keynote at Smart Data Online 2016 July 13.

**Schöch, C.** (2013). Big? Smart? Clean? Messy? Data in the humanities. *Journal for Digital Humanities*. 2(3) <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>