Contribution ID: **55** Type: **Oral Presentation**

Piloting Interactive Data Science Learning Platforms through the Development of Cloud-based Computational Digital Notebooks

Tuesday, March 23, 2021 4:00 PM (20 minutes)

In a world presently mired in cascading lockdowns resulting from the spread of the COVID-19 pandemic, physical learning, communication, and collaboration have taken a colossal hit. Virtual access has become the need-of-the-hour, and the uses of cloud-based course content delivery, distance learning, and document collaboration are becoming increasingly ubiquitous. This paper introduces a novel method to allow students and faculty in the Humanities, Arts, and Social Sciences (HASS) to collaborate and interact through data analytic technologies using computational notebooks.

We demonstrate this approach using a digitally archived Legacy of Slavery (LoS) dataset from the Maryland State Archives (MSA) and illustrate the socio-technical challenges faced in establishing this learning environment. We provide a step-by-step process involved in accessing, developing, and integrating different infrastructure elements. The LoS in Maryland is a major initiative of the MSA. The program seeks to preserve and promote the vast universe of experiences that have shaped the lives of Maryland's African American population. Over the last 18 years, some 420,000 individuals have been identified, and data assembled into 16 major databases. These databases contain information unique to enslaved people's lives such as manumission records, certificates of freedom, census data, penitentiary records, etc.
One of this paper's primary objectives is to enable the digital representation of these culturally rich and sensitive collections ready to be analyzed and studied through contemporary scholars' lenses. This project aims to achieve this goal by making these databases available and accessible so that users can generate individual stories, glean insights, and possibly recover "erased" memories of the enslaved people. To achieve this goal, as a first step, individual dataset collections were prepared by downloading the databases and put through rigorous exploration, cleaning, and visualization process through coordination with interdisciplinary scholars composed of archivists, historians, computer scientists, and technology analysts. This project also illustrates the importance of a multidisciplinary approach to a unique set of digitized archival data with a specific focus on contextual aspects due to the data's historical value and sensitivity. The collaborative process used open-source and readily-accessible tools to create meaningful visualizations as an arrangement that flows together conducive for educators to teach. The visualizations use the spatial and temporal characteristics of the datasets to produce graphs and charts for a graphical view of the datasets. The visualizations constructed are responsive to present the data by instant connections to the datasets dynamically. Integration of these digital artifacts obtained from each dataset followed next on editors called "digital notebooks" that allow text and software code to co-exist and render in a single document coherently for instructors and students to follow the text with visual representations back-to-back. The "digital notebooks" are equipped with live examples of machine learning models and natural language processing on certain text-rich features of these dataset collections. The open-source nature of this project's setup and cloud-based distribution of these digital notebooks pave the way for students from underserved communities to take advantage of a unique way of learning and to perform hands-on work on marketable software tools, preparing them for a successful career.

The contributions of this paper to the field of digital humanities lie in the idea of providing an "always-on" cloud-based pedagogical environment for aspiring archivists and researchers worldwide to analyze, learn and unearth stories through data and fact-driven approach on the Legacy of Slavery in the State of Maryland.

Primary author: GNANASEKARAN, Rajesh Kumar (The University of Maryland)

Co-author: Prof. MARCIANO, Richard (The University of Maryland)

Presenter: GNANASEKARAN, Rajesh Kumar (The University of Maryland)

Session Classification: Humanities, Arts & Social Sciences Session

Track Classification: Humanities, Arts, and Social Sciences (HASS) Applications