Summary submitted with the GETAR proposal, 1619028 (and 1619371), that has been recommended for funding by **NSF**:

III: Small: Collaborative Research: Global Event and Trend Archive Research (GETAR)

Overview:
This project will devise interactive, integrated, digital library/archive systems coupled with linked and expert-curated webpage/tweet collections, covering key parts of the 1997-2020 timeframe, supporting research on urgent global challenge events and initiatives. It will allow diverse stakeholder communities to interactively collect, organize, browse, visualize, study, analyze, summarize, and explore content and sources related to biodiversity, climate change, crises, disasters, elections, energy policy, environmental policy/planning, geospatial information, green engineering, human rights, inequality, migrations, nuclear power, population growth, resiliency, shootings, sustainability, violence, etc. GETAR will leverage VT research on digital libraries, natural language processing, HCI, information retrieval, machine learning, discovery analytics, and Web archiving. In collaboration with the Internet Archive (IA), we will devise new methods for processing web archives, and comprehensive webpage collections for important End of Millennium (1997-2000) events. Beyond this period, when novel digital communication innovations were emerging as the main medium for the historical record, we will use collections of tweets and webpages we have, or will (collaboratively) construct, through 2020. Covering key parts of this timespan will enable comparisons, longitudinal studies of patterns and trends, and experiments with prediction. Coupled with event archives will be databases of sources, including individuals, organizations, and government agencies. We will advance the understanding of sampling methods, especially of sources, to control for bias; this is a necessity when events are discussed by large multilingual communities with amorphous structure. Domain experts will guide the construction and validation of interactive interfaces to help with event modeling--coupled with identification, development, curation, and utilization of collections. Building upon layers of interface components, services, and content management, tailored interfaces will be prepared and refined to address the needs in disciplines like Economics, History, Political Science, and

Sociology. There will be versions of Librarian Interfaces for collection development, curation, and reference. The latter will aid the development of a Public Interface. Our research into curation will guide the building of interfaces to be deployed in IA's K-12 Web Archiving Program, and for VT undergraduates. Research will extend work on modeling trends, events, and sources, to guide focused crawling, information extraction, tagging, and collaboration. Domain experts will leverage rich event models, exploiting the generality of the 5S framework (Societies, Scenarios, Spaces, Structure, Streams), extending from word, n-gram, topic, concept, and language models. One view of events will be in the context of movements, initiatives, programs, etc. We will exploit multiple levels of generality, e.g., 1) journalistic (who, what, when, where, why); 2) specific to the class of event (e.g., for an innovation, giving the inventor and global impact; for a climate disaster, tabulating the impact on life and infrastructure), and 3) event-specific. This will enable efficient assembly of knowledge bases, rapid prototyping of interfaces, gathering/curation of collections with high precision and recall, and flexible discovery in support of research and learning.

Keywords: digital libraries; Web archiving; NLP; global challenges; events/trends; HCI

Intellectual Merit:
This research will lead to innovations as we further integrate digital libraries and archives, along with collection development, curation, and utilization. Deliverables will include priceless collections, novel interfaces, and powerful tools and infrastructure. Interdisciplinary research advances will address: global challenges, digital humanities, digital libraries, web archiving, information retrieval, natural language processing, machine learning, and the construction of valuable interactive/collaborative interfaces.

Broader Impacts:
Studying and addressing important global issues, by scholars, the public, and K-12 students, will be enhanced through tailored interfaces coupled with important collections (e.g., related to climate, overpopulation, and/or globalization) that will be primary resources for understanding the modern world and its challenges, as well as initiatives, trends, and solutions.