

Lucy Nowell



National Science Foundation Office of Cyberinfrastructure, US

Inowell@nsf.gov

Now on assignment as a Program Director in the area of Data, Data Analysis and Visualization for the Office of Cyberinfrastructure at National Science Foundation (NSF), Lucy Nowell is a Chief Scientist from the Information Analytics group at Pacific Northwest National Laboratory (PNNL). She is also an alumna of the Virginia Tech Digital Libraries Research Laboratory, where she designed one of the first information visualization user interfaces for the Envision Project. Her research interests include long-term data preservation/archiving, user interaction with information in the context of massive data, usability engineering for information exploitation systems and digital electronic libraries, cognitive issues in user interface design, information visualization, intelligent user modeling and intelligent user interfaces, and information storage and retrieval.

As Program Director for Data, Data Analysis and Visualization in NSF's Office of Cyberinfrastructure, her program responsibilities include: [Sustainable Digital Data Preservation and Access Network Partners \(DataNet\)](#); [Community-based Data Interoperability Networks \(INTEROP\)](#); Software Development for Cyberinfrastructure (SDCI) and, [Strategic Technologies for Cyberinfrastructure \(STCI\)](#).

GRL2020 Position Paper

By the year 2020, the term “digital library” will be anachronistic, as all libraries will have digital collections. For research libraries, a major element of the digital collection will be the digital data that provides the foundation for science and engineering research and education.

For more than 2000 years, libraries have been trusted repositories for documents that have been the foundation of civilizations. Librarians have long been among the most literate members of the population, providing guidance about how to locate material of interest, as well as which documents could be deemed trustworthy and which should be challenged.

In the age of digital information, public libraries commonly provide Internet access to groups of people who would otherwise be excluded from participating in the global communication that the World Wide Web has enabled. Beyond providing access to systems, some libraries have become leaders in the movement towards digital libraries and schools of library science are increasingly schools of information or information science. That the documents of interest are digital in no way diminishes the need for trusted repositories for them nor the need for knowledgeable guides, as the “information superhighway” can take users through very bad neighborhoods.

A host of reports over the last decade make the case for creating long-lived repositories for digital data. The need is highlighted by the fact that last year for the first time we reached the point where annual production of digital documents surpasses the total global storage capacity. (<http://www.emc.com/collateral/analyst-reports/diverse-exploding-idc-exec-summary.pdf>) We can

no longer keep all that we create, and the task of deciding what to keep is not one to be undertaken lightly.

Finding reliable information in the face of such overabundance remains challenging. Search systems such as Google provide access to large volumes of information, commonly without consideration for the trustworthiness of data. Furthermore, the searches are biased by purchased priority listing and the preferences of other people, rather than objective assessment of content and value.

Where the documents sought are the digital data resulting from experiments or captured from sensor arrays, the sheer scale of a data set can boggle the mind, as we move beyond petascale to exascale data and computation. Data with geospatial and temporal attributes, such as that from seismic or environmental sensors, is irreplaceable. Data from clinical trials and other costly studies of human subjects is also well nigh impossible to replicate. But how many of these data sets can be preserved for decades? What must be done to maintain the usefulness of those that we do keep? Librarians and archivists have an important role to play in helping us know what to preserve and how to protect it while also providing open access.