

CIRSS

Center for Informatics Research
in Science and Scholarship

CIRSS conducts research
on information problems
that impact scientific and
scholarly inquiry.

cirss.lis.illinois.edu



GSLIS Education Programs

Preparing students for positions in data science, data analytics, data curation, data stewardship, data librarianship and related roles.

Specialization in Data Curation

http://www.lis.illinois.edu/academics/degrees/specializations/data_curation

Specialization in Socio-technical Data Analytics (SODA)

<http://www.lis.illinois.edu/academics/degrees/specializations/soda>

Research Events

e-Research Roundtable (ERRT)

Fostering research interactions in information problems in the conduct of research in the digital environment and development of research cyberinfrastructure.

CIRSS Seminar

Sharing our current research via research updates, guest lectures, and reruns of recent conference presentations.

Workshops

Data Curation and Access for the Digital Humanities

At the 2014 Digital Humanities at Oxford Summer School, Oxford, UK.

Data Curation Spring Institute

At the Hong Kong University of Science and Technology, Feb 2015.

Humanities Data: Curation, Analysis, Access, and Reuse

At the 2015 Digital Humanities at Oxford Summer School, Oxford, UK.

Site-Based Data Curation at Yellowstone National Park (SBDC)

Investigators: Carole Palmer (PI); Bruce Fouke, Geology (co-PI); Ann Rodman, Yellowstone (co-PI); Sayeed Choudhury, JHU (co-PI)
SBDC is developing a framework of policies and processes for the curation of "site-based" digital research data that responds to the needs of long-tail science researchers and site managers, and promotes coordination with libraries and data repositories.

Funded by: IMLS

Data Curation Education in Research Centers Program (DCERC)

Investigators: Carole Palmer (PI); Mary Marlino, NCAR (co-PI); Carol Tenopir, Tennessee (co-PI)
DCERC is a model for educating LIS master's and doctoral students in data curation through field experiences in research and data centers. DCERC brings students into the real world of scientific data curation, where they engage with current practices and challenges in data collection, delivery, management and preservation.

Funded by: IMLS

Workset Creation for Scholarly Analysis: Prototyping Project (WCSA)

Investigators: J. Stephen Downie (PI); Timothy Cole (co-PI); Beth Plale, Indiana (co-PI)

Working with the HathiTrust corpus of over 10 million volumes of text, the HTRC-based WCSA project aims to engage scholars in designing tools for exploration, location, and analytic grouping of materials so they can routinely conduct computational scholarship at scale, based on meaningful worksets.

Funded by: Mellon Foundation

Improving Information Retrieval by Consideration of Temporal Evidence in a Unified Model

Investigator: Miles Efron (PI)

Information retrieval (IR) systems are inherently temporal. Documents change, indexes acquire new documents, and systems answer or "field" queries differently over time. The vision of this project is to capitalize on this temporality to improve the models used for predicting document relevance.

Funded by: NSF

Kurator: A Provenance-enabled Workflow Platform and Toolkit to Curate Biodiversity Data

Investigators: Bertram Ludäscher (PI), James Hanken (PI, Harvard University); Timothy McPhillips (research developer)
Kurator is developing an extensible, open source toolkit that will enable collection managers and biodiversity researchers to compose their own workflows to automate data quality control and curation tasks. A graphical UI for building and running workflows will be developed for managers of natural science collections. The data curation workflow components that Kurator delivers will be usable both in scientific workflow systems and scripting languages such as Python.

Funded by: NSF-ABI

Collections and curation

e-Science

CIRSS

Digital humanities

Socio-technical data analytics

Selected Affiliated Projects

Predictive Modeling for Detection and Classification of Medical Entities and Facts

Investigator: Jana Diesner
Funded by: Intelligent Medical Objects

Collaborative Research: From Grant to Commercialization

Investigators: Vette Torvik (PI); Bruce Weinberg, Ohio State (partner PI)
Funded by: NSF, Mellon Foundation

Transforming Taxonomic Interfaces

Investigator: Michael Twidale (PI)
Funded by: NSF

Knowledge Sharing and Community Interaction on the VHA Data Portal

Investigators: Catherine Blake (PI), Maria Souden (PI, Veterans Affairs); Mike Twidale (co-PI)

The project aims to better understand essential socio-technical aspects of the Veterans Affairs (VA) Information Resource Center (VIReC) listserv and the information ecology in which it sits to maximize the findability and re-use of information generated by VA data users and to provide recommendations for improving its infrastructure.

Funded by: Veterans Affairs

Developing a Model for Socio-technical Data Analytics Education (SODA)

Investigator: Cathy Blake (PI)

<http://soda.lis.illinois.edu>

Combining theoretical foundations and hands-on experience, the SODA education model provides students with an understanding of the social and technical aspects of data analytics to prepare the next generation of leaders who can work with big data. Doctoral projects include automatically summarizing medical literature, and master's projects include assessing ozone impacts on asthma and identifying cultural clues from news stories.

Funded by: IMLS

