
Xuemao Wang
Vice Provost for Digital Scholarship
Dean and University Librarian

James Lee
Associate Vice Provost for Digital Scholarship
Associate Dean of Libraries
Director, Digital Scholarship Center

Kristen Burgess
Operational Manager
Digital Scholarship Center
University of Cincinnati

University of Cincinnati Quick Facts

Location: Cincinnati, Ohio
Number of Buildings: 118 facilities on 476 acres
Majors & Programs: 414 degree programs, 262 minors and certificates
Athletics: NCAA Division I; American Athletic Conference
Mascot: Bearcat
Colors: Red and Black
Famous Alumni & Faculty: Astronaut Neil Armstrong; President and later Chief Justice William Howard Taft; Eula Bingham, environmental scientist and one-time head of OSHA; Albert Sabin, developer of the oral polio vaccine; and prima ballerina Suzanne Farrell
Students from: 50 states and 114 countries
Living Alumni: over 300,000 with approximately half (more than 140,000 residing in the greater Cincinnati region).

Additional information available on the UC Fact Sheet.
University of Cincinnati Libraries' Strategic Direction
Journey to Transformational Change

The University of Cincinnati Libraries will become the **globally engaged, intellectual commons of the university**—positioning ourselves as the hub of collaboration, digital innovation and scholarly endeavor on campus.

www.libraries.uc.edu
Our Journey Towards a Great Public Research University – Boldly *Bearcat*

UNLEASHING OUR VISION

LEADING URBAN PUBLIC UNIVERSITIES INTO A NEW ERA OF INNOVATION AND IMPACT
THEMATIC AREAS
FOR STRATEGIC INVESTMENT

DATA SCIENCES

HEALTH & WELLNESS

CREATIVITY, EXPLORATION & JUSTICE IN A DIGITAL WORLD

HUMAN-ENVIRONMENT INTERACTIONS
UC Research 2030 Plan

**Objective:**

**National Prominence**

**Goal:**
TOP 25 PUBLIC RESEARCH UNIVERSITY

**Guiding Principle:**
Galvanizing our mission to serve the public good

**Invest in Success**

- Recruit and retain top talent
- Innovate the research infrastructure
- Cultivate & grow programs of excellence

**Invest to Advance**

- Urban futures pathway
- Rethinking the where
- Coalition for change

**Objective:**

**Impactful Research**

**Goal:**
Improving people's lives

**Guiding Principle:**
Foundational partnerships to solve real-world problems
DSC: Mission

- Core mission - To break silos and cross wires across the university. We work at the intersection of data science, the arts and humanities, and the libraries.
- Academic Center (Libraries + Arts & Sciences)
- Mellon Digital Integrator (One of six, projects with eight colleges).
- UC Digital Futures Anchor Team
• We are a technical catalyst: technology to activate new research

• Machine Learning and Human-Interpretable Data Visualization on Large Unstructured Datasets (Text, Image, Sound, Video)

• We translate between disciplines that rarely interact in order to connect content experts with technical experts.

• We provide resources and infrastructure to nurture research questions and collaborations that slip between the cracks of colleges and funding agencies.
UC's Digital Catalyst

**TEACHING**
- Intro coursework & workshops
- Student research in faculty team labs
- Job placement & digital training

**CCM**

**CCHMC**

**CECH**

**DAAP**

**CEAS**

**Allied Health Sciences**

**Social Science**

**DAAP**

**Architecture**

**Social Science**

**Planning**

**Languages**

**Digital Futures**

**Natural Science**

**Social Justice**

**A&S**

**Business**

**Digital Futures**

**Office of Research**

**Technology**

**Education**

**1819 Innovation Hub**

**IT@UC**

**Libraries**

**Computers**

**Medicine**

**Engineering**

**Nursing**

**Pharmacy**

**Design**

**Politics**

**CORE SERVICES**
- Digital tools development & training
- Faculty development
- Digital skills for analog expertise & archives

**RESEARCH**
Networked Research: Leverage different funding models and different research outputs for common team goals.

**Mission:** To create new areas of research and discovery across the university by applying AI + data visualization to human problems.

**Vision:** The DSC will be a distinctive global leader in transdisciplinary digital research.
UC Libraries provides access to a wide range of Research Data and GIS services and resources for the campus community. Informationists and librarians are available to assist researchers in managing and preserving research data, finding and acquiring external data, and in utilizing GIS techniques and software. The library also provides a variety of computing and collaboration spaces to support researchers.
DSC: Who

- **DSC**
  - James Lee, Asst Vice Provost for Digital Scholarship and Asst Dean of Libraries, Director of DSC
  - Kristen Burgess, Operational Manager
  - Lindsay Nickels, Program Coordinator
  - Ezra Edgerton, Data Visualization Developer
  - Erin McCabe, Digital Scholarship Library Fellow

- **RDS**
  - Amy Koshoffer, Asst Director of Research & Data Services
  - Tiffany Grant, Asst Director of Research & Data Services
  - Rebecca Olson, Business and Social Science Informationist
  - Don Jason, Health Informationist
  - Ted Baldwin, Director, Science & Engineering Libraries
  - Dorcas Washington, Data Analyst Specialist

- **Graduate Students (English, Computer Science, Business Analytics)**

www.libraries.uc.edu
The DSC has assembled research groups that genuinely span multiple disciplines, with people trained to think very differently about every step in the research process.

Teams are composed of true partners across entire research lifecycle:
- Formulation of research questions
- Pitching grant proposals
- Dataset cleanup and manipulation
- Data analysis and visualization
- Argument formation
- Publication of findings

In 2017, the DSC received $900,000 from the Andrew W. Mellon Foundation to expand this mission.
UC awarded a $700K grant from The Andrew W. Mellon Foundation

The renewal grant will advance and expand the Digital Scholarship Center’s "catalyst" model
How?

1. **Technical**: Our platform adapts machine learning approaches to any text and image dataset for research projects. We apply these methods in a discipline-specific way.

2. **Human**: We assemble teams to nurture these unconventional transdisciplinary research questions and partnerships – every collaborator has different goals and culture.
How?

1. **Tools**: Our technical platform applies deep learning on any text and image dataset for our partners. We open up these methods to collaborators (e.g. medicine, law, journalism, design).

2. **We connect**: faculty research questions + machine learning and data visualization. We assemble teams to nurture these unconventional transdisciplinary research questions and partnerships.

3. **Vision**: From Catalyst to Digital Integrator, a full stack infrastructure to support grand challenge projects by assembling transdisciplinary teams that apply our techniques to their field.

4. **Create new opportunities**: $3.1 Million in grants, 3 Books + 10 Articles, commercialization (CincyTech, P&G, 84.51°), community advocacy over past 3 years.

---

**Digital Integration: Technical Mission**

**Data Management**
- Open Science Framework
- Open Datasets
- Machine Actionable Data Management
- “Gentle Introduction to Data”: Human Welcome

**Data Analysis: Digital Humanities to Bioinformatics**
- Multimodal Deep Learning / Machine Learning
- Data Visualization Interfaces
- Mixed Methods: EDA + CDA
- Quantitative and Qualitative Data

**Digital Outcomes and Products**
- Publications
- Conference Presentations
- Reference Datasets
- Project Websites / Apps
- Grants
- Machine Actionable Data Management

---

- High Performance Computing
- Storage
- Data Structures:
  - Format,
  - Fields,
  - Metadata,
  - Machine Readable

- Team Science Culture
- Writing + Environmental Scans
- Method Translation
- Student Training and NextGen MA / PhD
Human Centered ML Needs Data Visualization
Machine Learning Platform: Model of Models (MoM)
Machine Learning Platform: Model of Models (MoM)

- Two machine learning strategies used to observe the latent patterns in large corpora.
  - Topic modeling (Latent Dirichlet Allocation – LDA)
  - Word embeddings (word2vec, BERT)
- Aggregates multiple models in parallel to compare word usage across the parallel models.
- Clusters integrate topics from the 6 models into an aggregated “model of models”:
  - Confirm consistent topics across all models
  - Reveal underrepresented topics that may not have appeared in a single model representation.
- Distributed parallel approach increases user confidence and interpretability of our models by bringing the most stable topics to the top tier of the model results.
Machine Learning Platform: Trust the Human for Validation

- Internal validation: Topic Coherence

- External validation:
  - Subject matter expert tagging of randomized 20% (N>1000) corpus
  - Blind human coder panel – Percentage agreement

- Replicability:
  - Pattern recognition capabilities of NLP methods as an information retrieval – and not a black-box classification – approach.
  - Provide models capable of evaluation by our panel of multiple independent coders.
  - Parallel replicates in each model (6–20 runs)

- Hybrid ML approach:
  - Human judgment of subject matter experts to verify and tag the model result – outperforms a purely machine-based analysis.
  - Semi-supervised learning for classification.
Interoperability with Datasets, Medium and Large

- HTRC Extracted Features, JSTOR Data for Research, Chronicling America, Text Creation Partnership, Harvard Case Law.
- Social Media (Twitter, Instagram, Reddit).
- Small corpora: we'll help you read them.
The Impact of Respiratory Viral Infection on Wheezing Illnesses and Asthma Exacerbations

Overview: Respiratory viral-induced wheezing illnesses in young children, viral bronchiolitis, is a LRTI typically associated with cough, tachypnea, retractions, and diffuse wheezing and rales [8, 9]. Bronchiolitis is a leading cause of hospitalizations in the first year of life, accounting for an estimated 120,000 infant hospitalizations annually [10]. In infants, the etiologic agents of bronchiolitis and other viral respiratory infections associated with wheezing include respiratory syncytial virus (RSV), human metapneumovirus (hMPV), human bocavirus (hBoV) [11, 12, 14, 15]. RSV causes epidemics of bronchiolitis and typically circulates in temperate climates during November to April with peaks in the winter months [11, 14, 16]. In tropical climates, peaks are related to temperature and level of rainfall [17]. RSV infects the majority of children during their first year of life and essentially all children show evidence of RSV infection by age 3 years [18]. The initial RSV infection is typically the most severe, causing lower respiratory tract disease, such as bronchiolitis, in 20% to 50% of infants [11, 18, 19]. Other viruses such as human parainfluenza virus (hPIV), adenovirus, rhinovirus, and respiratory syncytial virus (RSV) are found with reported rates of HMPV [15, 16, 17].
Model of Models: User Interface

Adaptable visualization outputs based on a single underlying model.
Model of Models: User Interface

Adaptable visualization outputs based on a single underlying model.
MoM: Digital Humanities
MoM: Social Sciences
Digital Scholarship and Academic Health

- Partnership with College of Medicine, CCHMC Biomedical Informatics.
- Text Mining Electronic Health Records, Scientific Literature, Grant Databases, Social Media, Imaging.
- A Two-Way Street: Teaching STEM about qualitative data.
Uncertain Diagnosis Project
CCHMC Hospital Medicine Division

Uncertain Case

**Unclear** diagnosis at this time, but **differential** would include **post-viral gastroparesis/ileus**, **although** severe intermittent **abdominal pain** would not be consistent with that diagnosis. **Could** have intermittent **intussusception** or **volvulus** with a lead point of an enlarged lymph node in the setting of recent **viral gastroenteritis**. **Renal colic** is a possibility with the description of "writhing" in pain, but pain is not localized to the back or flanks and no blood of other abnormality seen on UA. **Biliary colic** **could be considered**, **although** would be unusual in her age range and without associated with food. **Appendicitis** remains on the **differential** as was not visualized on ultrasound, but exam findings not consistent with the diagnosis.
Using Linguistic Trends to Improve Patient Outcomes
Relative impact of adverse events and screened symptoms of posttraumatic stress disorder and depression among active duty soldiers seeking mental health care.

Symptoms of depression and posttraumatic stress are among the most studied psychiatric difficulties among soldiers. Such symptoms have been linked to a history of adverse events among both civilians and combat veterans. There is a scarcity of research on this topic that can be applied to an active duty clinical population. Intake screening data were reviewed for 1,926 soldiers presenting to an outpatient mental health clinic to identify variables, including history of potentially traumatic experiences, associated with screened symptoms of posttraumatic stress disorder (PTSD) and depression. Demographics such as age, gender, and military rank, as well as number of adverse childhood experiences were significant predictors of screened PTSD and depression. A history of deployment to a combat zone predicted screened PTSD, but not depression. The role of childhood abuse as a risk factor is explored in more detail in the next section.
Second Phase: Multimodal ML
Second Phase: Jupyter Notebooks Pipeline
Let's Work Together

- Model of models platform: https://modelofmodels.io
- DSC website: http://dsc.uc.edu

Second Grant Objectives:
- 15 Subgrants through Mellon Foundation Award using our technologies.
- Expand use cases and projects deploying MoM and Jupyter library.
- Experimental use of MoM for data services.
- Partnerships with external collaborators engaged in digital scholarship.

Acknowledgements:
We are grateful to the Andrew W. Mellon Foundation, Public Knowledge Program for their support.