

CAS

Computational Archival Science

The DCIC is pioneering advances in *computational treatments of archival and cultural content*.

See our CAS portal for the latest developments:

<http://dcicblog.umd.edu/cas/>

What is CAS?

An interdisciplinary field concerned with the application of computational methods and resources to large-scale records/archives processing, analysis, storage, long-term preservation, and access, with the aim of improving efficiency, productivity and precision in support of appraisal, arrangement and description, preservation and access decisions, and engaging and undertaking research with archival materials.

CAS Founding Partners:

Richard Marciano, U. Maryland

Mark Hedges, King's College London (UK)

Vicki Lemieux, U. British Columbia (Canada)

Maria Esteva, Texas Advanced Computing Center (TACC)

Michael Kurtz, U. Maryland

Bill Underwood, U. Maryland

Greg Jansen, U. Maryland

Mark Conrad, National Archives and Records Administration (NARA)

curatelab

Hornbake South 4110

Digital lab for group learning, collaborative design, and hands-on digital curation project development (23 seats, 3 interactive screens, 12 workstations with 12TB of storage).



digitizationlab

Hornbake South 4110D

Document scanning, image manipulation, and archival ingestion facility for group projects.



serverfarm

Atlantic Building

On-campus virtual machine farm for research data processing, storage, and hosting (15TB storage, 2 Dell servers, VMWare-powered).



cloudlab

Amazon Cloud

Dashboard-enabled virtual computing lab in the cloud for creating Windows/Ubuntu instances using Amazon Web Services (AWS).



dataCave

UMD Cyberinfrastructure Center at the Rivertech Bldg

DRAS-TIC

Digital Repository At Scale That Invites Computation (To Improve Collections): a petascale archival storage and preservation repository (based on the DRAS-TIC open-source software [NoSQL Cassandra database] and computational infrastructure (4 Dell nodes).



dcic

digital curation innovation center

<http://dcic.umd.edu>

Mission:

Be a leader in the digital curation research and educational fields, and foster interdisciplinary collaborations using Big Records and Archival Analytics with public / industry / government partnerships.

Goals:

Sponsor interdisciplinary projects that explore the integration of archival research data, user-contributed data, and technology to generate new forms of analysis and historical research engagement, particularly in the arenas of social justice, human rights, and cultural heritage.

Motto:

"Integrating Education and Research"



COLLEGE OF
INFORMATION
STUDIES

Digital Curation Training

Digital Curation Fellowships & Internships

The iSchool has several Fellowship opportunities for students in digital curation and archives. These include a collaboration with the **National Agricultural Library (NAL)**; extensive project work with the **National Park Service (NPS)**; and a scholarship established in honor of **Bruce Ambacher**, retired senior archivist and iSchool faculty member. Also, student internship opportunities through the DCIC's Educational Partnership with public and private sector organizations are available.

Digital Curation Certificate Program (DCIP)

The DCIP is a three course, fully online program designed for working professionals who need training in next generation cloud computing technologies, tools, resources, and best practices to help them evaluate, select, and implement digital curation solutions.

Interdisciplinary Hands-on Digital Curation

In the **curate** labs have the opportunity to develop team-based hands-on extracurricular digital curation projects. The goal is to develop new digital skills, conduct interdisciplinary research, and explore professional development opportunities at the intersection of archives, big data, and analytics through the projects we offer.

Funding

"Brown Dog"

A CIC Big-10 \$10.5M NSF/DIBBs-funded collaboration with U. of Illinois NCSA Supercomputing Center and industry partners (NetApp and Archive Analytics Solutions). This project aims to help accelerate the development of digital curation processes and services and create a data observatory to provide access to Big Records training sets and teach students practical digital curation skills.

"Curate Cloud"

A \$300K IMLS-funded project that helped develop and launch online professional education for digital curation professionals, and is also developing as an open-source research and educational platform, the VCL (Virtual Computing Lab), to remove barriers to access for curation tools and resources.

"DRAS-TIC Fedora"

A \$240K IMLS-funded project that will research, develop, and test software architectures to improve the performance and scalability of the Fedora Repository. This project will apply the new Fedora 5 application programming interface (API) to the repository software stack called DRAS-TIC. We will work with Fedora community partners that will help us develop use cases and performance expectations.

People

Research & Teaching Staff:

Richard Marciano	<i>Director & SALT Lab Director</i>
Ken Heger	<i>Research Associate & Digitization Lab Director</i>
Katrina Fenlon	<i>Research Associate</i>
Ricardo Punzalan	<i>Research Associate & ARC Lab Director</i>
Adam Kriesberg	<i>Research Associate & Lecturer</i>
Noah Dibert	<i>Assistant Director</i>
Greg Jansen	<i>Research Software Architect</i>
Mary Kendig	<i>Research Associate & Instructor</i>
William Underwood	<i>Affiliate Professor</i>
Victoria Lemieux	<i>Affiliate Professor</i>
Maria Esteva	<i>Affiliate Professor</i>
Michael Kurtz	<i>Director (Emeritus)</i>



Research Affiliates:

U. Maryland

Tammy **Clegg**, Jesse **Johnston**,
Katy **Lawley**, Trevor **Owens**, Jenny **Preece**, & Katie **Shilton**

External

Bruce **Ambacher**, Natalie **Baur**, John **Burns**, Andrew **Lau**,
& Scott **Madry**

Doctoral Students:

Andrew Casertano & Will Thomas	<i>(Advisor: Marciano)</i>
Caitlin Christian-Lamb & Edward Summers	<i>(Advisor: Punzalan)</i>

Contact:

Contact us at dcic.ischool@gmail.com



Projects

Justice, Human Rights, & Cultural Heritage:

1. Global Journeys, Local Communities

A student- and professional society-driven project to collect information documenting payment of pensions to American veterans living overseas. The project documents migration patterns, the flow of money, health conditions, and family connections prior to World War I.

2. Managing White House Correspondents Association (WHCA) Pool Reports

This project is a partnership between the DCIC, UMD's School of Journalism and UMD Libraries. Students will collaborate in creating a digitized Presidential Press Pool Archives.

3. Legacy of Slavery

This project is a partnership with the Maryland State Archives, where students will transcribe digitized records involving Census Records and Certificates of Freedom, analyze the data and visualize it to enhance access for users.

4. Japanese-American WWII Camps

In partnership with the US National Archives, the project explores the auto-redaction of WWII Japanese-American Camp data. NLP and NER text processing techniques are explored through the development of extraction workflows.

5. Mapping Inequality

In partnership with Johns Hopkins, Virginia Tech, and U. of Richmond, a national collection of New Deal *redlining* records is being crowdsourced (these unique records capture racial, ethnic, and economic conditions).

6. The Human Face of Big Data

A student-led project that will create access and collaborative opportunities around historically and socially-significant heterogeneous datasets rooted in urban renewal housing records for a number of cities.

Cyberinfrastructure for the curation & management of digital assets at scale:

7. Big Data Archives and Analytics

See project descriptions for "Brown Dog" and "DRAS-TIC Fedora" under funding.

8. Enhancing User Access to Big Data Archives

In partnership with the Medical Heritage Library (a digital curation coalition) students will participate in developing user-friendly interfaces and workflows, and indexes to journals and medical terminology.