Computational Tools and Strategies in Enhancing Access to Cultural Big Data Collections

Richard MARCIANO & Ryan COX
marciano@umd.edu & ryan.cox@maryland.gov
Director & Research Archivist
Digital Curation Innovation Center (DCIC) & Maryland State Archives (MSA)

Tuesday, April 30, 2019
Columbus Metropolitan Library, Main Library, Columbus, Ohio
What are Computational Strategies?

Based on the concepts of Computational Archival Science (CAS), both case studies use visualization and analytical tools to connect and display data in new ways. The goal is to create transparency in cultural Big Data.

Case studies involve:

- automating the detection of personally identifiable information (PII) in Japanese-American World War II Incarceration Camps
- penetrating the complex pre-Civil War slave system in Maryland
What is CAS?

**GOAL:** Explore computational treatments of archival and cultural content

**PORTAL:** [http://dcicblog.umd.edu/cas/](http://dcicblog.umd.edu/cas/)

**GOOGLE GROUP:** computational-archival-science@googlegroups.com

A transdisciplinary 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.**

*Foundational Book Chapter:* May. 2018

**Book:** “Advances in Librarianship – Re-Envisioning the MLIS: Perspectives on the Future of Library and Information Science Education”.

**Book Chapter:** “Archival Records and Training in the Age of Big Data”
The Emergence of Computational XXX’s

- **XXX=Social Science**
  - “Investigating social and behavioral relationships and interactions through: social simulation, modeling, network analysis, and media analysis”, Wikipedia

- **XXX=Biology**
  - “The science of using biological data to develop algorithms or models to better understand biological systems”, Wikipedia

- **XXX=Journalism**
  - “Finding and telling news stories, WITH, BY, or ABOUT algorithms”, Nick Diakopoulos

- **XXX=Archival Science**
  - The Focus of this seminar…
Mission

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

- 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 engagements.

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 archival, retrieval and discovery/preservation and access decisions, and engaging and advancing research with archival materials.

CAS Founding Partners:
Richard Marciaclo, U. Maryland
Mark Hedger, King’s College London (UK)
Vicki Moulton, U. British Columbia (Canada)
Maria Sotera, Flex-Research Computing Center (FRC)
Michael Katz, U. Maryland
Bill Underwood, U. Maryland
Greg James, U. Maryland
Mark Conrad, National Archives and Records Administration (NARA)

CurateLab
Hombake South 411B
Digital lab for group learning, collaborative design, and hands-on digital curation project development (20 seats; 5 executive screens, 12 workstations with 128 of storage).

Digitization Lab
Hombake South 411DD
Document scanning, image manipulation, and archival injection facility for group projects.

Server Farm
On-campus virtual machine farm for research data processing, storage, and hosting (112 servers, 2 Dell servers, VMware powered).

CloudLab
Amazon Cloud
Desktop-optimized virtual computing lab is the Parallels Virtuozzo Containers/Amazon Web Services (AWS).

DataCave
CIR Cyberinfrastructure Center of the RiverTech Bldg.

Digital Repository (Azure Public clouds Compute for Sc Immersive Collections) is a scalable archival storage and preservation repository (based on the OAI-ORE open source software [SRDQ, Cassandra database] and computational infrastructure (4 Dell nodes).

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 engagements.

Motto:
"Integrating Education and Research"
An example:
Mapping Inequality - a focus on Big Data [Racial Zoning]

UMD Student Team:
Mary Kendig
Myeong Lee
Sydney Vaile
Maddie Allen
Martin Almirón
Jhon De La Cruz
Shaina Destine
Erin Durham
Darlene Reyes
Benjamin Sagay
Richard Bool
Historical Context

- Home Owners Loan Corporation 1930’s - 1940’s
  - Rated neighborhoods by racial makeup
  - Areas without loans fell apart
- 1950’s Urban Renewal targeted areas for clearance
- Result: Mass displacement
- RG 195: Federal Home Loan Bank Board, HOLC, 1933 - 1951
- Contains Maps, Neighborhood Surveys, Loan Information
Mapping Inequality

Documents
- Each survey corresponded to city map
  - **Green**: White / Wealthy = Best
  - **Blue**: White / Working = Still Desirable
  - **Yellow**: Foreign / Increase in PoC = Declining
  - **Red**: Black and Hispanic = Hazardous

Collection Statistics
- 150 Boxes
- Over 10,000 surveys alone
- 250 cities
Neighborhood description for: Boyle Heights in L.A.:
* Area D-53
* “Red” area

“Boyle Heights remained one of the most heterogeneous neighborhoods in the city for decades and it was a center of Jewish, Mexican and Japanese immigrant life in the early 20th century, and also hosted large Yugoslav and Russian populations.”

Wikipedia, 6/17/2011
AREA DESCRIPTION
Security Map of LOS ANGELES COUNTY

1. POPULATION:
   a. Increasing Slowly Decreasing Static
   b. Class and Occupation Jewish professional & business men, Mexican laborers, WPA workers, etc. Income $700 to $2000 and up
   c. Foreign Families 50% Nationalities Russian, Polish, Armenian, Jews... d. Negro 1%
      Slaves, Greeks, American Mexicans, Japanese and Italians
   e. Shifting or Infiltration Subversive racial elements increasing.

8. DESCRIPTION AND CHARACTERISTICS OF AREA:
   Terrain: Level to hillside with generally favorable grades and comparatively few construction hazards. Land improved 90%. This is a "melting pot" area and is literally honeycombed with diverse and subversive racial elements. It is seriously doubted whether there is a single block in the area which does not contain detrimental racial elements, and there are very few districts which are not hopelessly heterogeneous in type of improvement and quality of maintenance. Schools, churches, trading centers, recreational areas and transportation are all conveniently available. Many of the thoroughfares are arterial in character and constitute traffic hazards. This area is wholly in the City of Los Angeles. It is hazardous residential territory and is accorded a general medial red grade, although in many parts slum conditions prevail. The Federal Government, in conjunction with the city government are undertaking a slum clearance project covering 41 areas in the extreme northeast part of the area.

Mapping Inequality

http://mappinginequality.net
A. Historical Lab Notebooks

Paper-based Lab Notebooks:
• Used in science research
• Represent a record of:
  - observations
  - experiments
  - ideas
  - notes
  - formulas
  - data

Electronic Lab Notebooks:
• patient medical records
Learning Goals

- Archival Practices
- Computational Thinking Practices
- Ethics and Values Considerations
Motivation for Introducing Computational Thinking into Library and Archival Studies Curriculum:

- A basic understanding of the characteristics of digital materials is important for future libr. & archivists.
- Archival collections are increasingly composed of digital materials.
- The tools and practices associated archival activities are increasingly dependent on computing.
- The way users interact with archival collections reflects the increasingly computationally-mediated nature of our world.
- For today’s learners to succeed in future archival tasks, it is essential that computational thinking is included as part of their training.

Approach:

- Develop computational thinking enhanced lesson plans for archival topics that could be used by iSchool faculty to introduce computational thinking into their courses.
- Build an international networked community of iSchool faculty and Library and Archives practitioners to engender these capabilities?

Deep learning software to help botanists
Botanical specimen categorization at museums (5 million specimens)
Two big data analytics questions:
1. With what accuracy can a trained neural network sort mercury-stained plant specimens from clean ones? [90-94%]
2. With what accuracy can machine learning algorithms recognize members of two similar plant families? [96-99%]
“Responding to recent and exciting changes in scholarship that include the increasing use of digital data and computational methods”

“expanding and novel cross-disciplinary data reuse; new emphases on transparency and reproducibility; concern for preservation and long-term usability; and the rise of new forms of communication including data publication and open access publishing.”

“The successful candidate will... support new research methods being used by researchers as well as lead the adoption of new methods.”
ResearchWorks convened practitioners and researchers to shape a research agenda that charts engagement with data science and computational methods.

**Goal:** make the case for building networks and partnerships between libraries and data science and computational methods.

**E.G.:**
- Machine Learning multiplies connections – allowing for expanded discoverability.
- New Tools increase access to collections – allowing for progress on global information equity.
- A Range of Methods are used to analyze collections at scale – allowing for actionable insights that support sustainability and the realization of core values.

**Goal 2:** Identify key challenges matched with questions, methods, actions and grounded by carefully considered ethical commitments.

Beginning of a movement?
The new ways in which the public and researchers wish to engage with archival materials, are disrupting to traditional archival theories and practices.

The application of computational methods and tools to the archival problem space needs to be further explored.

The contextualization of records also needs to be explored, whether through:

- capturing metadata,
- enhancing records by semantic tagging,
- linking records with other records,
1. Computational Thinking (CT)
2. CT-Archives Mapping to Archives
   * Japanese American WWW Incarceration Camps
3. Publishing / Sharing Computational Stories Using Digital Notebooks
1. What is Computational Thinking?

“... a form of problem solving that uses modeling, decomposition, pattern recognition, abstraction, algorithm design, and scale”

---

**Viewpoint**

**Jeannette M. Wing**

**Computational Thinking**

It represents a universally applicable attitude and skill set everyone, not just computer scientists, would be eager to learn and use.

Computational thinking builds on the power and limits of computing processes, whether they are executed by a human or by a machine. Computational methods and models give us the courage to solve problems and design systems that no one of us would be capable of tackling alone. Computational thinking confronts the riddle of machine intelligence: What can humans do better than computers? and What can computers do better than humans? Most fundamentally it addresses the question: What is computable? Today, we know only parts of the answers to such questions.

Computational thinking is a fundamental skill for everyone, not just for computer scientists. To reading, writing, and arithmetic, we should add computational thinking to every child’s analytical ability. Just as the printing press facilitated the spread of the three Rs, what is appropriately inextricable about this vision is that computing and computers facilitate the spread of computational thinking.

Computational thinking involves solving problems, designing systems, and understanding human behavior, by drawing on the concepts fundamental to computer science. Computational thinking includes a range of mental tools that reflect the breadth of the field of computer science.

Having to solve a particular problem, we might ask: How difficult is it to solve? and What’s the best way to solve it? Computer science rests on solid theoretical underpinnings to answer such questions precisely. Stating the difficulty of a problem accounts for the underlying power of the machine—the computing device that will run the solution. We must consider the machine’s instruction set, its resource constraints, and its operating environment.

In solving a problem efficiently, we might further ask whether an approximate solution is good enough, whether we can use randomization to our advantage, and whether false positives or false negatives are allowed. Computational thinking is formulating a seemingly difficult problem into one we know how to solve, perhaps by reduction, embedding, transformation, or simulation.

Computational thinking is thinking recursively. It is parallel processing. It is interpreting code as data and data as code. It is type checking as the generalization of dimensional analysis. It is recognizing both the virtues and the dangers of aliasing, or giving someone or something more than one name. It is recognizing both the cost and power of indirect addressing and procedure call. It is judging a program not just for correctness and efficiency but for aesthetics, and a system’s design for simplicity and elegance.

Computational thinking is using abstraction and decomposition when attacking a large complex task or designing a large complex system. It is separation of concerns. It is choosing an appropriate representation for a problem or modeling the relevant aspects of a problem to make it tractable. It is using invariants to describe a system’s behavior succinctly and declaratively. It is having the confidence we can safely use, modify, and influence a large complex system without understanding its every detail. It is...
“Computational thinking is the thought processes involved in formulating problems and their solutions so that the solutions are represented in a form that can effectively be carried out by an information-processing agent.”

(Gun, Snyder & Wing, 2010)

**Computational Thinking Includes:**
- Abstractions and pattern generalizations (including modeling and simulations)
- Systematic processing of information
- Symbol systems and representations
- Algorithmic notions of flow of control
- Structured problem decomposition (modularizing)
- Iterative, recursive, and parallel thinking
- Conditional logic
- Efficiency and performance constraints
- Debugging and systematic error detection

(Grover & Pea, 2013)
## CT-STEM Practices Taxonomy

- **Data Practices**
  - Collecting Data
  - Creating Data
  - Manipulating Data
  - Analyzing Data
  - Visualizing Data

- **Modeling & Simulation Practices**
  - Using Computational Models to Understand a Concept
  - Using Computational Models to Find and Test Solutions
  - Assessing Computational Models
  - Designing Computational Models
  - Constructing Computational Models

- **Computational Problem Solving Practices**
  - Preparing Problems for Computational Solutions
  - Programming
  - Choosing Effective Computational Tools
  - Assessing Different Approaches/Solutions to a Problem
  - Developing Modular Computational Solutions
  - Creating Computational Abstractions
  - Troubleshooting and Debugging

- **Systems Thinking Practices**
  - Investigating a Complex System as a Whole
  - Understanding the Relationships within a System
  - Thinking in Levels
  - Communicating Information about a System
  - Defining Systems and Managing Complexity

- **- David Weintrop**
3. Automating the Detection of Personally Identifiable Information (PII) in Japanese-American WWII Incarceration Camps

Richard Marciano
William Underwood
The U.S. military targeted "all persons of Japanese ancestry," including both U.S. citizens and noncitizens, as a threat and removed them from their homes in the exclusion zone—designated areas on the West Coast.
Tule Lake, CA
(Incarceration Camp)

Arcadia, CA
(Assembly Camp, Santa Anita Racetrack, stables)

Pomona, CA
(Assembly Camp, LA Fairgrounds, racetrack, stables)
The records of the WRA (Record Group 210 from 1941-47) at the National Archives in Washington D.C. and Maryland, are comprised of over 100 series with motion picture films, drawings of incarceration centers, photos, maps, correspondence, yearbooks, rosters, etc.

Series 51 & 52 have immense value for survivors of the camps, their families, and historians, yet they are still not accessible.

Series 51, the “Internal Security Case Reports” from 1942 to 1946, comprises narrative reports prepared by camp investigators, police officers, and directors of internal security, relating cases of alleged “disorderly conduct, rioting, seditious behavior,” etc. at each of the 10 camps, with detailed information on the names and addresses in the camps of the persons involved, the time and place where the alleged incident occurred, an account of what happened, and a statement of action taken by the investigating officer.
364: Topaz UT (1%)
1,202: Poston AZ (5%)
1,578: Gila River AZ (6%)
 763: Granada CO (3%)
 533: Heart Mountain WY (2%)
2,146: Manzanar CA (9%)
2,343: Minidoka ID (9%)
15,648: Tule Lake CA (63%)
 468: Rohwer/Jerome AK (2%)

25,045
Graph Database Modeling and Visualization: Movements of Satsuki Ina’s Family within the Camp System

Graph Database Modeling and Visualization: People Deported to Tule Lake Clustered by City of Origin

Interactive Map of Tule Lake

Spring 2015
A. Collecting Data
B. Creating Data
C. Manipulating Data
D. Analyzing Data
E. Visualizing Data
F. Designing Computational Models
G. Programming
H. Developing Modular Computational Solutions
I. Troubleshooting and Debugging
G. Preparing Problems for Computational Solutions
H. Choosing Effective Computational Tools
I. Assessing Different Approaches/Solutions to a Problem
J. Assessing Computational Models
K. Constructing Computational Models
L. Thinking in Levels
M. Communicating Information about a System
N. Defining Systems and Managing Complexity
O. Investigating a Complex System as a Whole
A. Creating Data

“The increasingly computational nature of working with data in” archival science “underscores the importance of developing computational thinking practices in the classroom.” “Part of the challenge is teaching students that answers are drawn from the data available.” “In many cases” archivists “use computational tools to generate data... at scales that would otherwise be impossible.”
<table>
<thead>
<tr>
<th>Last Name</th>
<th>First Name</th>
<th>Birth Year</th>
<th>Original State</th>
<th>Gender</th>
<th>Birth Place</th>
<th>Family No</th>
<th>Individual No</th>
<th>File Number</th>
<th>Assembly Center</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABE</td>
<td>FRANK</td>
<td>1910</td>
<td>CA</td>
<td>M</td>
<td>CA</td>
<td>24067</td>
<td>24067A</td>
<td>208156</td>
<td>None</td>
</tr>
<tr>
<td>ABE</td>
<td>FRANK</td>
<td>1940</td>
<td>CA</td>
<td>M</td>
<td>CA</td>
<td>24067</td>
<td>24067</td>
<td>201888</td>
<td>None</td>
</tr>
<tr>
<td>ABE</td>
<td>FRANK</td>
<td>1905</td>
<td>CA</td>
<td>M</td>
<td>Honolulu county</td>
<td>8605</td>
<td>08605A</td>
<td>950783</td>
<td>Fresno</td>
</tr>
<tr>
<td>ABE</td>
<td>FRANK</td>
<td>1913</td>
<td>CA</td>
<td>M</td>
<td>Oregon</td>
<td>18050</td>
<td>18050B</td>
<td>805536</td>
<td>Santa Anita</td>
</tr>
<tr>
<td>ABE</td>
<td>FRANK</td>
<td>1881</td>
<td>CA</td>
<td>M</td>
<td>Sakhalin</td>
<td>34424</td>
<td>34424C</td>
<td>207865</td>
<td>None</td>
</tr>
</tbody>
</table>

---

**WRA Form 26 register**

“Japanese-American Internee Data File”

NARA AAD

---

**Final Accountability Rosters (FAR)**

Box 8 -- #269

---

**3-23-45 A-1067 INFRAC. PROJ. REG.**

**ABE, Frank Tomo**

1514-A

The above was put in project jail for military marching, blowing of bugles, display of Japanese emblems. Occurred in the colony.
“Computational tools make it possible to efficiently and reliably manipulate large and complex” archival holdings. “Data manipulation includes sorting, filtering, cleaning, normalizing, and joining disparate datasets.”

<table>
<thead>
<tr>
<th>Japanese Name</th>
<th>Last Name</th>
<th>First Name</th>
<th>Other Name</th>
<th>Incident Date</th>
<th>Year</th>
<th>Age</th>
<th>Residence ID</th>
<th>Family Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Abe</td>
<td>James</td>
<td></td>
<td>10/5/1942</td>
<td>1942</td>
<td>20</td>
<td>2803-A</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>Abe</td>
<td>Makoto</td>
<td></td>
<td>10/7/1942</td>
<td>1943</td>
<td>43</td>
<td>1206-A</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>Abe</td>
<td>Sakichi</td>
<td></td>
<td>11/2/1942</td>
<td>1942</td>
<td>62</td>
<td>5315-B</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>Abe</td>
<td>Shigeki</td>
<td></td>
<td>5/1/1943</td>
<td>1943</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>Abe</td>
<td>Kiyotake</td>
<td>Joe</td>
<td>11/4/43</td>
<td>1943</td>
<td></td>
<td>2904-D</td>
<td>40562</td>
</tr>
<tr>
<td>N</td>
<td>Jensen</td>
<td>Lloyd</td>
<td>H.</td>
<td>11/16/43</td>
<td>1943</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
C. Analyzing Data

“There are many strategies that can be employed when analyzing data for use in” an **archival context**, “including looking for patterns or anomalies, defining rules to categorize data, and identifying trends and correlations.”

- We used NER software to extract metadata from the incident cards. This was done with the open source GATE. This is based on pattern matching through recognition rules. The matching rules are often refined through iterative tuning.
- For example, a rule for recognizing a person’s name would be based on a lastname, followed by a comma, followed by a Japanese firstname, followed by an Anglo first name in parentheses. As we process additional cards we would note that there are other styles of names, so the pattern would be generalized account for stylistic variations. If the pattern is made to be robust enough it will eventually work on all of the instances of names.
- GATE, General Architecture for Text Processing, [https://gate.ac.uk/](https://gate.ac.uk/)
D. Visualizing Data

“Communicating results is an essential component of understanding archival data and computational tools can greatly facilitate that process. Tools include both conventional visualizations such as graphs and charts, as well as dynamic, interactive displays.”

Box 8  WRA Form 26  FAR Tule LAke
E. Designing Computational Models

“The ability to create, refine, and use models of phenomena is a central practice.” “Models can include flowcharts and diagrams.” “Part of taking advantage of computational power… is designing new models that can be run on a computational device.” “There are many reasons that might motivate designing a computational model, including wanting to better understand a phenomenon under investigation, to test out a hypothesis.” “Students… will be able to define the components of the model, describe how they interact, decide what data will be produced by the model.”
An important practice is the ability to create new or extend existing computational models. This requires being able to encode the model features in a way that a computer can interpret.

FOR EACH of the 10,000 cards:
  IF cardname is Japanese:
    MATCH cardname in the FAR registry
    IF MATCH is true:
      COMPUTE date difference
      IF difference > 18:
        RELEASE card
    ELSE:
      DO NOT RELEASE card: possible PII
  ELSE:
    MATCH cardname in the Form 26 registry
    IF MATCH is true:
      COMPUTE date difference
      IF difference > 18:
        RELEASE card
    ELSE:
      DO NOT RELEASE card: possible PII
    ELSE:
      PII Status NOT DETERMINED
  ELSE:
    RELEASE card: WRA Staff – Not Internee
“Enabling students to explore” **archival problems** “using computational problem solving practices such as programming, algorithm development, and creating computational abstractions.” “The ability to encode instructions in such a way that a computer can execute them is a powerful skill for investigating” **archival problems.** Programs include ten-line Python scripts.”
H. Developing Modular Computational Solutions

“We make use of abstraction and functional programming through the use of modular components such as:
- PII_DateCheck(),
- FORM26_lookup(), and
- FAR_lookup().

This allows for reusable chunks of code that can be tested locally. The larger program is the composition of these modules, which makes it both more readable and maintainable.”
I. Troubleshooting and Debugging

“Troubleshooting broadly refers to the process of figuring out why something is not working or behaving as expected. There are a number of strategies one can employ while troubleshooting a problem, including clearly identifying the issue, systematically testing the system to isolate the source of the error, and reproducing the problem so that potential solutions can be tested reliably.”

- To facilitate group debugging, we use an interactive server-based shared version of Jupyter Notebook.
  - “The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.”

- Project Jupyter, see: http://jupyter.org/
Educators are rapidly adopting Jupyter Notebooks for:

* teaching
* use in the classroom
* developing teaching materials
* creating computational stories

See: https://cases.umd.edu
3. Opportunities for Collaboration

1. Best Practices Exchange (BPE)
Computational Tools and Strategies in Enhancing Access to Cultural Big Data Collections
• April 30, 2019, Columbus, Ohio
https://bpexchange.wordpress.com/2019-conference/

2. Records Management Journal – Themed call for papers
‘Technology and records management: disrupt or be disrupted?’
• Extended: May 1, 2019
• Full paper submitted: September 1, 2019
• Review, revision and final acceptance: January 31, 2020

3. AERI (Archival Education and Research Institute) 2019
Developing a Computational Curriculum Framework for Archival Education
• July 11, 2019, Liverpool, UK
https://aeri2019.com/

4. ARA (Archives & Records Association – UK & Ireland) Conference
Shaping Digital Recordkeeping Competence
• August 28, 2019, Leeds, UK
https://conference.archives.org.uk

5. Computational Archival Science Workshop (CAS#4)
IEEE Big Data 2018, December 9-12, Los Angeles
http://dcicblog.umd.edu/cas/ieee-big-data-2018-3rd-cas-workshop/
CONTACTS

marciano@umd.edu
@umdDCIC
http://dcicblog.umd.edu/cas/
computational-archival-science@googlegroups.com