Heuristics for Assessing Computational Archival Science (CAS) Research: The Case of the Human Face of Big Data Project

Myeong Lee, Yuheng Zhang, Shiyun Chen, Edel Spencer, Jhon Dela Cruz, Hyeonggi Hong, Richard Marciano
Is that CAS as a new field?
Computational Archival Science

- Proposed trans-disciplinary field
- Combines computational and archival thinking
- Applies to large-scale records/archives processing
- Employs computational methods and resources
- Supports traditional archival functions and goals
Purpose of CAS

- Undertake research with archival materials
- Apply existing collective knowledge to understand how new technologies change archival practices
- Discover implications of these changes for archival science and the use and preservation of records
- Improve efficiency, productivity and precision
Problem:
It is unclear what constitutes a CAS study

Proposed Solution:
- Guidelines and Checkpoints Needed
- Use Heuristics to map studies on the CAS spectrum
Computational Thinking

- Iterative process
- Heuristic reasoning used to discover and solve problems
- Has been adopted by many fields

Image source: https://cspathshala.org/
## Theoretical Blending

<table>
<thead>
<tr>
<th>Combination Type</th>
<th>Boundaries</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-disciplinary</td>
<td>Combination of two separate fields of knowledge</td>
<td>Remain stable</td>
</tr>
<tr>
<td>Inter-disciplinary</td>
<td>Intersection of two fields</td>
<td>Blurred</td>
</tr>
<tr>
<td>Trans-disciplinary</td>
<td>Recombination of elements from two fields</td>
<td>Merged</td>
</tr>
</tbody>
</table>

CAS is transdisciplinary because it combines archival and computational sciences.

Image source: https://thedailyomnivore.net/
Computational thinking and transdisciplinarity are two dimensions of the conceptual CAS spectrum.

Positioning each CAS study on the conceptual CAS spectrum

Perceptual Map of CAS Spectrum

Figure 1: A perceptual map of CAS spectrum.
### Proposed Heuristics

How our project fits into CAS:

<table>
<thead>
<tr>
<th>Heuristic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifying the kinds and depth of computational</td>
<td>The extent to which computational thinking is practiced in solving archival problems, and how advanced it is.</td>
</tr>
<tr>
<td>thinking</td>
<td></td>
</tr>
<tr>
<td>Blending theories from both fields</td>
<td>Whether theories from both archival science and computational science fields are blended, and whether a new theoretical basis is created based on the fusion.</td>
</tr>
<tr>
<td>Positioning research in and around CAS studies</td>
<td>Given the kinds/depth of computational thinking and theoretical blending as a two-dimensional spectrum, and initial CAS studies presented in the foundational paper, a study can be placed on the perceptual map of the CAS spectrum.</td>
</tr>
</tbody>
</table>

1965-1978 Asheville, North Carolina:
- Urban renewal project to transform “blighted” neighborhoods into modern housing areas
- Affected 4000 people and nearly 1300 households of East Riverside neighborhood called “South Side”

Human Face of Big Data:
- Records of Southside property acquisition process:
- Each record represents one property’s documents
- Each record contains 20 to 200 pages of legal documents
- Project has possession of records of 1000 properties
- Project’s purpose is to show the property acquisition process
Digital Curation

For Historical Map (Created in 1973):
- Paper Map
- Scanning & Adjusting
- Geo-Referencing
- Geo-Tracing

For Urban Renewal Documents:
- Top-Down Approach:
  - User Persona
  - User Scenarios
- Data Modeling
- Transcribe Data From Documents Manually
System Design

- **Interface Design:**
  - Hand Drawings for Interface Design
  - The Interface Wireframe

- **Database Design:**
  - Database Schema:
    - Event-Based
    - Spatio-temporal Model
  - Open Source Relational Database System (PostgreSQL)
System Design Process

- Implementation:

Front-End

Back-End

Tools

Prototype
System Prototype

Popup: To see exactly what happened to a parcel with all information.

Search Function: To search for a name and/or a parcel street name.

Color Coding: To see how the entire city was renewed during the project period.

Statistical Results: To see trends of activities.
Heuristic Analysis

Is the Human Face of Big Data Project on the CAS spectrum?

1. Does it employ computational thinking processes?
   - Yes - uses software development practices, data visualization, data management and event-based modelling and HCI models

1. Does it blend theories from separate fields?
   - Yes - combines archival concepts with system design and agile development to create a new body of knowledge

3. How can the project be positioned within the realm of CAS themes?
   - System design and VSD are more about humanizing and specializing archival systems rather than generalizing.
   - The features make system design and VSD unique CAS themes.
Discussions:

- It is possible to map a CAS-related study on the CAS spectrum, and to provide a reference point of how it is situated in this emerging field.

- The advancement level of computational thinking can be subjective to some degree.

- The CAS-related studies need to be evaluated from diverse perspectives of computational thinking.
Conclusion and Future Work

- Our heuristics are not perfect, but still useful in assessing what constitutes a CAS study.
- The spectrum of computational thinking needs to be modeled more precisely and further conceptualized.
- The heuristics need to be examined and refined through diverse case studies.
- We invite researchers to use and refine these heuristics through their research work.
Thank You!

Any Questions?

http://DCIC.umd.edu

Email:  dcic.ischool@gmail.com