

# On Topic Evolution, Selection, and Future Work

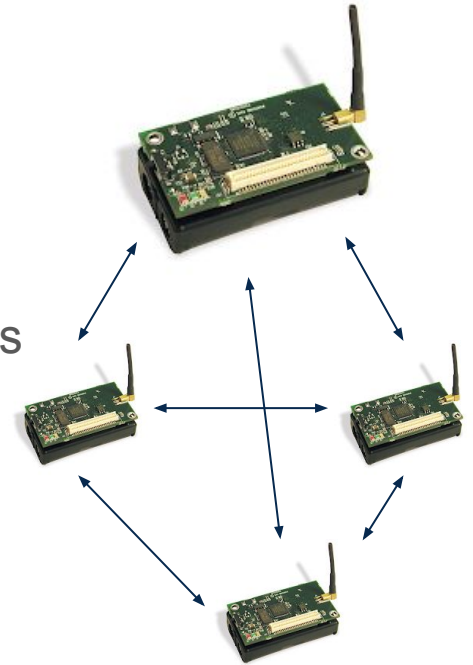
**Mat Kelly, PhD**

Assistant Professor, Information Science  
Drexel University  
College of Computing & Informatics  
mkelly@drexel.edu  
<https://matkelly.com>



# A Research Trajectory By Example

- Computer Science (CS) undergrad
- Both MS & PhD in CS (2019), as well
  - Eventual focus re:web archive
- Started MS in NSF-funded research in embedded systems
  - Emergency retasking of sensor networks
  - Wanted to get hands-on with hardware & low level programming
  - Piqued my interest in *decentralized systems*



# A Research Trajectory By Example

- **Advisor** and I transitioned labs
  - Networking Research Group (**her** research area)
  - ...to *Web Science and Digital Libraries (WS-DL)* Research Group
- First project (**ArchiveFacebook**) w/ WS-DL, still MS student
  - Voluntarily work to update software to be more robust, adaptable
  - Eventually served as the basis for my MS Thesis



Facebook Native Profile Download



Archive Facebook Archiving Session Result



# Step Backward, as a CS undergrad...

- Web Designer
- Web Accessibility Analyst / Developer
- Managing Editor at a newspaper
- Job post-BS legally ripping artist's digital promotional materials

# Step Backward, as a CS undergrad... **Self-Identification**

- Web Designer / Programmer
  - design fundamentals, presentation, data/content management
- Web Accessibility Analyst / Developer
  - working with bleeding edge standards, ensuring equity of access
- Managing Editor at a newspaper
  - time & personnel management, commercial business sense
- Job post-BS legally ripping artist's digital promotional materials
  - data liberation, circumventing unjust barriers

# Incorporate Your Unique, Past Experience

- Worked as web developer w/ focus on web accessibility (e.g., §508)
  - The Archival Acid Test, published @ ACM/IEEE JCDL '14
- Funded final portion of my BS working as a managing editor of a newspaper
  - Served as the basis for a self-support grant while a PhD student
- Identify the facets of the topic that interest you
  - Me: decentralization, data liberation, digital privacy, interoperability
- Get your hands dirty, determine your niche (here: via programming)
  - ArchiveFacebook → archiving → private web archiving



**Dodging the Memory Hole  
2016 Scholarship Projects**



# My Dissertation\* Research Questions (RQs)

- **RQ1:** What sort of content is difficult to capture and replay for preservation from the perspective of a Web browser?
- **RQ2:** How do Web browser APIs compare in potential functionality to the capabilities of archival crawlers?
- **RQ3:** What issues exist for capturing and replaying content behind authentication?

---

- **RQ4:** How can content that was captured behind authentication signal to Web archive replay systems that it requires special handling?
- **RQ5:** How can Memento aggregators indicate that private Web archive content requires special handling to be replayed, despite being aggregated with publicly available Web archive content?
- **RQ6:** What kinds of access control do users who create private Web archives need to regulate access to their archives?

Answered prior to candidacy exam

Themed around **data liberation** and **interoperability**

Novel problems that would arise because of my work

Informed by standards and practice to new domain

# Research Products

- Publications
  - conference proceedings, journals, tech reports, blog posts
- Interdisciplinary dissemination
  - conferences, org. meetings, guest presentations
- Software\* - to support but also facilitate my research topic
  - ...which served as the basis for my **graduate funding** (thanks PhD advisors!)



**NATIONAL  
ENDOWMENT  
FOR THE  
HUMANITIES**

\* All free and open-source @ <https://github.com/machawk1>





# Ongoing Research

- Central theme:
  - Computational negotiation & navigation in dimensions beyond time
- Has extended to computationalizing historical ontologies
  - Library of Congress Subject Headings (LCSH) 1910-present
- Decentralized persistent identifiers
  - Archival Resource Keys (ARKs)
- Coalescing the targets of these subject areas
  - While web archives use a URI as the key for context...
    - single dimension (time)
  - ...terms in historical vocabularies have more complex evolutions
    - e.g., splitting, deprecation, combination



**ARK Alliance**

Home of the Archival Resource Key (ARK)



# Converging on Unification, Seeking to Marry...

- Linked data (à la RFC7089: Memento)
- Historical vocabularies (LCSH over time)
- Persistent identifiers (particularly the decentralized sort)
- Semantic association (supported by LCSH exploration)
- Web archives (aspects beyond a URL over time)



**More info, pubs, software, etc.**