



**International Research Collaboration
Network in Computational Archival Science**

Mark Hedges

Department of Digital Humanities, King's College London

Who are we?

Who are we?

- Department of Digital Humanities, King's College London
- “Pioneering” department in digital humanities
- A broad view of DH (‘big tent’):
 - Role of technology in arts, humanities, and social sciences
 - Impact of digital tech on society and culture
 - Digital culture
 - Digital economy and innovation
 - Digital archives, content, assets
- But we are not an archives or LIS department/school

Collaboration with DCIC on CAS

- Symposium at DCIC in April 2016
- Organised various workshops (e.g. at IEEE Big Data conferences)
 - Activities this week as well
- Worked together of defining and refining ‘computational archival science’, publications.
- First jointly-funded project – IRCN-CAS

IRCN-CAS Network

Aims of the funding scheme

- Funded by AHRC (Arts and Humanities Research Council) in UK
- “Network Grant”
- To support new collaborations, exchange of ideas, crossing boundaries.
- Funds ‘networking activities’: conferences, workshops, symposia, reciprocal visits
- UK–US Collaborations in Digital Scholarship in Cultural Institutions



Arts & Humanities
Research Council

Core Partners of Network

- In the UK:
 - King's College London (Dept of Digital Humanities)
 - The National Archives
- In the US:
 - University of Maryland (Digital Curation Innovation Center)
 - Maryland State Archives
- These are the **core** partners – the Network will strengthen and also extend partnerships to facilitate further research in this area.

Why are we doing this? Motivations and objectives

- Build up a research network in CAS – in particular aspects relating to *digital scholarship*.
- Build *systematic* collaborations that leverage existing work
- Focus on the use of computational methods for *contextualisation* of records in archives
- How to embed in actual archival practice (e.g. at TNA)
- Holistic view, bringing together: archivists, LIS and DH schools, data scientists, computer scientists, scholars who use archives (e.g. historians)
- We do different things, but we all learn from one another, and we all benefit – experiences learned from DH

ENRICH DESCRIPTIONS

- Extract: dates, title & first para, keywords. Chapter by events & changes.
- Rich snippets and generous interfaces

COLLECT MORE METADATA

- Multiple dates: created, modified, probabilistic
- Widen capture – technical, harvesting, creators' knowledge

CLUSTER BY TOPIC

- Cross-link similar subjects across collections
- Classify / tag / categorise records for filtering
- Graph UI representations for graph data

JOIN SIMILAR RECORDS

- Join up duplicates from different sources, join multiple versions e.g. in web archive & catalogue
- Confidence and uncertainty

STRUCTURE & PROVENANCE

- 7 pillars of metadata
- More granular creator info e.g. depts. within OGD
- Personalised context?

LINK COLLECTIONS

- Internal & external
- Retrieve relevant results from across collections
- 'Trusted' external links: BBC, Hansard, Web archives ... TNA wiki

Activities, Events, Timeline

- Funded project runs Feb. 2019 – Jan. 2020
- Main events:
 - Initial meeting/symposium (April 2019)
 - Data event at The National Archives (June 2019)
 - Data event at Maryland State Archives (October 2019)
 - **Final symposium in London (January 2019)**
- Other events as opportunities arise:
 - AERI panel (Archival Education and Research Initiative) – July 2019
 - ARA panel (Archives and Records Association) – Aug. 2019
 - IEEE Big Data workshop in Dec. 2019 (Los Angeles)

First data workshop

- Worked with The (UK) National Archives' collections.
- Interdisciplinarity. Bring experts from different disciplines around the table: Raise new questions, find new solutions, develop a common language and engagement.
- Experimentation and collaboration.
- Opening the black box of digital research: exploring data, understand the methodological and theoretical challenges, decision-making and all steps involved in the research process.

Working Groups and Outcomes

- Worked in small groups.
- The groups were interdisciplinary.
- Enabled cross-disciplinary practices and experimentation: Explore challenges and opportunities, find commonalities.
- Captured the discussion.

Groups

UK Cabinet Papers

- Topic modelling

UK government web archive

- Network Analysis
- How is the government web linked together at different “points” in time?

War Diaries (official Unit War Diaries from WW1)

- Information extraction, text mining, visual analytics

UK legislation data (XML)

- Visualisation, visual analytics
- Analysis of structure of the legislation as a corpus
- Interrelationships, diachronic views

Video and image archives

- Videos from 2012 London Olympics and UK Supreme Court; watermark archive
- Neural networks

Documenting the process

- Selecting an approach
- Decisions and decision-making
- Processes of collaboration
- Obstacles and challenges
- Interdisciplinary working
- How to take things forward from exercise to production

Outcomes

- Sustainable *Network*
- Continuation activities
 - Other grants
 - Publication, white papers
 - Extending network, ...
 - New collaborations
- Website: <https://computationalarchives.net/>
- Final event: <https://www.turing.ac.uk/events/computational-archival-science-cas-symposium>

Thank you

Contact details/for more information

Mark Hedges

Department of Digital Humanities

King's College London

mark.hedges@kcl.ac.uk