

AURA Workshop 3, Abstracts and Biographies

Andrew Prescott, 'Can Archives make AI Better?'

The huge scale of born-digital archives means that archivists will inevitably have to deploy AI in their future management. AI will play a vital role in appraising born-digital records, analysing their data, generating metadata for discovery and ensuring their sustainability. However, commercial and government deployments of AI frequently exhibit racial and gender bias, because of inadequate training sets and lack of diversity among developers. Will this pose problems for the deployment of AI in archives? How does a technocratic view of the role of AI in archives conflict with the recent growth in community archives? Possibly the use of AI in exploring and helping preserve community archives can contribute to the development of a more community-oriented, democratic and ethical AI.

Biography:

Andrew Prescott is Professor of Digital Humanities in the School of Critical Studies at the University of Glasgow. He was from 1979-2000 a curator in the Department of Manuscripts at the British Library. He has worked in libraries and digital humanities units at the University of Sheffield, University of Wales Lampeter and King's College London. He was from 2012-2019 Theme Leader Fellow for the AHRC Strategic Theme of Digital Transformations.

Larry Stapleton, 'The INSTYE Cooley Laboratory, Waterford Institute of Technology: A Survey of Key Themes, Recent Research and Future Directions'

The Cooley Lab was established in 2020 following four years research on digital archives and human-centred systems development for knowledge repositories. It was named after Professor Mike Cooley, following a donation from the Cooley family of his library and private correspondence, which constitutes a national treasure of great importance. In 2019 the Cooley family kindly gave WIT permission to use Professor Cooley's name on the new INSTYE-Cooley Research Laboratory housed at the Digitisation Hub of the Luke Wadding Library. At a time when the onward march of digital technology is transforming working life and human relationships, Professor Cooley's vision has never been more important and the INSTYE Laboratory team seek to continue the Cooley legacy in their work. WIT information systems research and postgraduate teaching programmes embody his vision of a world in which human dignity is enhanced by advanced technologies. This presentation reviews the recent work of the laboratory and identifies key developments for systems engineering research in the digital humanities with special reference to the care and management of digital cultural heritage.

Biography:

Dr Larry Stapleton is a leading international expert in the development and social impact of advanced digital systems. He has lectured at Waterford Institute of Technology for over 30 years where he is the founder and Director of the Centre for Information Systems and Techno-Culture (INSYTE) which organises postgraduate research and teaching programmes for over twenty years. In his consulting practice he has advised private companies, NGOs, business development agencies, universities, regulators, the European Commission, and governments both inside and outside the EU. He holds

several international professor positions and has won “Best International Professor” awards, publishing over 140 academic peer-reviewed publications in digital systems, culture and related topics. He holds many international academic positions on various boards, including several journal editorial boards and is editor of the Springer journal “Artificial Intelligence and Society”. He is technical Board member at the International Federation of Automation and Control (an organisation comprising several thousand members across over fifty countries) and Honorary Chair for the Technical Committee for Technology, International Stability and Culture as well as Coordinating Chair for the Social Effects of Automation and Control Systems. In 2020 he won the prestigious “IFAC Outstanding Achievement Award” which was presented at the IFAC 2020 World Congress in Berlin.

Angeliki Tzouganatou, ‘AI, openness and participation in digital cultural archives’

During the past decades, Galleries, Libraries, Archives and Museums (GLAM) have been grappling with emerging technologies in an effort to both attract and build new audiences, as well as to improve organisational and management needs. Artificial Intelligence (AI) promised to make people’s lives easier and more dynamic, while opening up cultural heritage data online in a more interactive way. In the GLAMs landscape, AI has manifested in various ways, ranging from intelligent search of text and image recognition, to digital storytelling experiences, conversational interfaces (i.e. chatbots, and virtual agents), to name a few. Advances towards the accessibility and re-usability of cultural data, have been realized by the Findable, Accessible, Interoperable and Reusable (FAIR) guiding principles (Wilkinson et al. 2016), recently adopted by GLAMs, aiming towards a fair data ecosystem. Moreover, progressions concerning the opening up of digital archives have been conducted via Application Programming Interfaces (API) as well, creating new opportunities to be flourished, and the possibility for new projects to be connected with the creative industries.

The times of merely “consuming” big data have been replaced by a call to use and re-use these enormous amounts of data creatively, in a fair way, fostering meaningful participation. The talk will discuss human-centered approaches to AI for digital cultural archives, inquiring AI’s role in terms of openness and participation. How is AI compatible with these values, and what does it take to envision inclusive intelligent systems for digital cultural archives?

Biography:

Angeliki Tzouganatou is a Marie Skłodowska-Curie POEM (Participatory Memory Practices) PhD Fellow at the Institute of Anthropological Studies in Culture and History at the University of Hamburg. She studied Archaeology & History of Art at the National and Kapodistrian University of Athens and the University of Bologna, and completed the MSc in Digital Heritage at the University of York. She has work experience in various cultural heritage institutions and has been involved in research projects such as the EMOTIVE project and ‘Co-Production Networks for Community Heritage in Tanzania’, as a research assistant in visualization and interpretation of cultural heritage via digital media technologies.

Jennifer O’Mahoney, ‘The role of born digital data in confronting a difficult and contested past through digital storytelling’

The Waterford Memories Project is an oral history driven study in digital humanities, publicly documenting survivor narratives of the Magdalene Laundries and Industrial Schools in the South-East of Ireland. These institutions existed as part of a system of coercive confinement, which refers to a wide range of institutions used to confine both children and adults whose “crimes” were to act against the strict and punitive moral codes of the period, poverty or mental illness. These institutions existed until the mid 1990s.

The State archives of the Commissions of investigation into the Magdalene Laundries and Industrial Schools are inaccessible to researchers and members of the public; the Religious Orders which ran these institutions have also embargoed their archives. These restrictions to historical data ensure a fractured and incomplete narrative of the interconnected cultural, social, and economic trends, which allowed the Industrial Schools to exist until the mid-1970s and the Magdalene Laundries to exist until the mid-1990s.

The oral histories are born digital data, in the form of audiovisual recordings, which are openly disseminated via the project website. This talk will consider the role of the born digital survivor testimony in challenging the silencing and inaccessibility of both the State and Religious Orders’ archives; confronting a difficult and contested past through (digital) truth-telling. Culture can become fossilised in systems and continue to perpetuate the silencing of survivors; it is therefore essential to consider how the openly available digital testimony contributes to the framing of cultural discourse around our history of coercive confinement in Ireland.

Biography:

Dr Jennifer O’Mahoney is a Lecturer in Psychology and a Chartered Psychologist with the Psychological Society of Ireland. She is Co-Director of both the Analysing Social Change Research Group and Crime and Justice Research Group at WIT, and a Senior Researcher at INSYTE (The Centre for INformation SYstems and TEchno-culture) at WIT. Her research focuses on how victimology and trauma are remembered and narrated; the relationship between memory and cultural heritage in digital humanities; and activism and social change.

Giorgio Maria Di Nunzio, ‘Bais and Fairness in AI: New Challenges with Open Data?’

AI tools are becoming pervasive in our lives, and we use these tools because they are “better,” on average, than a human decision-maker. Should not an algorithm be fair and free of human biases? (Borenstein and Howard, 2021). In order to develop an inclusive and “fair” AI, we need to address some issues that may emerge from the point of view of gender, ethnicity, personal and social development of Artificial Intelligence algorithms.

In this presentation, we want to analyze the current situation about methodologies and techniques aimed at identifying and treating such possible biases introduced by AI tools. In fact, algorithms shouldn’t be burdened with the biases derived from our lived experiences, but an algorithm is made by humans, and humans make designing, programming, and evaluation mistakes. Can we ensure that AI tools aren’t biased? Can we define “fairness”? We will discuss these issues in the context of massive open born-digital archive records.

[Borenstein and Hoard, 2021] Borenstein, J., Howard, A. Emerging challenges in AI and the need for AI ethics education. *AI Ethics* 1, 61–65 (2021). <https://doi.org/10.1007/s43681-020-00002-7>

Biography:

Giorgio Maria Di Nunzio is Associate Professor at the Department of Information Engineering of the University of Padua (UNIPD), Italy. He is the Representative of UNIPD for the European Open Science Cloud (EOSC) Association. He is also the Representative of the Digital Library and Open Science Project, Member of the Technical and Scientific Committee of the University Library Center, and Member of the Inclusion and Disability Commission of the Department of Information Engineering at UNIPD.

Mathieu D’Aquin, ‘AI for archives and collections: From processing metadata to analysing content’

In this talk, I reflect on the experience in several projects with the aim to support humanities researchers in exploiting large collections through the application of AI techniques. A common aspect of those projects is the need for dedicated approaches that embed not only the technical process of AI, but also the knowledge, methods and practices of the domain. Through examples in literature, iconography and music, I discuss in particular the necessary dialogue between domain experts and AI experts, and how this dialogue can help evolve both the research methodologies in the humanities and AI techniques in a way that goes beyond the specific applications.

Biography:

Mathieu d’Aquin is established professor of data analytics, director of the Data Science Institute and site-director of the Insight SFI Research Centre for Data Analytics at NUI Galway. His research focuses on data and knowledge engineering challenges, through solutions that combine traditional artificial intelligence methods, semantic web approaches and machine learning. He has applied this research in a large variety of application domains, including healthcare, education, smart cities and the humanities.

Bram Van Der Werf, ‘What comes next? AI for discovery or destruction?’

In the age of big data, digital archives are big data and AI has the potential to unlock them. Meanwhile, archivists and digital preservation researchers consider AI as a means to help them tackle digital archives as paper archives, with the analogue world constraints of selection and destruction. Using AI to automate the appraisal process and reduce the growing backlogs of born-digital collections is probably the most attractive solution for both governments and their archives. This would help them “capture the truth”, destroy the rest, and perpetuate the euphemism of appraisal in the digital age. However, there are no constraints to keeping everything digital. It is a myth that keeping data long-term is costly. So, why use AI for selection? Make sure you keep everything and let future generations use AI for discovery and re-use!

Biography:

Bram van der Werf is an independent consultant. He contributed to a number of European heritage initiatives: as technical director of Europeana, executive director of the Open Planets (Preservation) Foundation and advisor-project leader of PrestoCentre on behalf of the Dutch Sound and Vision Archive.

Bram has an academic training in quality management and rich practical experience in the IT industry. He has a deep understanding of the societal challenges that come with the digital transformation and is able to unpack complex trends for heritage professionals.

Mark Bell and Jenny Bunn, ‘Archives and AI: What now?’

This presentation will seek to make explicit the sort of processing, or reasoning over records, that is currently taking place, and to question the extent to which this processing (by both human and machine agents) can and should be seen as analogous. In a climate of increasing demands that AI should explain its own process, it will start by reviewing the sort of explanations that have been offered by human archivists and researchers in regard of the processing and processes that they carry out upon records and archives. This will then be contrasted with an explanation of some of the automated techniques which are now being applied to assist in reasoning over records.

The purpose of this comparison will be to cut through hyperbole and to prepare a more grounded understanding of what we are talking about now, from which discussions about what comes next can continue. To be sure some of these new and automated ways of processing appear to happen very quickly and faster processing, the argument goes, means that you can process more. Just because you can though doesn't necessarily mean that you should. Even when there are external pressures on you to do so, even when you seemingly have no choice but to do so, it is still perhaps wise to consider (however briefly) how much and in what ways the forms of processing in which you are reinventing your older ones are or are not analogous.

Mark Bell Biography:

Mark Bell is a Senior Digital Researcher at The National Archives, UK, where he has worked since 2014. Prior to joining the Archives he worked for almost 20 years with large scale databases in a number of private and public sector organisations. His research interests include the application of AI to archival collections, probabilistic record linkage, and blockchain.

Jenny Bunn Biography:

Jenny Bunn is Head of Archives Research at The National Archives. She started her career working in a number of archive services including those at The Royal Bank of Scotland and the V&A Museum, before undertaking a PhD on the subject of archival description in 2007-11. From 2010 she worked at University College London, teaching and latterly directing the MA in Archives and Records Management programme, before moving into her current role in 2020.

Dr Adam Nix, ‘Finding light in dark archives: Using AI to connect context and content in email’

The landscape of digital archival discovery is still emerging, and the approaches future historians will take when using digital sources still unclear. Paradoxically, this is not helped by the fact that many born-digital collections remain dark, in part while questions of how they should be effectively made available are answered. Currently, historical methodology largely assumes the use of original, pre-digital sources, something that guides much of the theory around effective source criticism and the often tacit practice

of historical knowledge generation. We seek to build on work that unpacks the historical-use implications that unfamiliar archival technologies (including artificial intelligence) have on these traditions, and how new digital practices might be best integrated into this discipline. We focus particularly on our own project's interest; contextualisation within born-digital (email) sources and reflect on the technical and methodological questions our work here has raised. We are particularly interested in how effective iterative search and multiple pathways for discovery can be used to open contextually opaque collections. Such access is likely to leverage a collection's structural and content characteristics, as well as targeted archival selection and categorisation. We ultimately suggest that, by combining relatively open user-led interfaces with machine-led alternatives, digital archives can provide environments suited to both the translation of traditional historical norms and the integration of more novel opportunities for discovery.

Biography:

Adam is a Lecturer in Business and Management at De Montfort University. He holds a PhD in Management from Aston Business School, and his thesis used born-digital sources to research Enron's manipulation of California's deregulated energy markets in the early millennium. For the past year, he has been co-investigator on the AHRC-funding project Contextualising email archives, which seeks to make organisational email more accessible for historical research.