InterPARES 2024 SUMMER SCHOOL

June 24-28, 2024
San Benedetto del Tronto (AP)

Palazzo comunale di San Benedetto del Tronto – Town Hall of San Benedetto del Tronto
https://goo.gl/maps/VqC8HP4BHUQ4jPtz7

MINISTERO DELLA CULTURA
Soprintendenza Archivistica e Bibliografica delle Marche
InterPARES TrustRise
unimc UNIVERSITÀ DI MACERATA l'umanesimo che innova
REGIONE MARCHE
CITTÀ DI SAN BENEDETTO DEL TRONTO

Overview

The InterPARES Summer School presents an overview of the findings of 25 years of research about trustworthy digital records creation, maintenance, preservation, use, and access across generations of technology, discussing both theory and best practice for digital records and archives management.

Day 1 – Monday, 24 June

9:00-12:00 - Luciana Duranti, *Authentic Digital Records*
(includes 30 minute break at 10:30)

InterPARES tested the traditional concept of record in the digital environment, determining the characteristics of digital records as well as their necessary and sufficient attributes. This session will discuss the findings of InterPARES with respect to what is a record in the digital environment, how to maintain a record authentic and how to ensure that its authenticity will be verifiable over time.

12:00-13:30 - *Lunch*

13:30-14:45 - Hrvoje Stancic, *Technological Authentication*

The digital era brought new challenges to the archives. The underlying archival theory and concepts are the same for analogue and digital records. However, the technical manifestation of digital records, their variety, speed of creation, volume, and volatility require new computational approaches, especially with regard to record authentication and the concept of original. Digital signatures bring another layer of complexity to the long-term preservation of digital records, because the certificates they rely upon expire much sooner than the records' retention period. Blockchain and distributed ledger technologies (DLT) can support records' integrity, confirm their sequence, enhance non-repudiation, but also help preserve validity of digital signatures. The TrustChain model, resulting from the InterPARES Trust project research, establishes trust among voluntarily interconnected institutions (i.e. network nodes) and enables confirmation of digital signatures' validity and records' integrity even after the expiration of the signing certificates. The use case of the system for authenticating analogue university diplomas by connection to the blockchain will be shown.

14:45-15:15 - *Break*

15:15-16:30 - Victoria Lemieux, *Diplomatic Criticism of Blockchain-based Recordkeeping Solutions*

The discipline of "diplomatics" – originating in the seventeenth century to systematically test the authenticity of medieval documents – has more recently been adapted to the study of digital records and their systems. In establishing the necessary elements for the long-term preservation of authentic records, archival diplomatics provides one possible (and powerful) analytic framework and methodology for analyzing the trustworthiness of records, including those to be found in blockchain and distributed ledgers. Regardless of the type of blockchain and distributed ledger system under examination, each relies upon trust in the ledger and in the records the ledger contains. Yet each type of blockchain and distributed ledger system still has limitations when judged against archival diplomatic standards of records' trustworthiness, which demands the accuracy, reliability, and authenticity of records. By gaining an understanding of the elemental requirements for trust in records (and in record systems), there is hope that the designers of blockchain and distributed ledger systems might continue to improve the evidentiary quality of blockchain records and recordkeeping. To explore this theme, students will engage in hands-on exercises drawing from solutions encountered in her role as founder and co-lead of Blockchain@UBC, and discussed in her book *Searching for Trust*, incorporating "prospective" application of diplomatic concepts in order to promote improved design of blockchain systems proposed for recordkeeping.

Day 2 – Tuesday, 25 June

9:00-10:15 - Muhammad Abdul Mageed, *Natural language processing for archivists*

An introduction to AI models and tools being developed in the UBC NLP lab and their application in archives.

10:15-10:45 - *Break*

10:45-12:00 - Basma Makhlouf Shabou, *Information governance maturity assessment: concepts, tools and recent developments*

12:00-13:30 - *Lunch*

13:30-14:45 - Tracey P. Lauriault, *Data as artifacts and as records*

In this session participants will learn about the relationship between metrology and data, to consider data beyond normalized technological understandings. Examples will be drawn from the InterPARES 2 General Studies and concepts of accuracy, reliability and authenticity in the sciences, and The I Trust AI case studies about digital twins, and smart grids. The session will end with a discussion of whether data are records.

14:45-15:15 - *Break*

15:15-16:30 - Michel Barbeau, *Problem-Solving Using Knowledge Graphs*

Archives contain vast quantities of information. This presentation introduces knowledge graphs to capture the richness of the information in archives and enable advanced data navigation and analysis. **Topics:** Fundamental concepts (entity, attribute, value, and relationship). Construction of knowledge graphs (creating entities and relationships) from structured data (domain-specific knowledge) or unstructured data (text, images). Capturing complexity of data. Exploring knowledge graphs. Query language. Graph analytics. Demonstration. Problem-solving in archives (historical material).

Day 3 - Wednesday, 26 June

9:00 – 12:00 - Erik Borglund, *Records in the Cloud*
(includes 30 minute break at 10:30)

Modern cloud service as we see it today was established by Amazon in 2002, when it was used as an internal service. In 2006, Amazon offered the service outside the company. The cloud as technology and the related service providers have been a challenge for records and archives professionals, and others. The focus of this session is on the findings of the InterPARES fourth phase, "records in the cloud", but also the cloud technology challenges that one can find today will be discussed.

12:00-13:30 – *Lunch*

13:30-14:45 - Corinne Rogers, *Trustworthy Digital Preservation*

What is required to preserve trustworthy digital records over time and technological change? This is a question that InterPARES has addressed from the beginning. The first two InterPARES projects formulated and elaborated the concept of the chain of preservation, providing a basis for judging whether records' preservation is trustworthy. This session will examine the functional and data requirements for digital preservation derived from the chain of preservation model, and under the rubric of Preservation as a Service for Trust (PaaST), a comprehensive model developed in InterPARES Trust (phase 4) for integrating what is required in digital preservation with technical approaches for meeting such requirements.

14:45-15:15 - *Break*

15:15-16:30 - Joe Tennis, *Metadata*

The findings of InterPARES have established Benchmark and Baseline requirements for preserving authentic records in digital systems. These, along with the Chain of Preservation Model, have been used to create the InterPARES Authenticity Metadata (IPAM). This session introduces these findings from InterPARES and discusses their relevance in the emerging technological environment.

Day 4 - Thursday, 27 June

9:00 – 12:00 - Jessica Bushey, *Managing & Preserving Digital Images Collections*
(includes 30 minute break at 10:30)

In this session participants will learn about managing and preserving aggregations of digital images based on InterPARES research into digital image creation and recordkeeping practices. The role of metadata to capture information contributing to the authenticity and reliability of an image will be explored, along with standards and best practices for metadata and image formats for access and preservation. Images held in social media collections and the challenges these platforms present to access and preservation will be also discussed. The session will end with an exploratory discussion about the opportunities and potential obstacles in using Artificial Intelligence to manage and preserve digital image collections.

12:00-13:30 - *Lunch*

13:30 – 14:45 - Pierluigi Feliciati, *Trusted and easy access to records and archives*

This session will present and discuss the main concepts and issues related to Archival Reference and Access resulting from InterPARES. It will focus on access as presented in the main archival conceptual models and standards, the role of records managers and archivists in mediating between authentic records and users in digital environments, and on how the related activities could be successfully managed. One of the topics will be a user-centred approach in the phases of conception and development, considering the organization of user studies to conceive, build and maintain good archival digital services. Part of the educational activity will be discussing and evaluating some archival access services to better focus the notion of “quality” and the primary metrics for its evaluation.

14:45-15:15 *Break*

15:15-16:30 - Jim Suderman, *Privacy and Ethics in the Era of AI*

This session will peek behind the curtain of the sensational AI tools and techniques and their potential application in the archival field. These tools provide incredible possibilities to enhance appraisal, description and access to archival holdings. But, like other tools, their strengths and limitations must be understood in terms of how and why they were developed so as to mitigate the risk of undermining the fundamental values that drive archival work. Similarly, responsibly providing archival records for transformation into data to train AI models requires archivists to consider how that might affect subjects identified in archival records. This session will reflect on ethical principles from both the archival and AI communities and share practical experience gained within an InterPARES Trust AI study, especially with regard to privacy protection.

Day 5 - Friday, 28 June

9:00- 10:15 - Victoria Lemiuex, *Privacy-enhancing Technologies*

With increased concerns about data protection and privacy over the past several years, and concomitant introduction of regulations restricting access to personally information (PI), archivists in many jurisdictions now must undertake ‘sensitivity reviews’ of archival documents to determine if they can make those documents accessible to researchers. Such reviews are onerous, given increasing volume of records, and complex, due to how difficult it can be for

archivists to identify whether records contain personal information (PI) under the provisions of various laws. Despite research into the application of tools and techniques to automate sensitivity reviews, effective solutions remain elusive. Not yet explored as a solution to the challenge of enabling access to archival holdings subject to privacy restrictions is the application of privacy enhancing technologies (PETs) - a class of emerging technologies that rest on the assumption that a body of documents is confidential or private and must remain so. While seemingly being counterintuitive to apply PETs to making archives more accessible, we argue that PETs could provide an opportunity to protect PI in archival holdings whilst still enabling research on those holdings. In this session, students will be exposed to an overview of these technologies based on a scoping review and possible use cases and future research directions.

10:15-10:45 - *Break*

10:45-12:00 - Scott Cameron, *The Role of Paradata in Ensuring Trust in AI Systems*

Even as it promises revolutionary changes in the information and computing fields, artificial intelligence presents new risks to archivists and information professionals. Too often using opaque processes to arrive at their conclusions, AI systems present challenges for archivists accustomed to documenting human activities in contexts which value transparency and accountability. If unable to understand the techniques of AI tools used to manage archival collections, archivists will be unable to evaluate the strengths and limitations of AI technology solutions, and unable to account for their own actions as professionals.
This session will introduce the concept of paradata, presented as an analytical lens which may be used to assess the documentation needs of AI tools within archival contexts. Defined as information recorded and preserved about records' processing with AI tools, this session will illustrate the need for robust paradata documenting AI applications in archives and equip archivists and information professionals with the vocabulary necessary to articulate their needs for transparency and understandability for AI tools.

12:00-13:30 - *Lunch*

13:30 – 14:45 - Peter Sullivan, *AI tools for Audio Visual Archives*

Audiovisual materials can often be challenging for archival institutions to make available to the public. Accessibility laws for instance may require subtitles or audio descriptions be added if material is published online, which often may be laborious to produce by hand. Finding appropriate content may also be challenging as item-level metadata is required in order to index and search through an institution-run webportal. Still due to their unique nature, audiovisual archives may represent some of the most culturally valuable content archives might offer. As part of our ongoing work, we've looked deeply at how AI tools may play a role in addressing these challenges. While not a silver bullet, AI has improved greatly in recent years and in many situations may provide a suitable candidate for generating first-draft transcriptions of audio recordings, identifying the language or dialect of a spoken audio file, as well as improving existing metadata to aid in search engine discovery. We'll look at the ecosystem of tools that are now available, contextualize their proper use, and draw lines around existing gaps in their application.

14:45-15:15 - *Break*

15:15-16:30 - Emanuele Frontoni, Perganet: *Bringing the Past to Life - Leveraging Deep Learning and Appearance-Based Techniques in Archival Science*

In the rapidly evolving field of archival science, the advent of sophisticated computational methods, particularly those rooted in artificial intelligence, has opened new vistas for historians, archivists, and researchers. This speech, presented at the summer school, delves into the groundbreaking project "Perganet", which exemplifies the integration of deep learning and appearance-based techniques in archival science to revitalize historical documents and artifacts. The lecture begins by laying a foundational understanding of the challenges in archival science, focusing on the degradation of historical documents and the consequential loss of valuable information. It then introduces the audience to the core concepts of deep learning, emphasizing convolutional neural networks (CNNs) and their adeptness at handling visual data. This is followed by an exploration of appearance-based techniques, elucidating how these methods aid in the analysis and reconstruction of the physical and textual attributes of aged documents.

A central portion of the speech is dedicated to the 'Perganet' project. Here, we dissect the methodology employed, illustrating how deep learning algorithms are trained on vast datasets of historical documents to recognize and interpret patterns, damages, and ink fading. The presentation highlights the project's success in restoring and digitizing ancient manuscripts, making them more accessible and legible for research and public viewing. The lecture concludes by discussing the broader implications of such technologies in archival science. It posits that the integration of AI, particularly deep learning and appearance-based methods, is not merely a technical advancement but a transformative approach that enriches our understanding of history and cultural heritage. The potential for these technologies to democratize access to historical knowledge and foster a deeper connection with our past is underscored.

Faculty bios

Muhammad Abdul-Mageed - University of British Columbia
Dr Abdul-Mageed is a Canada Research Chair in Natural Language Processing and Machine Learning, and Associate Professor in the School of Information and Department of Linguistics (Joint Appointment), and Computer Science (Associate Member), at The University of British Columbia. His research is in deep learning and natural language processing, focusing on deep representation learning and natural language socio-pragmatics, with a goal to innovate more equitable, efficient, and 'social' machines for improved human health, safer social networking, and reduced information overload. Applications of his work currently span a wide range of speech and language understanding and generation tasks. He is director of the UBC Deep Learning & NLP Group, co-director of the SSHRC-funded I Trust Artificial Intelligence, and co-lead the SSHRC Ensuring Full Literacy Partnership Grant. He is a founding member of the Center for Artificial Intelligence Decision making and Action and a member of the Institute for Computing, Information, and Cognitive Systems.

Michel Barbeau - Carleton University
Michel Barbeau is a professor of Computer Science. He got a Bachelor's (Universite de Sherbrooke, Canada '85), a master's and a Ph.D. in Computer Science (Universite de Montreal, Canada '87 & '91). From '91 to '99, he was a professor at Universite de Sherbrooke. During the '98-'99 academic year, he was a visiting researcher at the University of Aizu, Japan. Since 2000, he has worked at Carleton University, School of Computer Science, Canada. His main topic of expertise is computer networks: architecture and protocols. Specific research interests include underwater communications and networks, flying drone networks, quantum networks and network control systems. Michel Barbeau is the director of the School of Computer Science at Carleton University.

Erik Borglund – Mid-Sweden University
Dr. Erik Borglund is a professor in archives and information science from Mid Sweden University, campus Sundsvall. Erik has been involved in InterPARES since 2012, as well as in the Digital Records Forensic Project (UBC). Erik's main research focus is on current recordkeeping in the crisis management domain.

Jessica Bushey - San José State University
Dr. Jessica Bushey is an Assistant Professor in the School of Information at San José State University in California, where she teaches courses on Reference and Information Services in Archives, and Preservation Management in Archival Repositories. Prior to joining SJSU, Bushey worked with municipal archives, university museums and archives, and international organizations to develop policies and procedures for managing and preserving digital images and audiovisual collections. Most recently, she led a rapid response social media collecting project and a digital oral history project at the Museum and Archives of North Vancouver (MONOVA) to document community responses to the COVID-19 pandemic.

Scott Cameron - University of British Columbia
As a graduate researcher with the InterPARES project, Scott Cameron has contributed to several research projects related to the themes of paradata, artificial intelligence, and accountability. He is a librarian and archivist and holds a master's degree in history.

Luciana Duranti - University of British Columbia
Dr. Luciana Duranti is since 1987 a Professor of archival theory, diplomatics, and digital records in the master's and doctoral archival programs of the School of Information of the University of British Columbia (UBC), in Vancouver, Canada, and, since 2011, Affiliate Professor at the University of Washington at Seattle, United States. Professor Duranti is Director of the UBC Centre for the International Study of Contemporary Records and Archives, and, since 1998, the Principal Investigator of the InterPARES research project. She has published extensively on archival and diplomatics theory and on the use of their concepts for understanding the products of new technologies. Since 2015, she is the Chair of the Canadian Government Standards Board committee for Electronic Records as Documentary Evidence.

Pierluigi Feliciati – University of Macerata
Archivist in the Italian National Archives from 1986 to 2007, Pierluigi Feliciati coordinated the Information Systems of the State Archives and their first Web portal. He is an associate professor of Archival and Information Science at the University of Macerata, where is the pro-rector for digital archives. In 2019 he was a visiting Professor at the Information School of the University of British Columbia (Canada), winning the Dodson Visiting Scholarship. He leads the UniMC research group of InterPARES Trust AI, is a member of the Research Steering Committee and coordinates two Reference and Access studies. He is the co-editor of the JLIS.it journal.

Emanuele Frontoni – University of Macerata
Dr. Emanuele Frontoni is Full Professor of computer science at the University of Macerata and the Co-Director of the VRAI Vision Robotics & Artificial Intelligence Lab. His research interests include computer vision and artificial intelligence, with applications in robotics, video analysis, human behavior analysis, extended reality and digital humanities. He is the author of over 250 international articles and collaborates with numerous national and international companies in technology transfer and innovation activities. He is also involved in several national and international technology transfer projects in the fields of AI, Deep Learning, data interoperability, cloud-based technologies, and big multimedia data analysis, extended reality and digital humanities. He served as expert for the EU Commission in the AI H2020 and Horizon Europe Calls and he is currently co-speaker of the European IPEI CIS (Important Project of Common European Interest - Cloud Infrastructure and Services) for the AI services of the next generation of European cloud – edge services.

Tracey P. Lauriault - Carleton University
Dr Tracey P. Lauriault is Associate Professor, Critical Media and Big Data, School of Journalism and Communications, Faculty of Public Affairs, Cross Appointed to Digital Humanities, and board member of the Institute for Data Science at Carleton University in Ottawa, Ontario, Canada. She is one of the founders of the field critical data studies, part of the group that spearheaded open data in Canada, and she developed the concept of Open Smart Cities. Ongoing research involves data preservation, digital twins, smart cities, data brokers and indigenous data governance. As a publicly engaged scholar, she mobilizes her research into data and technology policy across sectors. As a data and technological citizen, she examines large and small data and technology systems with the hope of making them more just, inclusive, equitable and environmentally sustainable. https://orcid.org/0000-0003-1847-2738

Victoria Lemieux - University of British Columbia
Victoria is an experienced and multidimensional academic, technology leader, and innovator. She currently holds a position as Professor of Archival Science at the University of British Columbia's School of Information, an affiliated faculty position in UBC's Department of Electrical and Computer Engineering, Faculty of Applied Science, and is a member of UBC's Institute for Computing, Information and Cognitive Systems. She is also founder and co-lead of Blockchain@UBC, the University of British Columbia's multidisciplinary blockchain research cluster. Her academic research focuses on risk to the availability of trustworthy records and how these risks impact upon transparency, financial stability, public accountability, and human rights. Victoria currently sits on the International Standards Organization's Technical Committee 307 (blockchains and distributed ledgers). She holds a doctorate from University College London (Archival Studies, 2002) and, since 2005, has been a Certified Information Systems Security Professional (CISSP). She is the winner of the 2015 Emmett Leahy Award for outstanding contributions to the field of records management, a 2015 World Bank Big Data Innovation Award and 2020 Blockchain Ecosystem Leadership Award. In 2020 she also was named one of Canada's Top 20 women in IT Security, and has twice been named a Sauder School of Business Distinguished Scholar. She is the author and editor of award-winning articles and books, including Financial Analysis and Risk Management: Data Governance, Analytics and Life Cycle Management (Springer, 2012); Building Trust in Information – Perspectives on the Frontiers of Provenance (Springer, 2016), Building Decentralized Trust – Multidisciplinary Perspectives on the Design of Blockchains and Distributed Ledgers (Springer, 2020), and Searching for Trust: Blockchain Technology in an Age of Disinformation (Cambridge University Press, 2022).

Basma Makhlouf Shabou - University of Applied Sciences and Arts Western Switzerland
Dr. Basma Makhlouf Shabou is professor, head of the archival science field at University of Applied Sciences and Arts Western Switzerland in Geneva School of Business Administration (HEG HESSO), where she is also leading the Master programme of Information sciences. She holds a PhD from the University of Montreal (EBSI-UdeM), as well as a

postgraduate degree in RM and a Bachelor's degree in Social Studies. She developed the national program of public records management of the National Archives of Tunisia. She has contributed to the teaching, design and/or revision of various archival programmes in different countries (University of Montreal; Sorbonne University; University of Mannouba; University of A'Sharqiyah; University of British Columbia; University of Liverpool; University of Lausanne; university of Geneva; Mid-Sweden University; University of Anger; University of Bern). Her research focuses on archival appraisal, defining and measuring data quality, information governance, information risk assessment and research data. She co-leads the DLCM 2 project, has presided over the OLOS Association since its creation in 2021, and is active in various expert groups (GREGI, Gira, various ICA bodies, etc.). She created the archival laboratory, ArchiLab, in Geneva.

Corinne Rogers - University of British Columbia
Dr. Corinne Rogers is the Project Coordinator for InterPARES Trust AI (UBC, 2021-2026), and previously InterPARES Trust (UBC, 2012-2019). She is an adjunct professor in the Information School at the University of British Columbia (diplomatics, digital records forensics, and digital preservation). She is Co-Convenor of the Working Group on Electronic Records as Documentary Evidence, Canadian General Standards Board. From 2018-2021 she was a Systems Archivist at Artefactual Systems, lead developers and organizational home to open source projects for digital preservation, AccessToMemory (AtoM) and Archivematica.

Hrvoje Stančić – University of Zagreb
Dr. Hrvoje Stančić is Vice-dean for organization and development, and full professor at the Faculty of Humanities and Social Sciences, University of Zagreb, where he teaches in the Department of Information and Communication Sciences at undergraduate, graduate and postgraduate levels. He has been Chair of archival and documentation sciences at the same Department since 2008. In the context of the 4th InterPARES project (2013-2019) he was Director of the European research team where he led a blockchain-related research study. At the Croatian Standards Institute, he is President of the mirror technical committee for development of ISO/TC 307 Blockchain and Distributed Ledger Technologies standard. In October 2021 he was awarded a bronze medal at the 19th International Innovation Exhibition for his innovation *TrustChain – A System for Preservation of Trustworthiness of the Digitally Signed Documents*. In 2022 he was awarded a silver medal at the 19th International Innovation Exhibition for his innovation *Blockchain-based diploma authentication system.*

Jim Suderman - City of Toronto
Jim Suderman recently retired from the position of Director of Information Access at the City of Toronto, where he oversaw the operations of the City's records management, archives, and information and privacy protection programs. Prior to that he was a Senior Archivist and the Coordinator of the Electronic Records Program at the Archives of Ontario. He has been a researcher in InterPARES 2 and 4, and is currently a co-investigator in InterPARES Trust AI (5).

Peter Sullivan - University of British Columbia
A doctoral student in the PhD program University of British Columbia School of Information, Peter has contributed in a substantial way to the writing of the research proposal for I Trust AI, and to the delivery of tutorials and workshops on AI for records. He provides support to various studies as needed. His doctoral research is on AI for archives.

Joseph T. Tennis - University of Washington
Dr. Joseph T. Tennis is a Professor of Information Science, Adjunct Professor in Linguistics, and a member of the Textual Studies, Computational Linguistics, and Museology faculty advisory groups at the University of Washington. He is also outgoing Associate Dean for Faculty Affairs, and Executive Director of Administrative Services at the University of Washington Information School. He served as President of the International Society for Knowledge Organization from 2014-2018. He was Chair of the Governing Board for the Dublin Core Metadata Initiative, where he has also served on the Usage Board, continuously since 2006. He has served on the *Library Quarterly* and *Knowledge Organization* editorial boards. He has been a member of InterPARES 2, 3, and 4 and is currently a co-investigator in InterPARES Trust AI. His research is on classification theory, information provenance, metadata versioning, ethics of knowledge organization work, descriptive informatics, and authenticity. He teaches courses in classification, metadata, and intellectual foundations of information science.