InterPARES 2023
SUMMER SCHOOL

July 7-11, 2023
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 MiC

Soprintendenza Archivistica e Bibliografica delle Marche

InterPARES TrustAI

uniMC UNIVERSITÀ DI MACERATA
l'umanesimo che innova

REGIONE MARCHE

CITTA' DI SAN BENEDETTO DEL TRONTO
PROVINCIA DI ASCOLI PICENO
FEE Italia

Day 1 – Friday 7 July

Artificial Intelligence to support the ongoing availability and accessibility of trustworthy public records - an overview of InterPARES Trust

9:00 - 16:00 Specialists from the fields of artificial intelligence and archival science will present current studies and discuss challenges and opportunities for bringing AI into all archival functions. This event is open to the public.

9:00   Gather and welcome

9:30   Luciana Duranti & Muhammad Abdul-Mageed, University of British Columbia
        The Whys and Hows in "I Trust AI": Objectives, methods, expected outcomes

10:00   Jason R. Baron, University of Maryland
        AI and Freedom of Information Laws: Using AI To Filter For Exempt Materials in Documents

10:30   Jim Suderman, City of Toronto (ret.)
        Clear values, murky responsibilities: considering the ethical pipeline of archival information and AI implementation

11:00   Umi Mohktar, Universiti Kebangsaan Malaysia
        AI for Records Classification

11:30   Pat Franks, San Jose State University
        Paradata: Documenting the AI Process for Transparency and Accountability

12:00   Peter Sullivan, University of British Columbia
        Applying AI tools in Archival Functions

12:30   Lunch

14:00   Hugolin Bergier, Regis University
        Intensional Logic in RiC-O

14:30   Emanuele Frontoni & Pierluigi Feliciati
        AI for Multimedia understanding: from computer vision to deep learning

15:00   Sanja Seljan, University of Zagreb
        Data Acquisition and Corpus Creation for Security-Related Domain

15:30   Moises Rockembach
        AI literacy and the future of Records Management and Archives

16:00   Round-table discussion and questions

16:30   Adjourn

Day 2 – Saturday 8 July

9:00-9:30 Luciana Duranti – An overview of the InterPARES research project and its products

9:30-10:30 Luciana Duranti – The Concept of Digital Record

InterPARES tested the traditional concept of record in the digital environment, determining the characteristics of digital records as well as the necessary and sufficient attributes of a record that must be captured and preserved to ensure that its nature remains intact. This session will discuss the findings of InterPARES with respect to what is a record in the digital environment.

10:30-10:45 *Break*

10:45-12:00 Corinne Rogers – Authenticity and Authentication

According to archival science, record authenticity is determined on the basis of provenance and documentary context, and is linked to the immutability of documentary evidence affixed to a medium. However, in the digital environment, content, structure and form are no longer inextricably linked, and a presumption of authenticity is made by investigation of the record's elements of identity and integrity. This session will present the findings of InterPARES on authenticity and authentication in the digital environment, and the ontology of trustworthiness to which authenticity belongs.

12:00-13:30 *Lunch*

13:30-16:30 Hrvoje Stancic – Technological Authentication
(includes 15 minute break at 14:45)

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. The digital era brought new challenges in authentication of records and the concept of (one) original. The digital signature further complicated long-term preservation of digital records, because their certification expires much sooner than the records' retention period. The blockchain and distributed ledger technologies (DLT) can support records' integrity, confirm their sequence, enhance non-repudiation, but also help preserve digitally signed records. 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.

Day 3 Sunday 9 July

9-12:00 Basma Makhlouf Shabou – Information governance maturity assessment : concepts, tools and recent developments
(includes 15 minute break at 10:30)

12:00-13:30 *Lunch*

13:30-16:30 Tracey P. Lauriault – Data as artifacts and as records
(includes 15 minute break at 14:45)

In this session participants will learn about the relationship between metrology and data, to think of data beyond normalized technological understandings, about different kinds of data and their ontological characteristics, including data in administrations, as national and historical artifacts and about data infrastructures. The session will frame concepts within contemporary systems such as smart cities, digital twins, and how data are often the root cause of many AI ethical and social justice issues. The session will end with a discussion of whether data are cultural and social artifacts, simply the outputs of scientific or administrative endeavours, and whether they are records.

Day 4 Monday 10 July

9:00 – 12:00 Erik Borglund – Records in the Cloud
(includes 15 minute break at 10:30)

Modern cloud service as we see it today was established by Amazon in 2002 where they used it as an internal service. In 2006, they offered the service outside of the company. The cloud as technology and the cloud where you find service providers have been a challenge for archival practitioners, records managers, and others. In this lecture the focus is upon the findings from the InterPARES "records in the cloud" project, but also a focus will be upon cloud technology challenges that you find today in 2023.

12:00-13:30 *Lunch*

1:30 - 16:30 Ken Thibodeau – Proteus Bound: Trustworthy Digital Preservation
(includes 15 minute break at 14:45)

Trustworthy digital preservation looks like an oxymoron, a contradiction. How can you keep something intact when it depends entirely on a network of loosely connected and very diverse things that change at different rates and in unpredictable ways? 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, which is not a way to preserve digital records, but a way of documenting how they have been preserved. It provides a basis for judging whether their preservation is trustworthy. InterPARES also looked more deeply into the problem, examining digital records in interactive, dynamic and experiential systems; that is, electronic records that have no precedents in hard copy. This research lead to a deeper understanding of what it means to preserve digital records. These different threads came together in the articulation of functional and data requirements for digital preservation under the rubric of Preservation as a Service for Trust (PaaST). PaaST provides a comprehensive model for integrating what is required in digital preservation with technical approaches for meeting such requirements. This session will review InterPARES research on digital preservation, its products and its impacts.

Day 5 Tuesday 11 July

9:00- 10:30 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.

10:30-10:45 *Break*

10:45 – 12:00 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. The presentation will include the first results of the InterPARES Trust AI study on "Users approaches and behaviors in accessing records and archives in the perspective of AI: a global user study". 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.

12:00-13:30 *Lunch*

1::30 – 4:30 Jessica Bushey – Managing & Preserving Digital Images Collections
(includes 15 minute break at 14:45)

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.

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](#).

Hugolin Bergier - Regis University

Dr. Hugolin Bergier is an Assistant Professor in Computer Science at Regis University. Prior to joining Regis, he was an Assistant Professor at the Catholic University of the West in France and worked as a developer and research scientist for Phase Change Software, a company which specializes in artificial intelligence for program analysis. He still does consulting there as part of the R&D department. His work at Phase Change includes research work on intellectual property, patenting and the development of a prototype in Prolog and Haskell. Dr. Bergier has a PhD from Sorbonne University (Paris). He is from Western France, where his family still makes Cognac. His research interests revolve around logic, mathematics and theoretical computer science with focuses on lambda-calculus, combinatory logic, intensionality and computation theory.

Jason R. Baron - University of Maryland

Jason R. Baron is a Professor of the Practice in the College of Information Studies at the University of Maryland, with current research interests involving the capture of new forms of electronic communications used in the government, and the use of AI in improving access to public records. He previously served as the first Director of Litigation at the US National Archives and Records Administration, and before that as a trial attorney and senior counsel at the US Department of Justice, where he represented the Archivist of the US in litigation involving the preservation of White House electronic records. He was a co-founder of the National Institute of Standards and Technology' TREC Legal Track, and is a past winner of the Emmett Leahy Award.

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.

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, Dr. Pierluigi Feliciati coordinated the Information Systems of the State Archives and the Web portal of the Italian Archives. He is an associate professor of Archival and Information Science at the University of Macerata, where he is the pro-rector for digital archives. In 2019, winter term II, he was visiting Professor at the information School of the University of British Columbia (Canada), winning the Dodson Visiting Scholarship. Since 2021 he leads the UniMC researchers of InterPARES Trust AI, member of the Research Steering Committee and coordinates the RA05 study on users' attitudes against archival digital services and AI tools. He is the co-editor of the JLIS.it journal, managing director of the "Capitale culturale" journal.

Patricia Franks - San Jose State University
Dr. Franks, Professor Emerita at San José State University, teaches courses in Enterprise Content Management and Digital Preservation. She is a Certified Archivist, Certified Records Manager, Information Governance Professional, and Certified Information Governance Officer, as well as a member of ARMA International's Company of Fellows. Franks is a member of the *CIGO Association Board of Director*s, a member of Preservica's *Digital Preservation Sustainability Council*, and the immediate past president of the National Association of Government Archivists and Records Administrators (NAGARA). Her research interests lie in emerging technologies, including artificial intelligence and blockchain distributed ledger technology, and their impact on records management and information governance.

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.

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 Department 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.

Umi Mokhtar - Universiti Kebangsaan Malaysia
Dr. Umi Asma' Mokhtar is a senior lecturer in information science at Universiti Kebangsaan Malaysia, Faculty of Information Science and Technology. Her areas of interest in research include electronic records management, function-based classification, information policy, and information security. Her articles have appeared in international and national periodicals, such as the International Journal of Information Management and the Records Management Journal. She is currently the Malaysian Team's lead researcher for the InterPARES Trust AI project.

Moises Rockembach - Federal University of Rio Grande do Sul
Dr. Moises Rockembach is Professor of Archival Science and Information Science at the Faculty of Librarianship and Communication at the Federal University of Rio Grande do Sul (Brazil). His expertise involves archives and records, digital transformation, digital preservation, digital ethics, platform studies, information science and digital humanities. He is research leader of the Research Group in Digital Preservation (Federal University of Rio Grande do Sul / Brazil). He is currently a Visiting Scholar at KU Leuven (Belgium), engaged in activities of the Mintlab (Meaningful Interactions Lab) and the KU Leuven Digital Society Institute.

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.

Sanja Seljan, University of Zagreb

Dr. Sanja Seljan is a full professor-permanent tenure, researcher, consultant, and Ph.D. supervisor in the Faculty of Humanities and Social Sciences, University of Zagreb. Her research interests are related to AI development: big data and information extraction, Natural Language Processing (NLP), machine translation and CAT tools, chatbots, data analysis and visualization, business analytics, machine learning and eHealth. Currently, she is working in the domain of health, finance, law, pharmaceutics, visual identity and national branding.

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 Professor, Associate Dean for Faculty Affairs, and Executive Director of Administrative Services at the University of Washington Information School, 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 on the Usage Board of the Dublin Core Metadata Initiative and served on the Governing Board until 2021. He has been a member of InterPARES 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.

Kenneth Thibodeau - Regis University

An internationally recognized expert in management of digital information, Dr. Thibodeau has served as Chief of the Records Management Branch of the US National Institutes of Health, Director the Electronic Records Archives and the Center for Advanced Systems and Technology at the National Archives and Records Administration, Director of the Department of Defense Records Management Task Force and Senior Guest Scientist at the National Institute of Standards and Technology in the US. A Fellow of the Society of American Archivists and winner of the Emmett Leahy Award, he has been a researcher in all five InterPARES projects.