The High-Performance Computing Chair

Transforming Knowledge Through Advanced Computing

Book of Abstracts

HPC CHAIR
High Performance Computing

University of Évora
26-28 May 2025

EURO²

International Meeting: "Transforming Knowledge Through Advanced Computing"

Meeting Website: https://indico.hpc.uevora.pt/event/126/

Organization: High Performance Computing Chair (https://catedrahpc.uevora.pt)

Scientific Organizing Committee

Carla Ferreira de Castro, University of Évora, Portugal (*)
Estelina Lora da Silva, University of Porto, Portugal (*)
Francisco Coelho, University of Évora, Portugal (*)
Domingos Barbosa, ENGAGE SKA & University of ÉvoraPortugal (*)
Ivo Santos, University of Évora, Portugal (*)
Jaime Serra, University of Évora, Portugal (*)
Miguel Avillez, University of Évora, Portugal, & Technical University Berlin, Germany (*) (**)
Rui Barroso, DECSIS, Portugal (*)
Salvador Pinto Abreu, University of Évora, Portugal
Sandra Boto, University of Évora, Portugal (*)

(*)The HPC Chair is a R&D infrastructure (based at the University of Évora), endorsed by Hewlett Packard Enterprise (HPE), and involving a consortium of higher education institutions (6), research centres (3), enterprises (11), a CoLAB, and a public organisation. The HPC Chair has a research team of 24 scientists (22 Ph.D. holders and two graduate students).

Contents

Transforming Knowledge Through Advanced Computing
26 May 2025
(11:30-19:00)

Digital Humanities - Innovation with Advanced Computing
28 May 2025
(09:00-17:00)

Session 1: Digital Innovation in Tourism

Session 2: Computational Models in Heritage

Session 3: Computational Models for Text Recognition and Reuse

Meeting Closure & Farewell

Transforming Knowledge Through Advanced Computing

26 May 2025

(11:30-19:00)

Session 1: Advanced Computing: HPC, AI & QC

Overview of HPC technologies and related fields (KT)

Pedro Alberto

University of Coimbra, Coimbra, Portugal

Email: pedro.alberto@uc.pt

The presentation will give a brief overview of current technologies used in HPC and their connection with related fields as AI and Quantum Computing. A reference to current initiatives in Europe addressing these matters will be also given.

Digital Twins, AI and the Future of Science (KT)

Mark Asch

Université de Picardie Jules Verne, Amiens, France

Email: mark.asch@u-picardie.fr

In this general talk, we will evaluate the growing influence of digital twins and AI on science and on the scientific approach. The principal proposal relies on a suitable *blending* of classical model-based science with data-driven machine learning. This is now possible thanks to three factors: data volumes, open-source software and affordable high-performance computing hardware, in particular GPUs (Graphics Processing Units). Finally, the recent "invasion" of LLMs (Large Language Models) represents an enormous challenge to the scientific world as well as to educators—this phenomenon requires our particular attention, and we will survey some initiatives and tendencies.

The Revolution of Quantum Technologies (KT)

Yasser Omar

IUniversity of Lisbon & Portuguese Quantum Institute, Lisbon, Portugal

Email: yasser.omar@tecnico.ulisboa.pt

In this talk, I will introduce Quantum Information Technologies in a non-technical way. I will emphasise quantum computation, review its state of the art, and discuss its impact in advanced computation.

Session 2: Processing & Analytics of Massive Volumes of Data

Beyond Silos: Building the HPC-Cloud Interoperability Bridge (KT)

Marco Puccini

CINECA, Bologna, Italy

Email: m.puccini@cineca.it

In the last few years, even thanks to AI breakthrough in computing world, a growing attention to the ever more sophisticated cloud technologies is raised in the HPC community. If the HPC systems give high solidity for massive processing loads thanks to the bare metal resources access, the flexibility and scalability of Cloud technologies comes up as useful for specific usage, expanding sometimes the potential of data. Cloud could also be very useful for deploying dedicated services (i.e. inference and/or MCP servers for AI activities) leveraging scalability, reliability and facilitating multiple remote accesses. Even if they are both based on the same bare metal hardware, HPC and Cloud are bearers of very different computing paradigms, in terms not only of resource management but even for data handling, services exposure, interoperability. Let's consider the difference between files and objects as organization and representation of data for instance. Meeting the challenge of hybrid computing is strategic for the near future and CINECA is putting a strong effort into building a bridge between those different, distant yet necessarily complementary worlds. This work aims to give an overview of such efforts from a Cloud engineer perspective.

ALICE data acquisition and processing in LHC Run 3 and Run 4 (IT)

Vasco Barroso

CERN, Geneva, Switzerland

Email: vmcb@cern.ch

ALICE (A Large Ion Collider Experiment) has undertaken a major upgrade during the LHC Long Shutdown 2 (2019-2021) with the goal of operating at a peak Pb-Pb collision rate of 50 kHz to explore rare physics processes for which large statistics are needed. To address the new requirements, a brand-new online computing system has been developed, commissioned and put into production. This talk will describe the newly implemented ALICE online data acquisition and processing system, specifically designed to handle the substantial raw data rate of ~3.5 TB/s generated by the detectors. It will describe the data flow from detectors front-end-electronics to persistent storage, the data processing stages with the extensive use of FPGAs and GPUs, the data storage and the operational experience during 2023 and 2024.

Advanced computing for reproducibility of astronomy Big Data science with a showcase of AMiGA and the SKA Science prototype (IT)

Julián Garrido

Instituto de Astrofísica de Andalucía (IAA - CSIC), Granada, Spain

Email: jgarrido@iaa.csic.es

In this talk, I will explain some of the key challenges related to workflow execution faced by the SKA Regional Centre Network. I will highlight how linked data technologies and metadata can contribute to reproducibility, and how these elements can serve as a foundation for developing smarter and more efficient Science Platforms. Additionally, I will present some of our past developments to tackle the reproducibility challenge within the SKA context, how they have impacted in the Spanish SRC, and how we have been applying various strategies to selected science astronomy studies using existing technologies.

Session 3: Advanced Computing in the Enterprise Environment

Review of current challenges in HPC & AI infrastructure and the working in progress HPE is doing for the HPC, AI and Quantum convergence (KT)

Gerardo Ares

Hewlett Packard Enterprise, Madrid, Spain

Email: gerardo.ares@hpe.com

In recent years, artificial intelligence has revolutionized the market due to its great potential to aid in production processes. The technologies used to develop it are emerging at a dizzying speed compared to what was previously known, and with this, they have brought challenges in cooling architectures and technologies. In this talk, we will discuss these new challenges, how HPE addresses them, and, in turn, present HPE's approach to the convergence of HPC, AI, and quantum computing.

It's not all good news: Quantum computing's threats to cybersecurity... and solutions (IT)

Rui Semide

N10GLED, Lda., Évora, Portugal

Email: rui.semide@entangled-space.com

Conventional e-mail security relies on a TLS handshake in which asymmetric algorithms (RSA or ECDHE) generate symmetric AES session keys. Because Shor's algorithm will render these public-key primitives tractable on a large-scale quantum computer, any TLS traffic captured today is vulnerable to retrospective decryption. Hybrid post-quantum TLS suites are being prototyped, yet recent cryptanalytic results—full breaks of the NIST PQC finalists Rainbow (signatures) and SIKE (key exchange) in 2022—show that newly standardized mathematic algorithms can fail abruptly. We therefore present Quantum Key Distribution (QKD) as a physics-based alternative. Using single-photon polarization states, the BB84 protocol derives shared keys whose confidentiality follows from the no-cloning theorem and Heisenberg uncertainty, rendering any eavesdropping detectable regardless of computational power. The presentation concludes with N10GLED's ongoing efforts to deliver Quantum Keys-as-a-Service (QKaS): a space segment featuring a weak coherent pulse CubeSat demonstrator and a ground segment deploying transportable agnostic quantum station feeding a key-management platform. Together, these initiatives aim to provide operational, physics-anchored key material for critical communications well ahead of large-scale quantum adversaries.

Fusion of Quantum Chemistry, AI, and HPC: A New Dawn in Computational Toxicology and Cheminformatics (IT)

Pedro Lopes, Beatriz Costa, & Ana Sofia Almeida

FastCompChem, Lda., Covilhã, Portugal

Email: pemlopes@fastcompchem.pt

FastCompChem is the scientific lead of the EIC Pathfinder project *QUANTUM-TOX*, which pioneers a novel approach to computational toxicology through the development of ESigns-a new class of molecular fingerprints. ESigns integrate quantum chemical descriptors of chemical reactivity and intermolecular interactions with artificial intelligence to address key limitations in current cheminformatics methods, including inadequate chemical space coverage and the opacity and complexity of traditional descriptors. This talk will present the conceptual framework behind ESigns, highlighting how quantum chemistry, artificial intelligence and high-performance computing (HPC) jointly enable chemically accurate yet computationally feasible predictions. We will detail strategies used to reduce computational costs while maintaining chemical accuracy and highlight broader applications of the approach beyond toxicology, signaling a shift toward more robust and interpretable predictive models in cheminformatics.

Solving Complex Scientific Problems (Discussion) & Meeting Summary

Advanced Computing and the Solution of Complex Scientific Problems (Discussion)

Roman Chertovskih[1] (Chair), Estelina Lora da Silva[1], António Canto[2], Alfredo Carvalho[2], Domingos Barbosa[2,3], & Miguel de Avillez[2,4]

[1] *University of Porto, Porto, Portugal*

[2] *University of Évora, Évora, Portugal*

[3] *ENGAGE SKA, Portugal*

[4] *Technical University Berlin, Berlin, Germany*

Email: roman@fe.up.pt

Rapid advancements in computational power, algorithms, and data storage are fundamentally reshaping how we acquire, process, analyze, and disseminate information. Key areas of focus include the role of high-performance computing (HPC) in accelerating scientific discovery and complex simulations, the paradigm shift introduced by artificial intelligence (AI) and machine learning (ML) in extracting insights from vast datasets, and the emergence of quantum computing with its potential to tackle problems currently intractable for classical systems. This transformation is evident in fields ranging from molecular chemistry, materials science, astrophysics modelling, geophysics, digital humanities, where advanced computing enables unprecedented levels of precision, prediction, and innovation. The discussion also touches upon the challenges and opportunities associated with this evolution, including ethical considerations, data governance, and the need for interdisciplinary collaboration to fully harness the transformative potential of these technologies. Ultimately, advanced computing is not merely augmenting human intellect but fundamentally redefining the very nature of knowledge creation and its application in solving global challenges.

Meeting Summary (KT)

Paulo Quaresma

University of Évora, Évora, Portugal

Email: pq@uevora.pt

A summary of the presentations and discussions during the meeting sessions ("Advanced Computing-HPC, AI, and QC", "Processing and Analytics of Massive Volumes of Data", and "Advanced Computing in the Enterprise Environment") and the final discussion on the use of "Advanced Computing in Solving complex Scientific Problems" will be presented looking into the different aspects under discussion and their importance in key areas of the society as a whole.

Square Kilometre Array - Cutting Edge Astronomical Facilities

27 May 2025

(14:00-18:30)

Session 1: SKA Science Regional Centres

ENGAGE SKA: from capacity building to a PT Science Regional Centre prototype (OT)

Domingos Barbosa

ENGAGE SKA & University of Évora, Portugal

Email: domingos.barbosa@uevora.pt

ENGAGE SKA is a Research Infrastructure of the National Roadmap for Research Infrastructures of Strategic Interest. IT sets up a capacitation and sustainability plan for a Green e-Science Infrastructure, that will act as a driver for smart and sustainable growth in Portugal taking radioastronomy as an Innovation Open Living Lab. ENGAGE SKA has established an innovation strategy "born from science" anchored in capacity building towards the participation in the construction of SKA and consolidation of national assets for optimal science exploration with an industrial leadership footprint. Its results contributed to the inclusion of a radioastronomy and ICT based roadmap into the main pillars of PT and EU digital strategies, as set in Portuguese Government program InCode2030 (for example) and in the SKA Portuguese White Book (2019, following presence of the Minsier of Economy in the SKA days 2018). ENGAGE SKA encompassed three type of results: the astronomical and fundamental science results, the compute and data science results, the industrial and social impact results including ICT, environmental and farming applications out of the technologies and methods developed. Currently, building on the contributed efforts to the developments of HPC infrastructure at University of Évora, ENGAGE SKA is preparing Portugal's contribution to the SKA Regional Centre Network. I will describe main achievements and future plans.

SRCNet: Designing a Global Data Ecosystem for the SKA Era (IT)

Jesus Salgado

SKA Observatory, Macclesfield, UK

Email: jesus.salgado@skao.int

The Square Kilometre Array (SKA) will produce unprecedented volumes of scientific data, demanding a transformative, distributed approach to data processing, access, and analysis. The SKA Regional Centre Network (SRCNet) is a collaborative international initiative aimed at addressing this challenge. This talk will provide an overview of the current status and evolving design of the SRCNet, highlighting its role as a federated infrastructure that supports science-ready data delivery, user-driven processing, and community engagement across member regions. Key topics will include architectural principles, interoperability frameworks, governance models, and the integration of emerging technologies to ensure scalability, sustainability, and scientific impact. We will also discuss recent progress of SRCNet v0.1 and main versions, as well as the roadmap toward full operational readiness in alignment with SKA Science Verification.

Developing UK SKA Regional Centre (IT)

Louise Chisholm

University College London & UK SKA Regional Centre, United Kingdom

Email: l.chisholm@ucl.ac.uk

The Square Kilometre Array (SKA) is set to revolutionise radio astronomy. The SKA Regional Centre Network (SRCNet) will develop and deploy a collaborative and federated network of SKA Regional Centres, globally distributed across SKA partner countries, to host the SKA science archive and analysis services. The SRCNet will make data storage, processing and collaboration spaces available, while supporting and training the community, to maximise the scientific productivity and impact of the SKA.

The talk will provide an update on how the UKSRC is developing its infrastructure and services both nationally, and as a part of the SRCNet, to maximise the opportunities from the SKAO for the UK astronomy community.

The Spanish SRC initiative: supporting national preparatory activities for SKAO and contributing to the SRCNet (IT)

Susana Sánchez Expósito, Lourdes Verdes-Montenegro, Julán Garrido, Laura Darriba, Javier Moldón, Manuel Parra-Royón, Jesús Sánchez, Maria Angeles Mendoza

Instituto de Astrofísica de Andalucía (IAA - CSIC), Granada, Spain

Email: sse@iaa.es

The Spanish SKA Regional Centre initiative (espSRC) began in 2018 with the main goal of supporting and strengthening scientific activities in Spain related to the SKA Observatory (SKAO) and its precursors. With a strong commitment to Open Science, the espSRC places particular emphasis on addressing the challenges of handling SKAO data to extract scientific knowledge in a reproducible way.

The initiative operates a cloud-based platform that has supported over 90 projects to date. These projects span a wide range of activities, including scientific data analysis, technical developments, and training events primarily focused on radio interferometric data processing and open science tools.

The espSRC has actively contributed to the SKA Regional Centre Network (SRCNet), coordinating one of the international teams established for the SRCNet development since 2022. SRCNet is now entering its implementation phase, with version 0.1 representing its first operational release. In February 2025, the espSRC became the first node to complete the deployment and integration of services required to join SRCNet v0.1.

This talk will provide an overview of the activities undertaken by the espSRC to date, highlighting its technical achievements, contributions to the SRCNet, and commitment to Open Science.

Session 2: Science with the SKA

The SKA Cradlke of Life Science Working group (IT)

Josep Miquel Girart

Institut de Ciències de l'Espai, CSIC, Spain

Email: girart@ieec.cat

In this talk I will briefly review the main topics of the Cradle of Life Science Working group, which are divided in five main sections: (1) Understand Planet formation, (2) Detect and characterize large molecules in star and planet forming regions, (3) Detection and characterization of exoplanets, (4) Search for ExtraTerrestrial Life (SETI) and (5) Understand of the Solar System and its origins. I will also briefly present the VOLS (VLA Orion Large Survey).

Cosmology with HI intensity mapping with the SKAO and precursors (IT)

José Fonseca

Centre of Astrophysics, University of Porto, Portugal

Email: jose.fonseca@astro.up.pt

The future Square Kilometre Array Observatory will perform a wide HI intensity mapping survey over 20,000 sqdeg from z = 0.35 - 3. Here we review intensity mapping with HI, the science that it will enable and challanges faced. We will also discuss the science with the ongoing HI IM survey with the MeerKAT (MeerKLASS) and all it has been learn to do Cosmology with HI IM in practice. These will provide complementary clustering information that have independent systematic uncertainties to those of optical surveys like LSST and Euclid. Such a multitude of synergies can further improve constraints significantly beyond what optical or radio surveys can achieve on their own.

The loudest Radio-Shadows: Blazars and the Assembly of Black Holes in the SKAO-VLBI era (IT)

Sonia Antón

University of Coimbra, Portugal

Email: santon@uc.pt

Blazars are a rare and powerful type of galaxy with a supermassive black hole at the center; they shine brightly across the entire electromagnetic spectrum, especially in radio. Using current SKAO precursor telescopes, we have already begun discovering high-redshift blazars—objects that help us understand not only how black holes grow, but also the early stages of the universe. At the other end of cosmic time, these same instruments are helping identify nearby systems with multiple supermassive black holes orbiting each other, offering clues to how galaxies and black holes evolve through mergers. These studies are just the beginning: with the full SKAO, and especially when combined with Very Long Baseline Interferometry (VLBI), we will be able to explore these extreme systems in far greater detail—mapping black hole growth across the universe with unprecedented clarity.

Software development and guidelines for the SKA and the SKA Regional Centres (IT)

Dalmiro Maia

Astronomical Observatory Professor Manuel de Barros, University of Porto, Portugal

Email: dmaia@fc.up.pt

The Square Kilometre Array (SKA) project aims to build the world's largest radio telescope, generating unprecedented volumes of scientific data. The SKA Regional Centres (SRC) are a collaborative and federated network, globally distributed across SKA partner countries, to host the SKA Science Archive. The SRC Network will make data storage, processing and collaboration spaces available, while supporting and training the community, to maximise the scientific productivity and impact of the SKA by:

- developing a scalable, prototype SRC Network that allows authorised users and teams to access and analyse SKA data;
- developing the software, architecture, policies and processes necessary for SRC Network operations;
- growing the prototype SRC Network, as new SRCs become available and expanded or new functionalities are developed, leading towards a fully operational and global Network.

In this talk we will make an overview the Minimum Production Service for the SRCs, the functionalities required and the implementation scenarios. The focus will be on Data Management, Federated Computing requirements and Storage.

Digital Humanities - Innovation with Advanced Computing

28 May 2025

(09:00-17:00)

Session 1: Digital Innovation in Tourism

Decision-Making in a Human-Agentic AI Organizational Socialization context: The case of Tourism (KT)

Nikolaos Stylos

University of Bristol, United Kingdom

Email: n.stylos@bristol.ac.uk

The rapid diffusion of AI technologies across industries has transformed service co-creation and decision-making processes. As AI advances in competence and integration, its impact on daily life and work continues to expand, transforming human-AI interactions. These interactions shape the ways decisions are made, and ultimately, how organizations respond to complex tourism environments. Bridging the dynamic and context-sensitive human ecosystem with the calculative and data-driven AI ecosystem requires meta-learning and co-evolution between human and AI agents (Stylos et al., 2025).
Lindgren (2024) argue that AI's role in decision-making should be viewed beyond discrete tasks—emphasizing AI's evolving social and cognitive interactions within complex infrastructures. Sociotechnical Systems Theory highlights human-technology interdependence in organizations, yet AI's ability to engage in discourse, negotiation, and socialization differentiates it from past technological shifts. Consequently, new theoretical models are needed to examine AI's integration into contemporary decision-making.

The tourism industry, as a complex open system, involves multiple stakeholders influenced by public and private interests, evolving business models, and emerging technologies. Traditionally, tourism research assumes humans as sole decision-makers, with technology playing a supportive role. An integrated human-AI ecosystem can challenge this notion, as decision-making would rather be conceptualized as socially interactive and continuously evolving.

While there are concerns about AI reducing human interaction in tourism, it also holds potential for sustainability and innovation (Majid et al., 2023). As AI agents increasingly engage in social interactions with human counterparts, their influence extends beyond technical functions and into managerial decision-making practices. They become participants in the broader process of knowledge creation, problem-solving, and adaptation within tourism organizations. This talk offers a direction for future research that centers on how these agents influence one another, and how decision-making in a human-agentic AI hybrid system is shaped.

References

Lindgren, H. (2024). Emerging Roles and Relationships Among Humans and Interactive AI Systems. International Journal of Human–Computer Interaction, 1-23, https://doi.org/10.1080/10447318.2024.2435693.
Majid, G., Tussyadiah, I. P., Kim, Y. R., & Pal, A. (2023). Intelligent automation for sustainable tourism: a systematic review. Journal of Sustainable Tourism, 31, 2421–2440.
Stylos, N., Okumus, F., & Onder, I. (2025). Beauty or the Borg: Agentic artificial intelligence organizational socialization in synergistic Hybrid Transformative Dynamic Flows. Tourism Management, 111, 105205.

Optimising resources and enhancing the experience for more sustainable tourism (IT)

Francisco Dias

Polytechnic Institute of Leiria, Portugal

Email: francisco.dias@ipleiria.pt

As humanity realises that the planet's resources are finite and therefore scarce, it is also redirecting the objectives of human development. Today, decision-makers with responsibility in the tourism industry are aware that one of the main imperatives of sustainable development is the transition to new, more sustainable practices, which requires continuous monitoring of the environmental, social and economic impacts of tourism activity, as well as more rational, data-based decision-making.
In this context, and as each of the stakeholders realises that they can contribute individually to the collective effort to make tourism more sustainable, they are faced with a new problem: where is the compass of sustainability? In other words, on what data should day-to-day decisions be based (whether at strategic or operational level) in the absence of an information system on the specific impacts of each action?

To respond to this problem and fill the information gap, a system is being developed (all-in-one platform and mobile app) that allows the environmental and social impacts of specific actions to be assessed (continuously or over time): carbon footprint, water footprint, social responsibility, good practices in energy, water and waste management, etc. This is a multifunctional tool that guides destination and company managers towards Resource Optimisation (RO), and at the same time offers tourists and residents the means to Enhance Experiences (EV), which is why it was named OR-VE. Its implementation is expected to be of obvious benefit to society as a whole and to the planet.

Lisbon What the Tourist Should See: A Geo-Spatial Literary Guide to Fernando Pessoa's Lisbon Using ArcGIS StoryMap (CT)

Rosália Guerreiro

ISCTE - University Institute of Lisbon, Portugal

Email: rosalia.guerreiro@iscte-iul.pt

This paper explores innovation in Digital Humanities through the advanced use of geospatial computing to create an interactive literary map based on the book What the Tourist Should See (1925), the touristic guide to Lisbon authored by Fernando Pessoa. Using ArcGIS StoryMap app, the project reconstructs the early 20th-century urban and architectural landscape described in Pessoa's text, offering an immersive digital experience that bridges literary heritage, historical cartography, and cultural tourism. By integrating literary interpretation with spatial analysis, the story map allows users to follow curated routes through the city, visualizing Pessoa's unique view of Lisbon as both a historical and poetic space. This digital guide demonstrates how advanced GIS tools can enhance the accessibility, interactivity, and educational potential of literary texts, and exemplifies how computing can transform cultural memory into a participatory, dynamic resource.

References

Bodenhamer, D., Corrigan, J., & Harris, T. (2010). The Spatial Humanities: GIS and the Future of Humanities Scholarship. Indiana University Press.

Esri (ArcGIS StoryMaps). Documentation and case studies. https://storymaps.arcgis.com/

Pessoa, F. (1992). Lisbon What the Tourist Should See. Livros Horizonte.

Piatti, B., & Hurni, L. (2011). Mapping Literature: Towards a Geography of Fiction. Cartographica Helvetica.

Data-Driven Monitoring of Regional Sustainable Destinations: A Case Study of the PISTA Digital Platform (IT)

Jaime Serra

University of Évora, Portugal

Email: jserra@uevora.pt

The imperative for sustainable tourism management has catalyzed a pressing need for advanced monitoring systems that offer detailed, data-driven insights into the complex interplay of environmental, social, and economic impacts engendered by tourism within specific regions . Data-driven systems are emerging as pivotal tools in this context, enabling stakeholders to track progress, identify potential challenges, and make informed decisions to foster sustainability within tourism destinations (Reinhold et al., 2023). These systems leverage real-time data acquisition, advanced analytics, and multifaceted visualization techniques to deliver a comprehensive and dynamic understanding of destination performance. The PISTA Digital platform serves as a crucial infrastructure, specifically designed to provide continuous, data-supported evaluations of sustainability initiatives. By offering a practical platform to check indicators, it promotes engagement with sustainable tourism challenges. The architecture of data-driven monitoring systems often incorporates a range of data sources, including surveys; API connection to the public statistics; mobility data that provide a holistic view of destination dynamics. Through the integration of these diverse data streams, decision-makers gain the capacity to assess the effectiveness of sustainable tourism policies, optimize resource allocation, and promote responsible tourism practices. The PISTA Digital platform distinguishes itself through its integrated approach, which emphasizes the roles of public administration, tourism businesses, and the local community in promoting sustainability. The platform's capacity to process and visualize multi source business data is crucial for delivering statistical analysis results to various stakeholders, including travelers and administrators, thereby supporting informed decision-making (Qin & Pan, 2023). The primary objective of this talk is to outline the methodology of the PISTA Digital platform, focusing on its data lake architecture and the process of data-driven collaboration among key stakeholders in the tourism industry. Additionally, a series of case studies will be presented, showcasing the application of sustainable tourism indicators and the available data visualisation options.

References

Qin, Z., & Pan, Y. (2023). *Design of A Smart Tourism Management System through Multisource Data Visualization-Based Knowledge Discovery.* Electronics, 12(3), 642.

Reinhold, S., Beritelli, P., Fyall, A., Choi, H. C., Laesser, C., & Joppe, M. (2023). *State-of-the-Art Review on Destination Marketing and Destination Management.* Tourism and Hospitality, 4(4), 584.

Session 2: Computational Models in Heritage

AI for Heritage: Challenges and Opportunities (IT)

Anna Foka

Department of Archives, Museums and Libraries, Uppsala University, Sweden

Email: anna.foka@abm.uu.se

Artificial Intelligence (AI) is rapidly reshaping the heritage sector, presenting both transformative opportunities and complex challenges for cultural heritage stewardship. This talk critically examines the growing influence of AI for heritage, focusing on how machine learning technologies are used to digitize, catalogue, and to interpret heritage collections. While AI enhances access, engagement, and research possibilities, it also risks perpetuating historical biases embedded in data and metadata. These biases, if unaddressed, may be amplified by AI systems, shaping public understanding of the past in problematic ways. Using case studies as example , I argue that effective, responsible use of AI in heritage contexts demands interdisciplinary collaboration among cultural heritage professionals, data scientists, and social scientists. By integrating technical and humanistic perspectives, the sector can identify and mitigate bias, develop more inclusive datasets, and embed ethical considerations into AI applications. Recommendations are offered for future education, research, and practice, emphasizing the need for ongoing monitoring, capacity building, and the development of ethical frameworks. Ultimately, the paper advocates for a holistic approach to AI for heritage ensuring that technological advancements may be harnessed to contribute to a more inclusive, reflective, and equitable engagement with cultural heritage.

Extracting Information from Grey Literature in Portuguese Archaeology: Project Progress and Prototype Development (CT)

Ivo Santos

University of Évora, Portugal & Universidad Complutense de Madrid, Spain

Email: ifs@uevora.pt

Portuguese Archaeological grey literature plays a critical role in the discipline, but remains somehow inaccessible due to its unstructured format, limited circulation, and non-compliance with FAIR principles. This project aims to address these challenges by implementing a systematic workflow for extracting and structuring information from 11000 Portuguese archaeological work reports.

The reports were processed using a hybrid approach combining rule-based approaches, generic Named Entity Recognition models, LLMs, and others. The extraction workflow yielded structured text sections, raster, vector, and domain-relevant Named Entities. All the information extracted is currently being enriched with other data sources, and integrated on a Knowledge Graph aligned with the CIDOC-CRM ontology.

To ensure a thorough review of all extracted data, we assume that the workflow should culminate in a web interface. This interface, using community accepted software, will allow users to query the data, but also to upload new data for processing and review existing elements. The confirmation of these reviews will, in turn, be used for future iterations of the project.

CODA - Centre for Digital Culture and Innovation (IT)

Vera Moitinho de Almeida

University of Porto, Portugal

Email: vmoitinho@letras.up.pt

The Centre for Digital Culture and Innovation (CODA) is a hub on Digital Humanities that stands out by its potencial to generate new dynamics in the Humanities and Social Sciences. Based at the Faculty of Arts and Humanities of the University of Porto (FLUP), CODA's main objective is to reinforce and stimulate trans and Interdisciplinary research between FLUP's R&D units and other (inter)national collaborations, through projects and activities that prove to be aggregating and demonstrative in the field of Digital Humanities.

In this talk, I will showcase some of the projects and activities we have been developing and carrying out over the past two years. CODA is funded by the Portuguese Foundation for Science and Technology (FCT), under the CEECINST/00050/2021 contract programme.

Session 3: Computational Models for Text Recognition and Reuse

Modeling Text Reuse: New Explorations of Paraphrase and Translation in Historical Arabic Books (IT)

Sarah Savant & Peter Verkinderen

Centre for Digital Humanities, The Aga Kahn University, United Kingdom

Email: sarah.savant@aku.edu & Peter.Verkinderen@aku.edu

This presentation explores the role that recycling played in making the Arabic tradition one of the world's largest. I will first introduce an original data set, created by the KITAB project team, of more than 1.6 million files documenting word-for-word relationships between thousands of books in Arabic from over a 1000-year period. The team adapted an algorithm by David Smith (Northeastern University) called 'passim', which relies chiefly on a complex statistical model, which I will explain. The books under study span every conceivable topic, from religion, philosophy, and language to history, geography, medicine, and astronomy, and were written by authors in a region spanning modern Spain to Central and South Asia. For most writers, reusing earlier works was the starting point for creating new ones. They created histories, for example, by updating earlier accounts of the origins of Islam and the Prophet's life for their own times. They abbreviated long works to make short ones, commented on short ones to make long ones, and hunted across general histories to create themed works. In these and many more ways, they produced an enormously intertextual tradition. The KITAB team has created a web application that enables scholarly investigation of these relationships and access to two visualisations, data and book files (KITAB-project.org for the application and all documentation).

Then, I will turn to our current work, just initiated, which follows on from the first phase. The team now is seeking to produce a series of computational models designed to detect and classify transformative text reuse (paraphrase, translation) within and between Arabic and Persian. We will use the resulting data to address previously unconnected problems that have vexed historians, ranging from the relationship between oral and written transmission practices to recovery of 'lost' texts. These cases are linked by ways that authors reused earlier works, but in ways that the team could not previously detect with passim. Key points for discussion will be our modeling process (reliant on AI and machine learning) and our method for working between computer science and humanities.

Arabic Handwritten Text Recognition: Machine Learning and the Future Study of Arabic-Script Manuscripts (IT)

Mathew Barber

Centre for Digital Humanities, The Aga Kahn University, United Kingdom

Email: mathew.barber@aku.edu

In the world today there are possibly more than half a million Arabic-script manuscripts. Modern research, however, only benefits from a small sub-set of this larger corpus. For many manuscripts little or nothing is known about their content, even though vast numbers have now been digitally imaged. To an extent, research continues to focus on a relatively narrow Arabic literary canon that – it has been argued - had emerged with the introduction of Arabic printing during the nineteenth century (Shamsy 2020). Even where editors and historians step beyond that canon, their research necessarily focuses on narrow subject areas. It is only with large-scale and effective Arabic-script Handwritten Text Recognition, that scholarship will truly understand the scope and depth of the Arabic written tradition and begin to address questions about its survival.

In this presentation, I aim to share the current state-of-the-art in Arabic-script HTR and present a near-future in which this method is applied to whole collections of manuscripts, in three parts. Part 1 will introduce the work of the Open Islamicate Texts Initiative (OpenITI) in developing approaches to HTR, underlining the importance of using open-source approaches. Part 2 will outline a vision for a large multi-library HTR effort that would enrich access to manuscripts and fundamentally change the way in which the discipline (and the public) engages with the Arabic written tradition. Part 3 will then present how this knowledge might be applied to questions of manuscript survival and to reconstruction. Although the surviving corpus is huge, it likely represents a small portion of the total number of Arabic-script manuscripts ever written. Once we have a better understanding of the contents of the tradition (through HTR), we will be better placed to explore the gaps within it. Using some case studies, I will demonstrate how data produced from large-scale HTR might be applied to the reconstruction of lacunae in individual manuscripts and to identify lost texts or genres. While the cases here focus on the Arabic script, the approach might be generalised to any language or written tradition. In other words, Arabic-script materials - owing to the scale and complexity of the tradition - are a good test case for a robust pipeline that might applied to other literatures.

References

Shamsy, Ahmed El. 2020. *Rediscovering the Islamic Classics: How Editors and Print Culture Transformed an Intellectual Tradition*. Princeton (N. J.): Princeton University Press.

History in the face of Digital Humanities: constraints and challenges (CT)

Helena Cameron[1,2] & Fernanda Olival[2]

[1] *Polytechnic Institute of Portalegre, Portugal*
[2] *CIDEHUS, University of Évora, Portugal*

Email: helenac@ipportalegre.pt

Before working in digital environment, historians need to balance what is their work and how it can be changed. This means to face the structural bases of historical research and, at the same time, deal with the complexity of digital transformation, to take profit of AI and HPC opportunities in this field.

This presentation aims to discuss the process of research work in Historical research in Portuguese language, as it represents additional challenges, similar to other languages, except English.

However, there are several constraints when it comes to the Portuguese language. Although a global language, it can still be considered "peripheral," as digital tools and resources in this domain are less developed than in English. The geographical dispersion, different patterns, and the historical evolution of the Portuguese language add another layer of complexity. This underscores the urgent need for specialized tools and strategies to ensure equitable access to technological advancements in the humanities for Portuguese speakers.

Addressing historical and linguistic complexities requires the development of approaches within interdisciplinary research teams. These teams, by designing customised computational solutions to tackle complex problems, can effectively leverage technology while promoting innovation and transparency.

Investing in computational development for the study of History in the Portuguese language is not just a priority, it's a necessity. This investment is crucial to ensure access to our rich textual heritage and to preserve the diversity of the language, regardless of its pattern or linguistic period.

References

Baron, A. 2011. *Dealing with Spelling variation in Early Modern English Texts*, Lancaster University (Ph. D. Thesis).

Bollmann, M. 2018. *Normalization of historical texts with neural network models.* Universitätsbibliothek Johann Christian Senckenberg (Ph.D. Thesis).

Edmond, J. (ed) 2020. *Digital Technology and the Practices of Humanities Research.* Cambridge, UK: Open Book Publishers.

Gabay, S., Barrault, L. 2020. *Traduction automatique pour la normalisation du français du XVIIème siècle.* TALN 2020, ATALA, France.

McGillivray, B. & Tóth, G. M. 2020. *Applying Language Technology in Humanities Research: Design, Application, and the Underlying Logic*, Palgrave MacMillan.

Modelling Data for Long-Term Archiving in the Swiss DaSCH Service Platform (DSP) (IT)

José Luis Losada Palenzuela

University of Basel, Switzerland

Email: joseluis.losadapalenzuela@unibas.ch

The Swiss National Data and Service Center for the Humanities (DaSCH) is a national infrastructure dedicated to the long-term archiving and accessibility of research data in the humanities. DaSCH serves as an open data repository and a simple virtual research environment. It focuses on ensuring the archival integrity and direct accessibility of data, following Open Research Data (ORD) and FAIR principles.

At the heart of DaSCH is the DaSCH Service Platform (DSP), particularly suited for handling text data and bitstream data (images, audio, and video) from various humanities disciplines. The platform supports qualitative data and offers access control to address copyright and other sensitive data concerns, while advocating for maximum openness. One of the key features of DSP is its ability to assign persistent identifiers (ARKs) at the object level, enabling precise citation of individual data items.

The foundation of DSP's data modelling flexibility lies in its use of RDF and ontologies. Each project defines its own ontology, extending the common knora-base. This allows the data model to be tailored precisely to the research subject matter. Consequently, models can vary significantly in complexity. A project might choose a relatively simple model, perhaps primarily archiving XML files as digital objects with basic metadata. Conversely, another project might implement a highly complex model, representing the intricate structure of, for example, a scholarly edition, including detailed registers of persons, places, and concepts, with extensive internal linking between these entities and the primary source materials.

In this presentation, I will address how researchers can approach data modelling for long-term archiving using DSP. I will discuss the conceptual foundations of the platform's ontology-based architecture and explain how tools such as the DSP ontology editor and DSP-Tools support the creation, validation, and management of custom data models. Drawing on examples from existing projects, I will illustrate how different modelling strategies can be adapted to diverse research needs.

References

DSP Documentation: https://docs.dasch.swiss
DSP source code and documentation: https://github.com/dasch-swiss
Research Projects: https://app.dasch.swiss

Meeting Closure & Farewell