CINECA

AI, Cultural Heritage, Art and Science

Workshop proceedings – March 27, 2026



EDITORS

Antonella Guidazzoli

Maria Chiara Liguori

AI, Cultural Heritage, Art and Science

Workshop proceedings – March 27, 2026

Edited by Antonella Guidazzoli and Maria Chiara Liguori

@ CINECA, 2026
www.cineca.it
visitlab@cineca.it

Copyright

I contenuti di questa pubblicazione sono protetti ai sensi della normativa in tema di opere dell'ingegno.

I diritti di questa pubblicazione appartengono a CINECA e ai rispettivi autori.

È consentito utilizzare, copiare e distribuire la presente pubblicazione e le sue parti senza fini di lucro.

Le note di copyright, gli autori ove indicati o la fonte stessa devono in tutti i casi essere citati nelle pubblicazioni in qualunque forma realizzate e diffuse.

Qualsiasi forma di link al presente sito, se inserita da soggetti terzi, non deve recare danno all'immagine ed alle attività di CINECA e degli autori della presente pubblicazione.

È vietato il cosiddetto deep linking ossia l'utilizzo non trasparente, su siti di soggetti terzi.

L'eventuale inosservanza delle presenti disposizioni, salvo esplicita autorizzazione scritta, sarà perseguita nelle competenti sedi giudiziarie civili e penali.

ISBN – 978-88-86037-72-3

DOI – 10.82341/6fd4-5d78

CINECA

SUMMARY

01 **FOREWORD** 5
Antonella Guidazzoli, Maria Chiara Liguori

02 **INSTITUTIONAL MESSAGE** 6
Giacomo Scillia

03 **THE ROLE OF DATA SERVICES IN AI FACTORIES FOR CULTURAL HERITAGE** 8
Arnaud Céol

04 **AI AND HISTORICAL NEWSPAPERS** 11
Giovanni Baisi, Fauzia Albertin, Federico Andrucci, Simona Caraceni, Giorgia Cardano, Alex Gianelli, Antonella Guidazzoli, Silvano Imboden, Angela Lanzarini, Maria Chiara Liguori, Gabriele Marconi, Luca Mattei, Rossella Pansini, Donatella Sforzini

05 **THE "OCTOPUS" PROJECT** 22
Roberto Balzani, Elena Musiani

06 **NEW QUESTIONS FOR CULTURAL HERITAGE: SOURCES, FORMATS AND INFRASTRUCTURES IN THE AGE OF ARTIFICIAL INTELLIGENCE - Panel** 23
Elena Musiani, Angelo Pompilio, Elena Cervellati, Giuseppe Di Modica, Mirko Degli Esposti
Giovanni Baisi (moderator)

07 **INTRODUCTION TO THE AFTERNOON SESSION: INTANGIBLE CULTURAL HERITAGE** 26
Alberto Salarelli, Francesca Zanella

08 **INTANGIBLE CULTURAL HERITAGE AND NEW TECHNOLOGIES: A LEGAL PERSPECTIVE** 28
Valentina Gastaldo

09 **DANCE, VIRTUAL REALITY AND CULTURAL HERITAGE: THE DOUBLE REALITY PROJECT OF THE CENTRO COREOGRAFICO NAZIONALE / ATERBALLETTO** 33
Fabio Ferretti

10 **DanXe: A DIGITAL ECOSYSTEM FOR BALLET HERITAGE. METHODS, TOOLS AND PERSPECTIVES** 38
Silvia Garzarella

11 EXPERIENCING THE ARCHIVE: DIGITAL COMMUNICATION PRACTICES FOR A SHARED INTANGIBLE HERITAGE 43
Elena Radaelli

12 ACKNOWLEDGEMENTS 50

CHAPTER 01

FOREWORD

Antonella Guidazzoli, Maria Chiara Liguori – VisitLab, CINECA

Three years in from the first Workshop on *AI, Cultural Heritage, Art and Science*, held in January 2023, the field of artificial intelligence continues to evolve at an extraordinary pace. However, despite the growing prevalence of AI-based tools and applications, the process of discovering the potential offered by these technologies across various spheres of knowledge is far from complete.

Case studies represent one of the main tools through which this exploration can progress. Artificial intelligence, in fact, cannot be compared to a tool designed for a single purpose, such as a hammer: it does not immediately suggest its own use, nor is it limited to a single function. On the contrary, it is an extremely flexible resource, capable of being adapted, combined and reinterpreted to meet a wide variety of needs, which are often initially unforeseeable.

A notable example is provided by *layout-parsing* tools, developed to automatically identify the different areas that make up a printed page and enable its accurate transcription. Although originally designed to solve a specific problem, these tools produce information that can prove invaluable in a wide range of research fields: from the analysis of advertising to the history of language, and even the study of historical meteorology. Similarly, a model trained to detect and analyse human movement can be applied in dance research, but also in the medical, biomechanical, or rehabilitation fields.

The general and adaptable nature of artificial intelligence makes the contribution of a broad, interdisciplinary community of researchers essential, each bringing different and specific skills, perspectives and questions to the table. It is precisely from the convergence of different disciplines that novel applications, new research methods and innovative ways of harnessing knowledge can emerge.

From this perspective, the Workshop reaffirms its role as a forum for meeting, discussion and interdisciplinary exchange, where practical experiences provide an opportunity to share findings, identify new avenues of research and continue to explore the potential of artificial intelligence in cultural heritage, art and science.



CHAPTER 02

INSTITUTIONAL MESSAGE

Giacomo Scillìa - Deputy Director, CINECA

September 1959. A newspaper page, its ink faded. Amid the crowded columns, an advertisement for a jeweller sits alongside one for a spirit and a job vacancy. Ten years later, in 1969, a group of Italian universities would come together to share a supercomputer and put computing power at the service of research: CINECA was born. Today that page and that story meet. An algorithm, running on CINECA HPC infrastructure, divides the page into rectangles in a few seconds. Then a scholar, who has spent a lifetime on those years among newspaper archives and hand-compiled spreadsheets, corrects it box by box. The machine proposes and the scholar decides. That simple gesture captures much of what this volume has to say. It also embodies much of what we at CINECA believe about the future of artificial intelligence in the cultural sphere.

The contributions collected in this volume reveal a field that has moved beyond its initial wonder. AI no longer surprises us with what it can do: it challenges us with what it demands. First of all, it demands data: carefully curated, well-documented and reliable data. As the experience of the IT4LIA AI Factory shows, the real barriers to adoption are no longer computational. They lie in the availability, quality and usability of data. Those humanities scholars who took part in the panel reached the same conclusion from a different perspective: what is lacking is not modelling capacity but scholarly-grade material. This challenges a widely held assumption. In the age of AI, the most valuable resource is the patient work of archivists, philologists, musicologists and dance historians: forty thousand stage directions, for example, individually transcribed from eight thousand librettos.

The second lesson is that keeping humans in the loop is not a temporary compromise, destined to disappear as automation advances. It is a fundamental design principle. The layout-parsing platform developed by our Data Visualization office, the DanXe toolbox and the Ecomic ecosystem share one principle: machines take on the repetitive work, while intuition, hypothesis, validation and interpretation firmly remain in human hands. The research process retains its structure; what changes is the time required to carry it out. When a model misclassifies half of the brands in a set of advertisements and yet saves months of work, this is not a failure but rather a new division of labour.

The third lesson takes us to a new frontier: cultural heritage that cannot be captured in writing alone. A dancer's movement, a puppeteer's gesture, the expertise of a traditional craft or the sound a nineteenth-century critic struggled to describe. Through projects such as DUETTO's Large Movement Model, the virtual stages of Double Reality and the marionettes of Marionette 3.0, AI meets forms of heritage that exists only through their performance. Here lie some of the hardest questions, but also some of the most fascinating possibilities. If authenticity rests on the continuity between a cultural practice and its community, then technology must serve that relationship and never replace it.

Looking ahead, I believe that the responsibility of an institution like CINECA begins with cultural heritage itself: we must turn it into infrastructure. Digitising collections is no longer enough. They must become structured, accessible cultural data that artificial intelligence can effectively use. This should happen in close partnership with museums, archives and libraries, not simply as technology users but as active partners in shaping the ecosystem. Supercomputing, AI Factories and data services fulfil their purpose only when they are accessible to those who preserve and safeguard our collective memory.

But the challenge goes beyond technology. Europe's cultural memory cannot simply become training

material for models developed elsewhere. Public, interoperable and shared infrastructures are nowadays an essential condition to preserve our sovereignty and, ultimately, our capacity for cultural self-determination.

Above all, however, this is about people. During the workshop we heard young people express their fear of becoming redundant. The contributions in this volume point to a different conclusion: these tools call for more knowledge, not less. There will always be need for people who can establish ground truth, formulate the right questions and interpret the answer. These are the skills in which we must invest.

In 1969, sharing a computer was an act of trust among universities. Today, sharing infrastructures, data and models is an act of trust among institutions, public administrations, communities and generations. Our mission remains the same, but its horizons have expanded: not only to provide computing power, but to ensure that our society's collective memory remains alive, challenged and shared. This includes the memory preserved in a 1959 newspaper page, which, thanks to a machine and a scholar working together, still has a story to tell today.

My sincere thanks go to Antonella Guidazzoli, Maria Chiara Liguori and the entire CINECA team of Data Visualization. I am equally grateful to the University of Parma, the University of Modena and Reggio Emilia, the Italian Ministry of Culture's Digital Library and all the contributors who chose to stand, as someone put it during the workshop, "on the side of the questions".

CHAPTER 03

THE ROLE OF DATA SERVICES IN AI FACTORIES FOR CULTURAL HERITAGE

Arnaud Céol - Fondazione ICSC, Centro Nazionale HPC, BIG DATA AND QUANTUM COMPUTING

Introduction

The European Union has articulated an ambitious strategy to position itself as a leading global actor in artificial intelligence, promoting a model grounded in trust, human-centric values and the protection of fundamental rights. Within this framework, AI Factories provide integrated environments that combine computational resources, data and expertise to accelerate AI adoption across sectors.

IT4LIA, the Italian AI Factory, builds on this vision through advanced infrastructures such as EuroHPC supercomputers and cloud platforms [1]. However, early interactions with user communities show that the main barriers to AI adoption are not computational, but data-related, particularly in terms of availability, accessibility, quality and usability. Datasets are often fragmented, heterogeneous and insufficiently described, while limited technical skills further constrain the ability of many organisations to exploit these resources [2].

This contribution focuses on two key aspects: data availability, related to discoverability and access, and AI data readiness, referring to the processes required to make data effectively usable in AI workflows. In this context, this contribution is both a presentation of the IT4LIA ecosystem and an invitation to participate. The development of a sustainable AI ecosystem depends on the active involvement of stakeholders, including cultural institutions, both as users and as contributors of data.

Data Challenges and Services in IT4LIA

Building on these challenges, IT4LIA positions data services as a central component of the AI Factory. In many cases, datasets exist but remain fragmented across institutions, stored in heterogeneous repositories, or subject to access constraints that limit their reuse. Beyond availability, issues of heterogeneity and quality further complicate data usage. Datasets are often created for preservation or documentation purposes rather than computational analysis, resulting in differences in formats, annotation practices and levels of completeness. Gaps, inconsistencies and biases may affect the reliability of AI models.

These challenges are compounded by a lack of technical and methodological skills. Many organizations recognize the potential of AI, but lack the expertise required to prepare data, manage workflows, or operate within HPC and cloud environments. This highlights the need for structured support addressing both data-related issues and competences.

In this context, IT4LIA provides a set of data services covering the entire data lifecycle, including data discovery, preparation and standardization, quality assessment, and secure management and access. Particular attention is given to facilitating the transformation of heterogeneous datasets into resources that can be effectively used within AI workflows.

More broadly, AI Factories act as enabling environments, where data, tools and expertise are integrated into a coherent framework. In this perspective, the concept of AI data readiness

provides a useful framework to structure these efforts.

AI-Ready Data and Governance

A central element of the IT4LIA approach is the recognition of AI data readiness as a prerequisite for the effective use of data in artificial intelligence applications. Rather than focusing only on dataset aggregation, the approach emphasizes the need to assess data maturity and support data owners in improving usability.

Achieving this involves a structured process, starting with data onboarding, where datasets are analysed together with their owners to identify key characteristics such as structure, provenance, and constraints. Emphasis is placed on metadata, whose harmonization and enrichment are essential for discoverability and interoperability, and on integration into AI workflows.

Data are therefore treated as evolving assets, requiring continuous quality assessment and iterative improvement. This includes addressing gaps, inconsistencies, and potential biases, while also building upon high-quality datasets and advanced use cases already developed within institutions and research projects.

The outcome of these efforts is the creation of *data products*, i.e. curated datasets prepared for AI applications and accompanied by documentation and usage guidelines. These reduce the effort required to start new projects, while also enabling the reuse and extension of existing mature solutions.

This perspective is particularly relevant in domains characterized by heterogeneous and distributed data sources, such as Cultural Heritage. It is important to underline that many initiatives in this field already demonstrate a significant level of maturity, as highlighted by projects and contributions presented during the workshop. In such contexts, AI data readiness is not a distant objective, but an ongoing process of consolidation, integration, and scaling of existing capabilities.

Overall, AI data readiness can be understood as a progressive and collaborative process that both supports less structured datasets and builds upon existing excellence, contributing to a more coherent and reusable data ecosystem.

Applications to Cultural Heritage

The Cultural Heritage domain presents specific challenges for the adoption of artificial intelligence, mainly due to the heterogeneity and complexity of the data. Collections typically include images, texts, structured records, and audiovisual materials, produced using different standards and practices. While this diversity represents a valuable resource, it also creates barriers to integration and reuse in AI workflows.

In this context, IT4LIA provides an enabling environment to support cultural institutions. By combining computational infrastructures with data services, the platform offers tools to access, prepare and analyse heterogeneous datasets in a coherent and scalable way, while remaining adaptable to the needs of museums, archives and libraries.

A key objective is to support the accessibility and reuse of digital collections. Although significant digitization efforts have been carried out, their impact depends on the availability of services for data discovery, preparation, and access. IT4LIA contributes by lowering technical barriers and facilitating the transformation of digital assets into usable resources for research and innovation, while respecting legal and ethical constraints.

The integration of data services with AI tools enables new forms of analysis, such as classification, transcription, and semantic enrichment, complementing traditional research practices. At the same time, IT4LIA acts as a collaborative environment, encouraging cultural institutions to participate both as users and as contributors of data, expertise and use cases.

This contribution is therefore intended as an invitation to engage with the platform, enhancing the visibility and usability of cultural collections while contributing to a shared ecosystem.

Conclusions

This contribution has highlighted the central role of data services in enabling the effective use of artificial

intelligence, emphasizing that computational infrastructures alone are not sufficient. A primary focus has been placed on aspects related to data availability, quality and usability, which continue to represent significant barriers across domains.

In the field of Cultural Heritage, these aspects are particularly evident. The richness and diversity of cultural data offer significant opportunities, but also require dedicated efforts to address heterogeneity, fragmentation, and accessibility. In this context, the approach pursued within IT4LIA provides an environment where data services, computational resources, and expertise can be combined to support the progressive transformation of digital collections into usable resources for AI.

At the same time, the development of a shared ecosystem depends on active participation from the community. Cultural Heritage institutions are invited to engage both as users and as contributors, bringing their data and expertise into the platform.

This perspective extends to a broader European context, where AI Factories, data spaces and infrastructures such as the European Open Science Cloud support collaboration and data sharing across domains [3]. Strengthening data services and fostering community participation represent key steps toward building interoperable and sustainable ecosystems for the responsible use of artificial intelligence.

Acknowledgments

IT4LIA AI FACTORY is funded by the Ministry of University and Research and the National Cybersecurity Agency. The initiative is co-funded by the European Commission through the European High-Performance Computing Joint Undertaking under Grant Agreement No. 101234224. Additional contributions come from Austria, Slovenia, as well as from CINECA, the Emilia-Romagna Region, the Italian National Institute for Nuclear Physics, the Italian Institute of Artificial Intelligence for Industry, Fondazione Bruno Kessler, and the ItaliaMeteo Agency.

References

[1] “AI Factories - The European High-Performance Computing Joint Undertaking (EuroHPC JU).” Accessed: May 07, 2026. [Online]. Available: https://www.eurohpc-ju.europa.eu/ai-factories_en

[2] “Artificial Intelligence in Italy: 65% of Organizations Have AI Projects Stalled by Skills Shortages.” Accessed: May 08, 2026. [Online]. Available: https://it4lia-aifactory.eu/news/https-www-it4lia-ai-factory-eu-ai-projects-italy-skills-shortages/

[3] "European Open Science Cloud." Accessed: October 16, 2026. [Online] Available: https://open-science-cloud.ec.europa.eu/

CHAPTER 04

AI AND HISTORICAL NEWSPAPERS

Giovanni Baisi, Fauzia Albertin, Federico Andrucci, Simona Caraceni, Giorgia Cardano, Alex Gianelli, Antonella Guidazzoli, Silvano Imboden, Angela Lanzarini, Maria Chiara Liguori, Gabriele Marconi, Luca Mattei, Rossella Pansini, Donatella Sforzini - CINECA

Introduction

Historical newspapers represent one of the most vital and comprehensive repositories of information regarding modern society. Beyond merely recording historical events, they serve as a dynamic mirror of contemporary social trends, cultural shifts, and public phenomena. Consequently, they constitute an important resource for researchers in the Humanities and Social Sciences.

However, unlocking the vast wealth of knowledge embedded within these archives presents significant methodological and technological challenges. While large-scale digitization campaigns have successfully preserved millions of pages, the resulting digital surrogates are often limited to raw images or unsegmented text transcripts. In most cases, current open-access portals to historical newspapers' digitisations lack structural metadata, failing to incorporate essential information regarding page layout, article segmentation, and individual column transcriptions. Because newspapers are intrinsically dense, multi-topic, and voluminous, manual data extraction and analysis are very time-consuming.

The Role of Artificial Intelligence and Interdisciplinary Collaboration

To address these limitations, the integration of advanced computational methodologies has become essential. Over the past few years, CINECA Data Visualization Lab has applied cutting-edge technological solutions to the field of Cultural Heritage. By leveraging a multidisciplinary team of experts specializing in Artificial Intelligence (AI), data visualization, and domain-specific research in humanities, it actively supports research initiatives dedicated to the analysis, preservation, and valorization of cultural assets.

This synergy between innovative technologies and domain expertise is further strengthened by a commitment to the principles of Open Science. By aligning technical workflows with European research standards, we ensure that the developed solutions are not only sustainable and reproducible but also directly tailored to the actual needs of the scientific community.

Within the framework of a special project with The Italian Ministry of Culture – Digital Library (Ecomic, Ecosistema digitale per la cultura), a continuous technological scouting activity was launched to identify [1], benchmark, and adapt state-of-the-art AI models to specific use-cases dedicated to HCR [2] and the layout analysis of historical newspapers [3].

Human-AI Collaboration and Data Preparation Platforms

A critical bottleneck in deploying AI for the Digital Humanities lies in the preparation of high-quality training and fine-tuning data. To bridge this gap, two specialized AI models were developed in tandem with data preparation platforms [4]. Recognizing that advanced machine learning pipelines can be intimidating or inaccessible to non-technical scholars, these applications were designed with intuitive, user-friendly interfaces that allow immediate operational deployment.

The development of these platforms is strictly anchored to two foundational pillars:

- **The Human-AI Collaboration Model**: Rather than aiming for fully automated, black-box automation, the platform fosters a relationship between human expertise and machine efficiency. The AI accelerates repetitive tasks (such as layout analysis and optical character recognition), while the researcher remains "in the loop" to validate, refine, and interpret the outputs.
- **Open Science Principles**: In compliance with modern European research paradigms, the methodologies are designed to be transparent, shareable, and reusable, thereby fostering democratization of data and collaborative research.

Structure and Contribution of this Paper

This paper aims to demonstrate how the integration of tailor-made AI models and data preparation platforms can revolutionize the heuristics of historical newspaper research. The present contribution is organized as follows:

- First, we provide a comprehensive review of the state of the art regarding currently available online resources and platforms for historical newspaper consultation.
- Second, we explore the transformative potential of AI tools in this domain, mapping technical capabilities onto realistic user journeys designed and co-created by researchers.
- Third, we deliver a technical and functional description of the developed data preparation platform, illustrating how its user interface supports seamless human-in-the-loop workflows.
- Finally, we present a case study demonstrating the practical application of this ecosystem: the creation of a specialized, structured dataset dedicated to historical advertising. Through this use case, we validate the effectiveness of our approach in transforming raw archives into structured, searchable, and analytically viable research assets.

Online resources and platforms for historical newspaper consultation

Before creating a newspapers consultation platform, several systems for newspaper consultation and retrieval were reviewed. The platforms were divided by a selection of criteria, such as their commercial or public/research purpose, the number of pages available, if they use OCR for text retrieval, and if they offer personalized services (Table 1). The commercial platforms that were taken into consideration were newspapers.com and newspaperarchive.com.

Newspaper.com appears to be a sub-service of Ancestry.com, a paid service for ancestry research, genealogy, and event history. Since it's a paid service, the investigation was not continued. **Newspaperarchive.com** has the same commercial research purpose, but it is not a sub-service of a more thematic, generalist portal. It gives industries and professionals access to published documentation, and as nespapers.com it has a commercial purpose and was not investigated for this research.

The public and research platforms that were examined are Gallica, Library Of Congress (LOC) Newspaper Navigator, Europeana Newspapers, MLOL, EBiblio, Internet Archive Newspapers, and La Stampa newspapers Archive.

Gallica uses a particular searches system, allowing users to consult, for example, entire editions of newspapers, magazines, and databases. As regards, for example, the research on *Le Figaro*, there are 32,120 issues for 100 years of history of the newspaper from 1854 to 1954, which in all likelihood respects the copyright law of the 70 years after which an intellectual work can change the level of sharing policies by becoming public access. For Gallica, in-depth search allows you to refine the results, which are presented according to the set criteria; OCR is revealed in part of the screen: it is possible to select the text, but without differentiation between articles on the page.

The situation is different if you search for full newspapers pages. In this case, the information will always be presented in PDF, but with easily zoomable text.

LOC Newspaper Navigator went into production in 2020 and was discontinued in 2025. It was an experimental web application that used machine learning to enable search and navigation based on the visual similarity of over 1.56 million images extracted from the Chronicling America database of digitized historical newspapers. The experiment was conducted in 2020 by Benjamin Lee as part of the Library of Congress's Innovator in Residence program and was the first machine learning application developed internally at the LOC. The Newspaper Navigator application was hosted by the Library of Congress from 2020 to 2025. In addition to the research application, the Newspaper Navigator experiment produced a dataset consisting of visual content extracted from 16,358,041 pages of historical newspapers in Chronicling America, as well as a machine learning pipeline and other content.

The Newspaper Navigator project has received significant recognition in the Digital Humanities context, including the Best Dataset award at the 2020 DH Awards and second place for Best Resource Paper at the 2020 CIKM conference. The Chronicling America – Historic American Newspapers collection offers access to a selection of digitized newspaper pages, created as part of the National Digital Newspaper Program (NDNP), a collaboration between the National Endowment for the Humanities (NEH) and the Library of Congress. The Chronicling America platform collects millions of newspaper pages from nearly every US state and territory, published from first editions, through 1963. For more information on current participation, see the list of NDNP grant recipients. Further newspaper pages are regularly added to the collection. Content exploration is made possible through an interactive map and timeline interface, accompanied by additional data visualizations that provide a deeper understanding of the breadth and coverage of the newspapers currently available in the collection.

The Europeana Historic Newspapers collection was created through the **Europeana Newspapers project**, an EU-funded project that ran from 2012 to 2015. This project created 10 million pages of full text content using Optical Character Recognition (OCR) on existing digital images provided by partner libraries, build the Historic Newspapers interface on The European Library, and provide newspaper headline information to Europeana. The material included in the Historic Newspapers collection was contributed by numerous European libraries. Much of the content is in the public domain, but some remains under copyright. Written permission from the contributing library is required before reproducing, distributing, or commercially exploiting any material from the collection. The European Library is interested in exploring other ways to redistribute the data with participating libraries, such as through APIs and open data. Europeana Newspapers is a project funded by the European Commission's ICT Policy Support Programme, as part of the Competitiveness and Innovation Framework Programme (CIP).

The newsstand of **Italian Media Library Online (MLOL)**, similarly to Spanish **EBiblio** was born between 2005 and 2008 as a collaboration between Horizons Unlimited srl of Bologna, the Province of Reggio Emilia, and then the Consorzio Sistema Bibliotecario Nord-Ovest (CSBNO). MLOL is a system for distributing any type of digital object, whether open access or commercial (newspapers, audio, video, text, paid databases, historical texts in image format, iconographic archives, audiobooks, digitized books and commercial eBooks, e-learning, real-time live-casting, etc.); a Digital Asset Management portal for managing all licensing and copyright issues; a national network of libraries, library systems, and other organizations that collaborate and share the costs of managing digital resources. MLOL aggregates commercial content (e.g., username/password-protected online resources), allowing its registered users direct access to resources without additional logins. This is where MLOL's "Newsstand" space was born, allowing users to browse the day newspapers or access publishers' online resources, according to their individual policies. This may involve situations where no more than a small fixed number of users can access the resource per day, or at any one time, or where all issues of a newspaper are visible to the newspaper reader. From this perspective, two broad typologies of newspaper consultation can be identified: or the newspaper has its own proprietary reader, or it relies on pressreader.com, a widely used global news platform.

The **Internet Archive** has a section for browsing newspapers. The database holds just over 2.5 million "objects," that is newspaper issues, which can be searched through a timeline or using text search tools to explore OCR-recognized text on pages. The

newspapers in this collection were digitized as part of a pilot project using microfilm and microfiche. After using a microfilm/microfiche scanner to create a digital image of each page, the resulting images were processed so that each reel is contained in a single, easily navigable file.

The ***La Stampa* Historical Archive** (also included in the Internet Archive) contains a digital copy of the entire collection of every edition of the newspaper, from the first issue on February 9, 1867, when it was called *Gazzetta Piemontese*, until 2005. All pages published over these 138 years are available in digital text and image format, as are the main articles (over 12 million), with the ability to search and highlight all the words they contain. The Archive, totaling 1,761,000 pages, spans from 1867 under the *Gazzetta Piemontese* masthead to 2005, including the years when the *Stampa Sera* afternoon edition was published. The pages are digitized from black-and-white microfilm (over 1,600 reels of approximately 1,000 pages each).

Table 1

Name	Commercial/research purpose	Capacity	Metadata	OCR	Login for personalization
Newspapers.com	Commercial	More than a billion pages	N/A	N/A	Yes
Newspaperarchive.com	Commercial	Millions of pages	N/A	N/A	Yes
Gallica	Public Research	Millions of pages	Yes, international standards	Yes	Yes
Newspaper Navigator	Public Research	Millions of pages	Yes, international standards	Yes	Yes
Chronicling America	Public Research	Millions of pages	Yes, international standards	Yes	Yes
Europeana Newspapers	Public Research	Millions of pages	Yes, international standards	Yes	Yes
MLOL	Commercial, free for the Public	Millions of pages	N/A	No	Yes
EBiblio	Commercial, free for the Public	Millions of pages	N/A	No	Yes
Internet Archive Newspapers	Public Research	Millions of pages	Yes, international standards	Yes	No
La Stampa Archive	Public Research	Millions of pages	Yes, international standards	Yes	No

In a comparison between the examined platforms, what became clear in all the systems we investigated is that the text of the OCR is unstructured, put as an undifferentiated list of characters, making the comprehension of the articles and the titles very difficult to handle without the view of the image of the related page. This limit made us think about the LPW platform that we built in order to facilitate the Artificial Intelligence system to be better in taking into consideration the structure of the page with titles, text of the articles, images and ads for an improved OCR output.

Historical Newspapers analysis

Recent advances in Artificial Intelligence have significantly improved document analysis. However, effectively exploiting these technologies for historical newspapers requires processing workflows specifically tailored to their unique characteristics, made of complex layouts, heterogeneous typography, and the variability of page designs across newspapers and historical periods, which cannot be adequately addressed by generic document analysis approaches. Furthermore, while the core processing stages are largely independent of the research objective, the subsequent information extraction workflow is inherently research-specific and must be designed according to the target research questions.

To address these challenges, we developed a core processing pipeline specifically designed for large-scale historical newspaper collections. The pipeline combines image selection, layout analysis, and Optical Character Recognition (OCR) to transform digitized newspaper page images into machine-readable transcriptions of individual articles together with their associated metadata. Layout analysis was performed using a fine-tuned version of LayoutParser [5]. Although this model represents the

state of the art for modular layout analysis–OCR pipelines and is one of the most widely used open-source solutions, previous evaluations have shown limited performance on historical Italian newspapers. To address this limitation, the model was fine-tuned on an annotated dataset of pages from *Il Resto del Carlino* [3; 6]. The resulting model (LPW_Carlino40) improves the detection of page regions across diverse historical newspaper layouts and recognizes five layout categories: title, body text, photograph, drawing/graph, and advertisement. For Optical Character Recognition (OCR), we employed the open-source Tesseract [7] engine. In contrast to layout analysis, qualitative evaluation showed that Tesseract produced sufficiently accurate transcriptions on the processed collections without requiring task-specific fine-tuning.

Together, layout analysis and OCR provide a common processing framework that transforms digitized newspapers into structured, machine-readable articles. The resulting articles constitute a common intermediate representation that can support a wide range of downstream research applications. However, extracting domain-specific knowledge requires dedicated workflows tailored to the research topic, moving beyond machine-readable text to produce structured, searchable digital resources.

To demonstrate the flexibility of this modular approach, three research use cases and workflow are presented: one focusing on historical advertisements, another on the evolution of the term *piovra* and a third on the analysis of narratives related to climate events.

Use case: Advertisements in historical newspapers

Advertisements within historical newspapers not only reflect the economic context of their time but also provide valuable insights into the role of print advertising in shaping narratives and societal perceptions across different historical periods.

This use case presents a research-oriented workflow designed to support the analysis of historical newspaper advertisements for studies in Contemporary History, Economic Studies, and Social Studies. Building upon historical research previously conducted using traditional methods [8], the proposed workflow leverages AI to automatically detect, extract, transcribe, and describe advertisements, enabling large-scale analyses of historical media landscapes and consumer behavior.

For this use case, the output is a searchable database designed to support the exploration and analysis of historical advertisements, while allowing each extracted advertisement to be downloaded together with its associated metadata. The proposed workflow automatically detects and extracts advertisements before generating descriptions, transcribing their textual content, and classifying them according to product category.

As shown in *Figure 1*, the core processing stages of input, selection, and data harmonization are followed by layout analysis, performed using the LayoutParser_Carlino40 model, and advertisement analysis, performed using Qwen3. All extracted and AI-generated information is then stored in the database, enabling efficient search, filtering, analysis, and download of the extracted advertisements and their associated metadata.

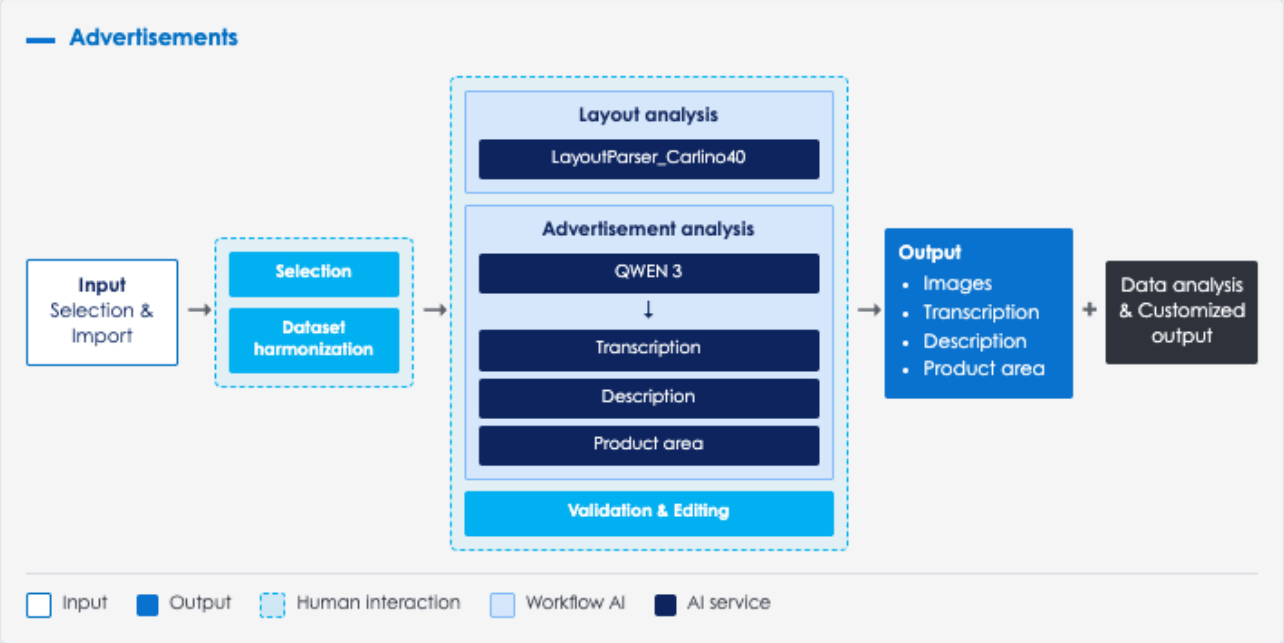


Figure 1: Information extraction workflow designed for the "Advertisement analysis" use case

Use case: Climate Events

Newspapers provide continuous accounts of heatwaves, floods, severe winters, and other weather phenomena, making them valuable resources for environmental and historical research.

This use case examines, from the perspective of climate researchers, the retrieval of historical information on climate events affecting the Italian territory. The objective is to create a dataset of historical weather-related newspaper articles by combining AI-based information extraction and analysis, thereby supporting researchers in reconstructing past meteorological events and investigating how they were represented over time.

For this use case, the output is a searchable database of articles related to climate events together with their associated metadata. The proposed workflow automatically identifies and transcribes newspaper articles. Climate-related articles are then identified through vocabulary matching using a domain-specific lexicon that combines terminology from the *Glossario Meteorologico* of ARPAL [9] and the *Nuovo Soggettario di Firenze* [10]. Since the extraction relies exclusively on vocabulary matching, without contextual analysis, some retrieved articles may contain metaphorical or other non-literal uses of the selected terms. To address this limitation, a Large Language Model is applied to filter the results and retain only articles referring to actual meteorological events.

Figure 2 illustrates the information extraction workflow designed for this use case. As in the previous use case, the core processing stages are followed by a research-specific information extraction workflow. Here, candidate climate-related articles are first identified through vocabulary matching and subsequently filtered using Qwen3 to retain only those whose primary topic is a meteorological event. The resulting articles and their associated metadata are then stored in the database, enabling efficient search, filtering, and analysis.

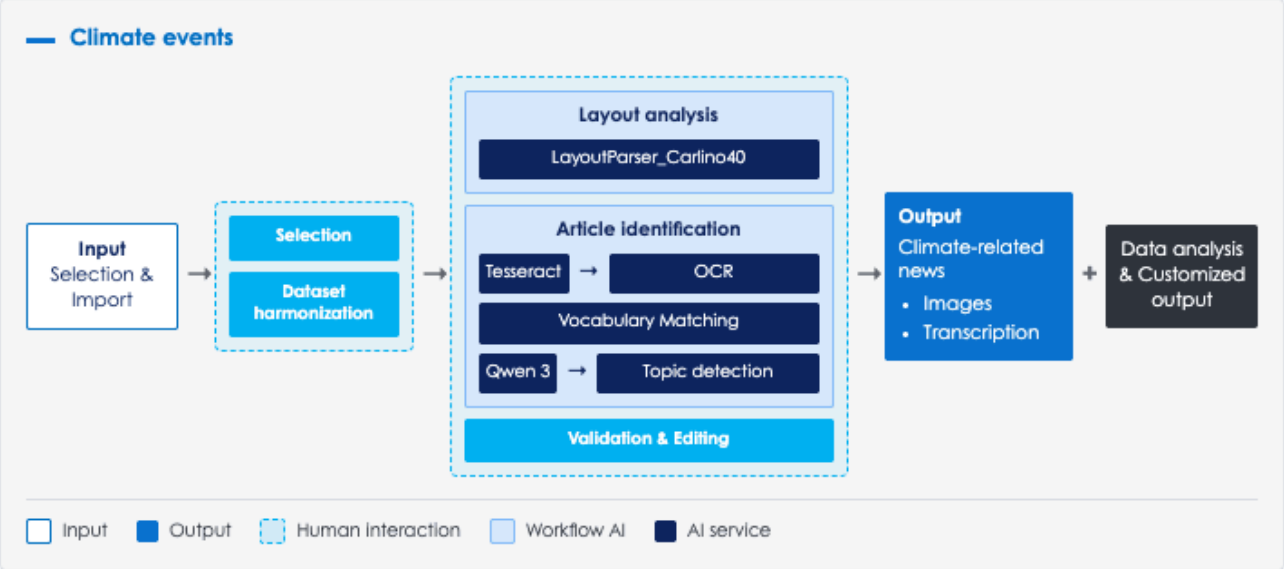


Figure 2: Information extraction workflow designed for the "Climate events" use case

Use case: the term *Piovra*

Historical newspapers provide an ideal source for tracing the construction of social imaginaries through the evolution of language.

This use case investigates, from the perspective of a researcher in Contemporary History, the evolution of the term *piovra* (plural: *piovre*), which entered the Italian lexicon as a calque of the French word *pieuvre*, coined by Victor Hugo in *Les Travailleurs de la mer* (1866), where the octopus acquired a new symbolic meaning associated with evil and oppression. Historical newspapers published between 1866 and 1900 provide an ideal source for tracing this process, as they played a central role in popularizing both the term and its imagery.

As in the previous use cases, the output is a database of articles containing the term *piovra*, together with their associated metadata. The proposed workflow automatically identifies and transcribes newspaper articles. Subsequently, relevant articles are identified through vocabulary matching, considering the target term, its linguistic variants, and common OCR errors, such as *u* instead of *v* or *e* instead of *a*.

Figure 3 illustrates the information extraction workflow designed for this use case. The core processing stages are followed by AI-based layout analysis, after which articles containing the target terms are identified through vocabulary matching.

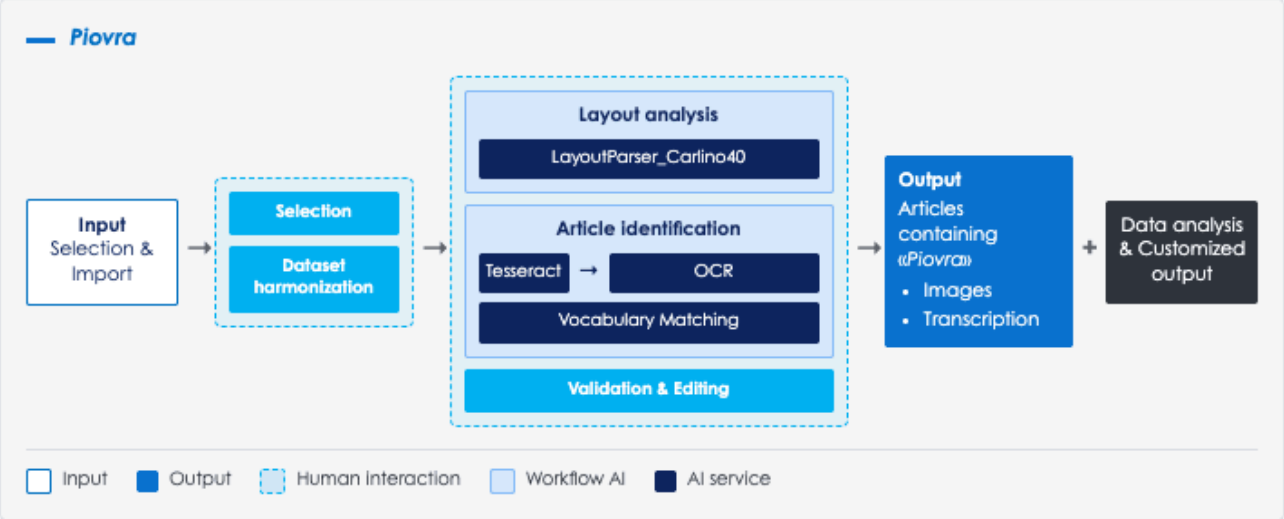


Figure 3: Information extraction workflow designed for the "Piovra" use case

The Layout Parser Workflow Platform and its Components

The data required for fine-tuning Layout Parser comprises a substantial corpus of around 3100 newspaper page images paired with precise layout descriptions. This necessitates the decomposition of the page area—excluding margins—into rectangular regions, which are subsequently classified into one of five distinct categories: Title, Text, Image, Drawing, and Advertisement.
To efficiently produce this ground-truth data, we developed a dedicated web application.

From an architectural standpoint, the application was deployed using the Kubernetes environment installed on **CINECA's ADA Cloud platform**.

Its two main components were saved as **Docker Images** in a **Harbor** private registry deployed on the same platform:

- **Frontend:** Developed with SvelteKit, an open-source component-based framework that drives the user interface
- **Backend:** Developed with Pocketbase, a lightweight open-source backend providing data persistence, user authentication, and custom API endpoints

Kubernetes also handles:

- **Routing:** makes the app accessible and directs communication between frontend and backend through **Ingress** objects;
- **Secrets:** keeps API keys (Harbor, AI services) secure.

The whole architecture was defined and deployed using **Helm**, the official package manager for Kubernetes.

For high-resolution visualization, the frontend integrates **OpenSeadragon**, enabling fluid, tiled navigation of large newspaper scans. Interactive annotation overlays are implemented with **D3.js** to render and manipulate guidelines and rectangular regions directly on top of the page view.

Interface Functionality and Editing Operations

Dataset preparation adheres to an AI-assisted, human-validated workflow which progressively transforms a scanned page into a validated training sample.

Data Input

Data within the platform is structured hierarchically: newspaper titles contain distinct projects, which in turn group individual page scans. Usually, a single project aggregates all newspaper pages published within a specific calendar month. Authorized users can ingest these resources and establish entire project structures directly via the interface tools.

Adding new projects includes a preprocessing step, applied to all selected pages:

- All images are converted to grayscale, keeping as much detail as possible.
- A heuristic algorithm identifies margins, columns, and separators on the page to generate editable guidelines.
- A pre-trained version of LayoutParser, invoked via external APIs hosted on dedicated infrastructure, generates a preliminary set of candidate areas.

Data Preparation

Once a project has been successfully uploaded to the platform, annotators can initiate the layout validation process:

- **Human Guideline Correction:** If the automatic structural guidelines are inaccurate, annotators may manually adjust them to prevent editing errors.
- **Human Region Validation:** Operators review and refine the proposed layout regions, adjusting boundaries, deleting false positives, inserting missed areas, and resolving fragmentation using the split/merge tools.

- **Usability Triage:** Pages with severe geometric constraints or digitization defects (such as severe scan skewing) are explicitly flagged as irregular to ensure they are excluded from model training.

The user interface is optimized for keyboard-and-mouse. Operators can insert a new region with a single click by pressing numerical keys 1 to 5 to select the category, prompting the system to detect the cursor location and automatically snap the boundaries to the nearest guidelines or adjacent rectangles. Supplementary editing capabilities are mapped to the following keyboard shortcuts:

- **Removal (Delete):** Removes the rectangle directly beneath the cursor (supports multi-rectangle selection and batch deletion).
- **Classification change (1–5):** Alters the semantic label of the active rectangle instantly.
- **Merge (j):** Joins two distinct, contiguous rectangles into a single unit.
- **Split (h / v):** Bisects the target rectangle either horizontally or vertically.

Ultimately, these validated annotations can be exported as a structured JSON file mapping page references to their respective coordinate regions and labels.

A case study on printed ads

Conducting research on historical newspapers presents a number of challenges, ranging from the sheer volume of available sources to the multitude of locations where they can be consulted—often in a fragmented manner across various sites. While digitization and the consolidation of materials on unified platforms have addressed the issue of scattered materials, searching through them still requires researchers to physically read the pages, one by one. Identifying useful content has always been an extremely time-consuming task, which—in the case of quantitative analysis—becomes even more so due to the collection and management of relevant data.

In my personal experience, the research I conducted on print advertisements issued in the years of the Italian economic boom—which led to the publication of the essays "La parità si acquista ai grandi magazzini?" [11] and "North and South: Advertising Prosperity in the Italian Economic Boom Years" [8], was based on the laborious creation of a series of spreadsheets compiled entirely by hand. For the daily newspaper *La Gazzetta del Mezzogiorno*, for example, the spreadsheet covering the period from June 1, 1958, to December 31, 1959, consists of 5,360 rows and 5 columns (product or service; publication date; ad size; local or national product or service; descriptive notes). That spreadsheet is accompanied by approximately 920 photographs of advertisements. Each advertisement was also classified according to 17 product categories defined at the outset. Collecting this data, therefore, involved not only physically consulting all the pages but also transcribing the relevant information, a process that was extremely time-consuming. The "size" column—useful for indirectly assessing the advertising investment made by various companies—required measuring every single advertisement. Similarly, the photos of the advertisements then had to be carefully renamed so they could be matched to the specific newspaper and the correct publication date.

The effort required inevitably placed limits on the data collection itself, forcing me to restrict the analysis to a sample of two daily newspapers over a total period of 24 months for each one. It was also decided not to analyze the pages dedicated to classified ads and to consider only those short ads appearing on other pages of the newspaper, limiting the analysis to the roughly ten healthcare services and job ads published daily—likely at a higher cost—in spaces outside the overcrowded classifieds section.

With the application of artificial intelligence solutions, everything changes. The layout parsing system is able to correctly identify the vast majority of advertisements, and the tool also provides the option to correct inaccurate selections. The bounding boxes applied in this way thus enable segmentation, as well as the automatic identification of the advertisement's dimensions. A multimodal LLM, specifically Qwen3, has then been applied to the individual images of each advertisement, using a prompt designed to generate a transcription of the text in the advertisement, a description of it, the identification of the brand promoting goods and services, and their classification within a specific product category. The product categories were

defined using the list of descriptions presented in Liguori [2].

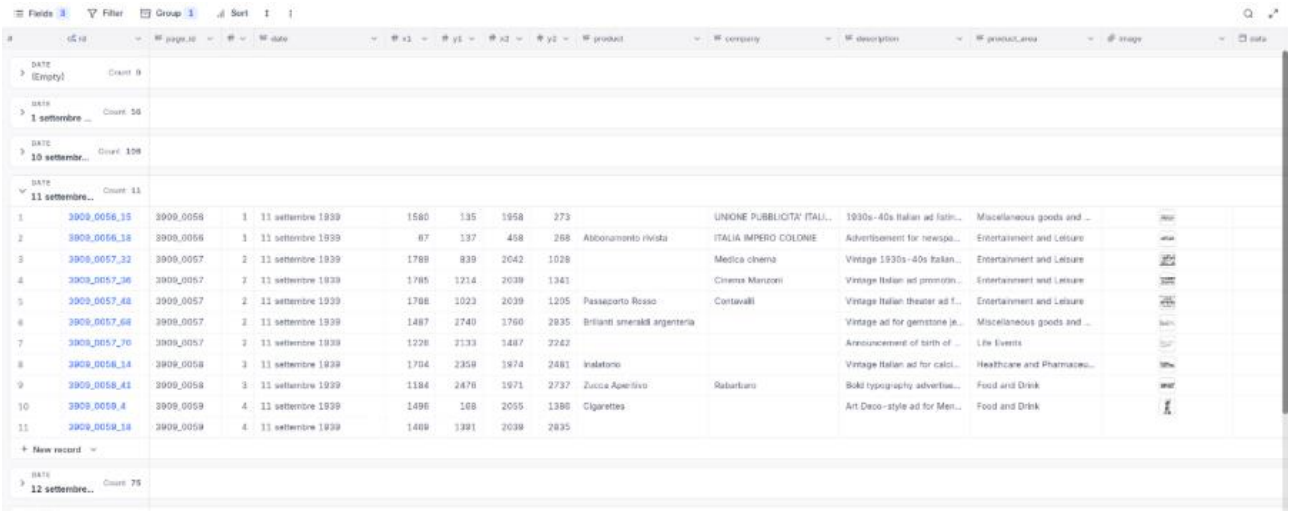


Figure 4 – An aggregate view of ads per date of publication.

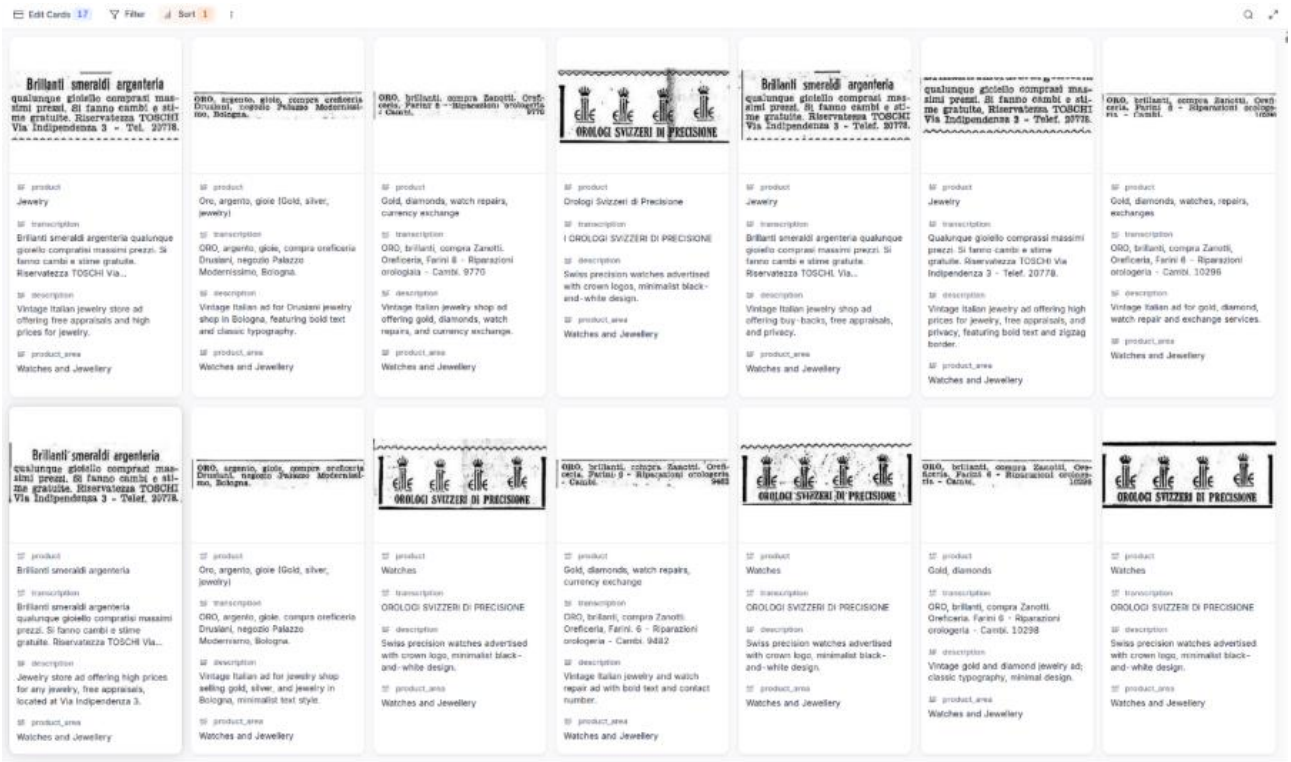


Figure 5 – An aggregated view of ads for Jewelry shops after the elaboration with a set of AI models

The results yielded generally accurate transcriptions of the texts, as well as accurate descriptions of the ad's content. However, the identification of the brand and product category yielded errors in over 50% of cases. Nevertheless, compared to manual data collection, this one involves applying a correction process to an already well-structured and populated spreadsheet; if it is possible to work directly on the original images, the image of the ad is also linked to each row, making the correction process quick and efficient. Furthermore, the majority of errors were detected in ads related to job offers and job searches, since the terms commonly used in such listings caused confusion in the AI model. This means that the other types of advertisements had a lower number of errors, making the correction process far easier. The ability to edit the results through a dedicated interface can

ultimately produce a correct document in a fraction of the time required for manual data collection, facilitating the research and analysis process for the scholar. It is then possible to explore the data in different ways, also visually, by aggregating it into galleries (Fig. 5) or organizing it according to the various fields in which the extracted information has been categorized—such as publication date, product category, or brand, further increasing the amount of correct information obtainable even from a rather large dataset (Fig. 4).

Conclusions

At the moment, LPW_Carlino40 is available in DPaaS as content processing tool and there are already efforts dedicated to the creation of further ground truth in order to train a new version compatible with other layouts of Italian historical newspapers.

References

[1] Antonella Guidazzoli, Maria Chiara Liguori (Eds.), AI e Beni Culturali. Tra ricerca e creatività; AI and Cultural Heritage. Between Research and Creativity Workshop proceedings February 9-10, 2024, CINECA, 2024 ISBN – 9788886037488, 2024, https://doi.org/10.1388/IIIWORKSHOPAIBC

[2] S. Imboden, A. Guidazzoli, G. Cardano, L. Mattei, F. Andrucci, D. Shemek, R. Pansini, F. Albertin, M.C. Liguori, An Open-Source Workflow for Handwritten Character Recognition, in S. Campana, D. Ferdani, H. Graf, G. Guidi, Z. Hegarty, S. Pescarin, and F. Remondino (Editors), DIGITAL HERITAGE (2025), Siena, https://doi.org/10.2312/dh.20253069

[3] S. Imboden, L. Mattei, G. Marconi, F. Andrucci, A. Gianelli, Fine-Tuning LayoutParser for the Analysis of Historical Italian Newspapers, in S. Campana, D. Ferdani, H. Graf, G. Guidi, Z. Hegarty, S. Pescarin, and F. Remondino (Editors), DIGITAL HERITAGE (2025), Siena, https://doi.org/10.2312/dh.20253275

[4] S. Imboden, G. Marconi, S. Caraceni, R. Pansini, F. Albertin, A. Guidazzoli, Layout Parser, come creare un dataset di qualità per allenare l'Intelligenza Artificiale, in Frontiere Digitali. Consortium GARR (2025), pp. 101-106. https://doi.org/10.26314/GARR-Conf25-proceedings-18

[5] SHEN Z., ZHANG R., DELL M., LEE B. C. G., CARLSON J., LI W.: Layout parser: A unified toolkit for deep learning based document image analysis. https://arxiv.org/abs/2103.15348, 2021. doi:10.48550/ARXIV.2103.15348.

[6] S. Imboden, A. Guidazzoli, Sforzini D., Mattei L., Marconi G., Andrucci F., Gianelli A., Caraceni S., Pansini R., Fatigati G., Albertin F., Baisi G., Cardano G., Liguori M. C., Lanzarini A.: *An Open-Source solution for layout parsing of historical Italian newspapers*. In Proceedings of 22nd Conference on Information and Research Science Connecting to Digital and Library Science - IRCDL 2026 (2026). In Press.

[7] Smith R., *An overview of the tesseract ocr engine,* In ICDAR '07: Proceedings of the Ninth International Conference on Document Analysis and Recognition (Washington, DC, USA, 2007), IEEE Computer Society, pp. 629–633.

[8] Liguori, M. C. (2015). *North and South: Advertising Prosperity in the Italian Economic Boom Years.* Advertising & Society Review, 15(4).

[9] REGIONALE PER LA PROTEZIONE DELL'AMBIENTE LIGURE, A.: Nuovo Glossario Meteorologico Arpal. https://www.arpal.liguria.it/tematiche/meteo/glossario-meteo-arpal.html

[10] BIBLIOTECA NAZIONALE CENTRALE DI FIRENZE: Nuovo Soggettario di Firenze. https://thes.bncf.firenze.sbn.it.

[11] M.C. Liguori, *La parità si acquista ai grandi magazzini? Boom economico e trasformazione del modello femminile*, in P. Capuzzo (a cura di), *Genere, Generazione, Consumi. L'Italia degli anni Sessanta*, Annali Fondazione Istituto Gramsci, n. XII, Roma, Carocci Ed., 2003, https://doi.org/10.1400/31003

CHAPTER 05

THE "OCTOPUS" PROJECT

Roberto Balzani - University of Bologna; Elena Musiani - University of Bologna / Université de Paris Nanterre

Proposed itineraries concerning heritage resources often focus on the development of analogies, references and connections, in keeping with the principle that digital heritage – in the cultural sphere – is, above all, dynamic rather than static. The potential for expanding connections and associations bears witness to the breadth of figurative languages, cultural influences and symbolic revivals, not only in the realm of art and imagery, but also in written documents and historical records. Where such proposals are developed with the greatest care, the results genuinely expand the range of cases documented by the sources. However, the research question, for the time being, remains, so to speak, 'traditional' or, at most, demonstrative and illustrative.

The 'Octopus' project – that is, the study of the spread of the term, which originated in France in 1866 and subsequently became established in Italian translation in the years that followed – represents a potentially significant shift because – contrary to what has been stated above – without digital data, this research, concerning the metamorphosis of the symbol and its gradual, almost magnetic acquisition of 'evil' meanings, would be unthinkable. It was, of course, already known that these concepts existed; how they had become established, however, was not. This is also because, unlike the origin of the word, we are not dealing here with a brilliant creative mind (Victor Hugo), but with a usage that spread virally, mediated by the press, partly through imitation (from France) and partly through spontaneous generation, based on associations of ideas whose context we can now identify but whose authorship is rarely known. In a later phase, during the 20th century, further contributions to its meaning gave way to the standardisation and simplification of contexts, alongside the term's growing popularity and its now firmly established place in the lexicon of the collective imagination.

Thanks to CINECA and the team led by Antonella Guidazzoli, it has been possible to examine this development on a large scale, using Il Corriere della Sera for the period when the term 'piovra' first appeared, and then Il Popolo d'Italia for the period of 'normalisation' and the political spread of the term in the 20th century (by the Fascists, for example). The research was made possible not only by identifying the frequency of the term, but also by reconstructing the precise context in which it was written.

This represents a potentially remarkable qualitative leap: the 'octopus' moves, so to speak, through the printed page, absorbing meanings; it is not (yet) directed by some external source, and therefore the analysis of the contexts it ultimately influences is, in fact, purely and simply identical to its precise meaning. If we disregard this meticulous work, we merely end up repeating clichés: yet what this workshop focuses on is, instead, the emergence and development of a cultural process of agglutinating meanings, aimed at enriching an imaginary that transforms the way we perceive and represent social, political and economic phenomena, in a sense 'radicalising' them – that is, emphasising their extreme manifestations for a wide variety of reasons. Why? Now we can start to grasp it.

The research question underlying this case would not even have been conceivable without the collaboration between analysts of large, coherent data sets and scholars interested in developing the heuristic potential of the digital humanities. From this perspective, the 'octopus' can be regarded as a pioneering experiment.

CHAPTER 06

NEW QUESTIONS FOR CULTURAL HERITAGE: SOURCES, FORMATS AND INFRASTRUCTURES IN THE AGE OF ARTIFICIAL INTELLIGENCE - Panel

Elena Musiani – University of Bologna, Angelo Pompilio – University of Bologna, Elena Cervellati – University of Bologna, Giuseppe Di Modica – University of Bologna, Mirko Degli Esposti – University of Bologna

Giovanni Baisi (moderator) - CINECA

Abstract

This chapter reports on the panel of the workshop "AI, Cultural Heritage, Art and Science", which asked five scholars of the University of Bologna not what artificial intelligence can already do, but what research questions can be asked once large digitised corpora and generative models are available. Four problems recurred across otherwise distant disciplinary positions: data quality, translation between formats, the silences of the sources, and the governance of infrastructures.

Introduction: siding with the questions

The morning session had been devoted largely to digitised newspapers, through two use cases: the semantic history of the term *piovra* in the Italian press, presented by Roberto Balzani, and the analysis of historical newspapers, presented by CINECA. Balzani said that this work puts researchers "on the side of the questions", and the panel started from there. The panel was therefore asked not to assess tools, but to identify the questions that become formulable when the object of enquiry is no longer only text: spoken and musical audio, audiovisual material, images, tangible and intangible heritage.

New questions, new networks

Elena Musiani opened from the perspective of the *Piovra* project, shifting the discussion from method to attitude. Reporting on a teaching experience, she observed that the generation that is often assumed to be the most receptive to these technologies had expressed a particular concern: the fear of becoming useless. Pressed on this point from the online audience, she reframed the issue: what is feared is not the loss of curiosity but the loss of a meaningful role. Far from rendering researchers unnecessary, these tools make possible investigations such as tracing the semantic evolution of a single lemma across a vast corpus of periodical publications, a task that would otherwise be practically impossible.

Her central argument concerned the kind of infrastructure that should be developed. The *Piovra* project had immediately prompted proposals from neighbouring institutions, with colleagues at the Istituto Gramsci drawing attention to their digitised collection of political posters. The goal is not simply to create a repository of seventy-five million digitised items, but to develop tools, systems, and an interdisciplinary ecosystem in which expertise generates questions, and questions, in turn, generate new lines of inquiry. She concluded with a cautionary remark: these systems must be trained to handle languages and categories that regarding gender, for example, are not merely binary, and metadata

practices are the point at which this requirement must be addressed.

Verbalising sound: librettos

Angelo Pompilio approached the corpus of nineteenth-century newspapers from a musicological perspective. Music criticism emerged in the 1830s and 1840s, and the critic's task was to render sound in language: describing an event which, in the Romantic conception, was valued precisely for expressing the inexpressible. The research question is therefore both lexical and rhetorical, requiring a broader documentary base that includes theatrical almanacs, publishers' catalogues, and the illustrated press.

Pompilio's proposal focused on aligning verbal and visual description. The work drew on a twenty-year compilation of a repertory of Italian melodrama from primary sources, chiefly librettos, undertaken on the premise that such systems are effective only when supported by high-quality data. Until the end of the nineteenth century, stage directions in librettos were essentially prescriptive, with scenographers expected to realize the instructions contained in the text. He has transcribed the stage directions of some 8,000 librettos, producing a corpus of approximately 40,000 descriptions. Museums, meanwhile, preserve the corresponding scenic sketches, whose catalogue descriptions do not always make their content readily identifiable. The proposed experiment is to relate these textual descriptions to iconographic features extracted from the images, testing whether the verbal record can be used to identify the corresponding visual material.

Saying the difficult to say: the Choreic Arts

Elena Cervellati extended the discussion to dance, considering not only performances but also the languages of movement themselves. Her case study focused on intercultural translation: tracing *Giselle* from its French origins to contemporary Italian versions reveals translations that are not merely linguistic, but also involve form, sensibility, and audience expectations. Ballet thus offers a clear example of a cultural form that endures precisely through change.

Two difficulties follow. The first is documentary. Writing on dance occupies a marginal place in the historical record: theoretical texts are already relatively scarce, while periodical writings are, as Cervellati noted, "micro-minor". As a result, digitisation has proceeded unevenly, and optical character recognition introduces further distortions.

The second difficulty concerns the mode of enquiry. Keyword searches alone are insufficient; what is needed is a conceptual interrogation of historical corpora, supported by the automatic identification of themes and lexical clusters associated with complex concepts. She concluded by pointing to two promising directions for future research: correspondence can help recover the voices of women spectators, largely absent from the periodical press, while visual materials such as sketches, lithographs, and portraits preserve part of dance's intangible heritage.

What the machines do, and what they do not do

Drawing on his involvement in the design of Ecomic, Giuseppe Di Modica responded to the concern raised at the opening of the panel. Current systems, he argued, lack the capacity for self-reflection and cannot question the directions that scientific inquiry should take. Large language models return answers that are statistically plausible, which is not the same as true. They operate on tokens rather than on any understanding of meaning or syntax. The research process (intuition, data collection, literature, hypothesis, experiment, validation) remains structurally unchanged, and what changes is the duration of some of its phases. Optical recognition performs strongly on print and weakly on handwriting, since humanistic script varies so widely that a model trained on one hand fails on another; the remedy is not a better model but a human who establishes the ground truth by transcribing and supplies it as an example.

Ecomic was designed as a democratic tool, allowing non-specialists to ingest data, configure processing pipelines, and extract the information they need. The wider asymmetry, however, is infrastructural: Europe

lacks major AI models of its own, with European data flowing instead towards those who use them to improve their models. He presented a platform of this kind, capable of growing incrementally in both services and data, as a condition for a European cultural sovereignty worth defending.

The city as Cultural Heritage

Mirko Degli Esposti acknowledged the partiality of his perspective: as a mathematical physicist working on urban complexity, he regards cultural heritage primarily as the built environment. As an example of this line of work, he described an experiment that draws on a local daily newspaper to reconstruct the climatic history of a city and its responses to extreme events. Local newspapers (his experiment uses *il Resto del Carlino* for Bologna, although the approach is transferable to any comparable publication) document floods, snowfalls, roof collapses, and the transformation of infrastructure and green spaces. The information they contain is not only semantic but also geographically and temporally situated.

Urban data, unlike the more structured holdings of libraries or music collections, are largely unstructured and combine a wide range of modalities, including sketches, cadastral records, LiDAR data, orthophotos, satellite imagery, photographs, video recordings, and survey responses. Current methods, he suggested, make it possible to extract information directly from this heterogeneous material, rather than requiring the prior normalisation of each source. Since textual semantics and visual information can be represented within a common latent space, detailed urban contexts may likewise be generated from such partial inputs.

His final point, offered as a provocation, concerned not the conservation of heritage but its consumption. Synthetic populations, artificial yet statistically representative populations used in simulation studies, have existed since the 1990s, yet large language models have introduced a qualitative leap: it is now possible to generate geolocated, socio-demographically detailed, and entirely artificial subpopulations defined by numerous parameters. Sixty thousand elderly holders of a museum card; forty-three thousand students distributed across a city: fictitious, yet realistic. By attaching a language model to each synthetic individual, one could, in principle, send them to a museum and have them complete questionnaires without burdening real citizens. He acknowledged that the proposal unsettles him as much as it excites him.

Conclusions

Four problems recurred:

- Data quality proved the binding constraint, converging Pompilio's insistence on good data with Di Modica's account of manual transcription: what is scarce is curated, scholarly-grade material, not modelling capacity.
- Translation between formats was the common object: sound into words, movement into description, semantics into image.
- The silences of the sources (unrepresented voices, minor literatures, binary metadata categories) mean that a corpus reproduces its own exclusions at scale.
- Governance asks who holds the data and the models.

The panel's most lasting contribution was methodological. Digitisation is a precondition rather than an achievement: what confers value on it is the construction of tools, of networks and of an ecosystem in which competences generate further questions. Against the fear of displacement, the panel returned an argument about the division of labour. These tools shorten some phases of research without removing any: intuition, hypothesis and validation remain human, and the phases the tools do enable demand more disciplinary work rather than less, since someone still has to establish the reference data on which the models depend, and someone has to curate and interpret the corpus. The panel accordingly produced no verdict on artificial intelligence, but working hypotheses that can be tested, from verbal-to-iconographic alignment for melodrama to the climatic history of the city extracted from the local press. The challenge is not to take sides for or against AI, but to turn such questions into research problems that can be tested.

CHAPTER 07

INTRODUCTION TO THE AFTERNOON SESSION: INTANGIBLE CULTURAL HERITAGE

Alberto Salarelli – University of Parma, Francesca Zanella - University of Modena and Reggio Emilia

Since the adoption of the 2003 UNESCO Convention, intangible cultural heritage has become the focus of a growing body of research addressing the relationships between cultural practices, their transmission and the institutions involved in their safeguarding. The handbook edited by Christoph Wulf in 2025 [1] brings together many of the questions that have shaped this field: how living practices can be recognised and safeguarded, what role museums and other institutions can play, and how communities participate in the transmission of their heritage. It also draws attention to the connections between these practices and wider concerns, including education, sustainable development and the effects of cultural activity on local economies.

The five domains identified by the Convention—oral traditions, performing arts, social practices and rituals, knowledge concerning nature, and traditional craftsmanship—suggest the breadth of the subject. They also remind us that intangible heritage cannot be separated from the people who practise it. Knowledge may pass through words, but also through movement, observation and shared experience. In each case, transmission allows a practice to continue while leaving room for it to change.

This living character has consequences for the institutions that work with intangible heritage. A museum may collect objects associated with a craft or a performance, while an archive may preserve photographs, recordings and written accounts. Such materials are indispensable, but their meaning depends on the practices and relationships from which they emerged. Safeguarding therefore involves more than keeping a record: it requires opportunities for people to practise, teach and reinterpret what they know. It also calls for attention to whose voices are heard when a tradition is documented or presented to the public. Communities are participants in these decisions, although their members may hold different views about how a practice should be represented or shared.

Digital technologies have expanded both the possibilities and the challenges associated with intangible cultural heritage. They can document gestures, connect dispersed sources and bring cultural practices to audiences who might otherwise have little opportunity to encounter them. Generative artificial intelligence calls for particular attention because it can produce new images, movements and representations from existing cultural materials. As discussions at previous workshops have shown, these developments require us to reconsider what it means to preserve, interpret and communicate a heritage whose vitality lies in continued practice. In this perspective, digital documentation is not simply a means of preserving intangible cultural heritage, but a form of mediation that shapes how cultural practices are represented, interpreted and made accessible. The contributions to this afternoon session address these questions from legal, artistic, archival and public perspectives.

Valentina Gastaldo examines the legal implications of treating intangible cultural heritage as living heritage. Her contribution considers the distance between the approach of the UNESCO Convention and legal frameworks developed largely around material objects. Through examples of traditional practices and performing arts, she explores questions of digitisation, artificial intelligence, cultural appropriation and the participation of communities in decisions about their heritage. Her analysis establishes an essential point for the discussion that

follows: technological possibilities need to be considered alongside the rights and interests of the people whose practices they represent.

The next two contributions focus on dance. Fabio Ferretti presents the work of the Centro Coreografico Nazionale/Aterballetto, where virtual reality has introduced new forms of engagement with the relationship between the dancing body, the spectator and the places in which performances are created. The "Double Reality" project and the DANSITE platform extend this exploration across cultural sites, while DUETTO investigates the possibilities of generative artificial intelligence in dialogue with human movement. Silvia Garzarella approaches ballet from a complementary perspective. Through DanXe, she examines how digital methods can bring together documents, costumes, audiovisual recordings and traces of movement, allowing researchers to study the relationships among sources and the transformations of ballet over time.

Finally, Elena Radaelli considers how an archive can communicate a performance tradition to audiences who did not experience it directly. Her discussion of "Marionette 3.0", developed around the archive of the Teatro delle Marionette di Gianni e Cosetta Colla, brings together data visualisation, storytelling and activities designed for younger audiences. The project asks how surviving materials can lead people towards an understanding of puppetry while acknowledging that no archive can reproduce the live event.

Across these contributions runs a shared concern with the relationship between cultural practices and the traces, descriptions and digital experiences created around them. Such forms of mediation can deepen research, expand access and create new opportunities for participation. Yet they also involve choices about what is documented, how it is represented and whose perspectives are taken into account. The contributions gathered in this session suggest that safeguarding intangible cultural heritage requires attention not only to the preservation of its traces, but also to the changing practices, relationships and contexts that give those traces meaning.

References

[1] Christoph Wulf (Ed.), Handbook on Intangible Cultural Practices as Global Strategies for the Future: Twenty Years of the UNESCO Convention on Safeguarding Intangible Cultural Heritage. Springer; 2025 ISBN – 9783031721229, https://doi.org/10.1007/978-3-031-72123-6

CHAPTER 08

INTANGIBLE CULTURAL HERITAGE AND NEW TECHNOLOGIES: A LEGAL PERSPECTIVE

Valentina Gastaldo – University of Parma

Abstract

This paper examines the tension between the notion of intangible cultural heritage, developed by the 2003 UNESCO Convention around the idea of "living heritage", and an Italian legal system still largely structured around tangible cultural property. After outlining the domestic regulatory evolution—from Article 7-bis of the Cultural Heritage Code to Law No. 152/2024—the paper considers three paradigmatic cases (the art of Neapolitan pizzaiuoli, traditional Cremonese violin craftsmanship, and Sicilian puppet theatre) and analyses the principal legal issues raised by new technologies: copyright, cultural appropriation and economic exploitation, digitisation, and artificial intelligence, with particular attention to cultural authenticity. It concludes that the role of law is not to restrain technology, but to govern its use in a manner consistent with the dynamic, relational, and community-based nature of intangible heritage.

Introduction: A Change of Paradigm

The relationship between intangible cultural heritage and new technologies is one of the most complex fields for legal analysis because it exposes the limits of categories traditionally developed for tangible cultural property. The starting point is the 2003 UNESCO Convention for the Safeguarding of the Intangible Cultural Heritage, ratified by Italy through Law No. 167, 27 September 2007 [1]. Article 2 defines intangible heritage as practices, representations, expressions, knowledge and skills that communities recognise as part of their cultural heritage, while Articles 11 and 15 assign communities an active role in its identification, safeguarding and transmission.

The Convention therefore introduces a fundamental change of paradigm: rather than a static object to be preserved, intangible heritage is understood as living heritage. Italian law, however, remains largely centred on the tangible cultural object. Article 7-bis of Legislative Decree No. 42/2004 recognises collective cultural identities covered by the UNESCO Conventions of 2003 and 2005, but extends the Code's protection primarily to their tangible manifestations when the requirements of Article 10 are met [2; 3; 4; 5; 6; 7; 8]. Protection therefore remains indirect and mediated.

Towards a Comprehensive Legal Framework? Law No. 152/2024

Law No. 44/2017 had already extended specific safeguarding measures under Law No. 77/2006 to Italian elements included in UNESCO's Representative List [9]. A more significant development came with Law No. 152 of 7 October 2024, which expressly recognises intangible cultural heritage as an identity-bearing and historically valuable component for individuals, local communities and the national community, and delegated the Government to adopt a comprehensive legal framework [10; 11; 3; 12]. The delegation period expired without the adoption of the delegated decrees. The promised comprehensive framework therefore remains unrealised.

The central difficulty is to avoid reproducing, for living practices, a "musealising" model designed for tangible objects. A further issue concerns coordination with regional legislation and registers,

which have developed unevenly over time. The question therefore remains whether—and through which instruments—positive law can protect heritage that cannot, by its nature, be reduced to things.

Three Paradigmatic Cases

The art of Neapolitan pizzaiuoli, inscribed on UNESCO's Representative List in 2017 [13], illustrates the distance between cultural recognition and exclusive legal protection. What is safeguarded is not a recipe as such, but a complex body of skills, gestures and practices transmitted within a community. Such heritage cannot readily be attributed to an individual author or protected through conventional copyright mechanisms.

Traditional Cremonese violin craftsmanship, inscribed in 2012 [14], presents a related problem. The practice is collective and intergenerational, while associations and public institutions can organise safeguarding and promotion without thereby creating an exclusive proprietary right over the traditional know-how.

Sicilian puppet theatre (Opera dei pupi), proclaimed by UNESCO in 2001 and included in the Representative List in 2008 [15], is particularly useful for introducing the technological dimension. Digital archives and augmented-reality initiatives can facilitate access and transmission, but they also reveal a central risk: the digital representation may come to be perceived as a substitute for the living practice that it was intended to document [16].

The Impact of New Technologies

Digital technologies have profoundly changed the documentation, transmission and enjoyment of intangible heritage. Digitisation can preserve oral traditions, endangered languages, rituals and traditional knowledge. International projects such as the UNESCO-category-2 ICHCAP centre and its ichLinks platform show how audiovisual digitisation can itself become a safeguarding instrument when supported by an appropriate institutional infrastructure [17]. European initiatives concerning cultural heritage data and projects, such as AlpiLinK, likewise demonstrate the growing importance of shared digital infrastructures [18].

Artificial intelligence adds a further dimension because it can not only preserve and organise cultural materials but also reinterpret and generate them. Digital archives created for safeguarding may themselves become training datasets for generative systems. Technology is therefore not neutral: fixing a living practice in digital form necessarily selects one version, removes it partly from its original context and may transform a contingent performance into an apparently authoritative representation.

Legal Issues

Intangible Cultural Heritage and Copyright

Copyright protects creative expression attributable to an author, whereas intangible heritage is often collective, anonymous and intergenerational. A recording of a traditional song may attract protection as a specific fixation or performance, while the underlying tradition may remain outside copyright protection. Italian law contains no comprehensive special regime for folklore or traditional cultural expressions. International discussions within WIPO on Traditional Cultural Expressions therefore remain particularly relevant [19; 20].

Cultural Appropriation and Economic Exploitation

"Cultural appropriation" does not correspond to a single legal category. It describes heterogeneous uses of cultural elements outside their communities of origin, sometimes without recognition, participation, or benefit-sharing. Existing instruments—copyright, unfair competition, geographical indications and contractual arrangements—can address specific situations but do not establish a general community right over a cultural practice. EU geographical-indication regimes protect names, reputation and product characteristics rather than the cultural practice itself [21; 22; 23; 24].

Digitisation: Between Safeguarding and Transformation

Digitisation preserves memory while simultaneously fixing it in a particular form. The legal problem is therefore not merely the preservation of files, but the governance of selection, access, reuse, and contextualisation. Community participation becomes crucial: documentation should not turn a living tradition into an immutable canonical version, nor

should digital accessibility automatically entail unrestricted reuse.

Artificial Intelligence and Authenticity

Artificial intelligence makes this tension particularly visible. Italian Law No. 132/2025 and Regulation (EU) 2024/1689 regulate aspects of artificial intelligence, but they do not independently protect the cultural relationship between training data and the community from which those data derive [25; 26]. Likewise, the Italian provisions on text and data mining operate within copyright law and do not grant communities a general power to authorise the use of unprotected traditional expressions [27].

The deeper issue is authenticity. In intangible heritage, authenticity cannot mean fidelity to an immutable original form, since living practices are continuously recreated. It refers instead to the continuity of the relationship between the practice and the community that recognises, transmits, and renews it. An algorithm may reproduce stylistic or musical patterns, but it cannot by itself establish the cultural meaning or community belonging of the result.

Some Indications from Constitutional and European Case Law

There is not yet a consolidated body of case law specifically addressing intangible cultural heritage and new technologies. Italian constitutional case law nevertheless illustrates the limits of a framework traditionally centred on tangible cultural property. In Judgment No. 118/1990, the Constitutional Court considered the historical and cultural significance of traditional activities in relation to the material objects in which that significance was embodied [8]. Subsequent judgments concerning the protection and enhancement of cultural heritage and the constitutional status of landscape protection reaffirmed the importance of cultural values, without establishing an autonomous legal regime for intangible heritage [28]. At European level, cultural heritage interests may also arise within broader human-rights frameworks, although no general judicial doctrine specifically governing digitally mediated intangible heritage has emerged [29].

The Role of Law

The function of law is not to impose a general prohibition on technology, but to govern its use consistently with the nature of intangible cultural heritage. This requires legally meaningful participation by heritage communities, procedures capable of identifying legitimate representatives and managing internal plurality, and instruments that go beyond the proprietary model of copyright. Transparency, data protection, cultural contextualisation and, where appropriate, benefit-sharing should accompany digital access and technological innovation.

Conclusions

The relationship between intangible cultural heritage and new technologies reveals a structural tension between innovation and safeguarding, dissemination and cultural identity, accessibility and simplification. Digital technologies can broaden access, support research and strengthen community memory; they can also crystallise dynamic practices, decontextualise community-based meanings and generate formally plausible simulations lacking cultural authenticity.

The central question is not, therefore, whether technology should be used, but under what legal, ethical and institutional conditions. Transparent procedures for community participation, specific protocols for digitisation and the use of cultural materials in artificial-intelligence systems, and duties of traceability and information can help reconcile innovation with safeguarding. The ultimate task of law is to make technological innovation compatible with heritage that exists through practice, memory and transmission.

References

[1] UNESCO, *Convenzione per la salvaguardia del patrimonio culturale immateriale*, Parigi, 17 ottobre 2003; ratifica italiana: l. 27 settembre 2007, n. 167.

[2] M. Fiorillo, *Dalle cose alle persone: declinazione e scenari del patrimonio culturale immateriale*, in federalismi.it, n. 20/2026.

[3] V. Gastaldo, *L'insostenibile leggerezza del patrimonio culturale immateriale*, in Aedon, n. 1/2025.

[4] A.L. Tarasco, *Diversità e immaterialità del patrimonio culturale nel diritto internazionale e comparato: analisi di una lacuna (sempre più solo) italiana*, in Amministrazione in Cammino, 11 dicembre 2008.

[5] A.L. Tarasco, *Il patrimonio culturale: concetto, problemi, confini*, Napoli, 2019.

[6] C. Barbati, M. Cammelli, L. Casini, G. Piperata, G. Sciullo, *Diritto del patrimonio culturale*, Bologna, 2017.

[7] Art. 7-bis, d.lgs. 22 gennaio 2004, n. 42, inserito dall'art. 1, comma 1, lett. c), d.lgs. 26 marzo 2008, n. 62 [Art. 7-bis of Legislative Decree No. 42 of 22 January 2004, inserted by Art. 1(1)(c) of Legislative Decree No. 62 of 26 March 2008].

[8] Corte cost., sent. 6 marzo 1990, n. 118 [Italian Constitutional Court, Judgment No. 118 of 6 March 1990].

[9] L. 8 marzo 2017, n. 44, recante modifiche alla l. 20 febbraio 2006, n. 77 [Law No. 44 of 8 March 2017, amending Law No. 77 of 20 February 2006].

[10] L. 7 ottobre 2024, n. 152, "Disposizioni in materia di manifestazioni di rievocazione storica e delega al Governo per l'adozione di norme per la salvaguardia del patrimonio culturale immateriale" [Law No. 152 of 7 October 2024, Provisions concerning historical re-enactment events and delegation to the Government for the adoption of rules safeguarding intangible cultural heritage - unofficial translation], Official Gazette (G.U.), No. 244, 17 October 2024, Arts. 10-11.

[11] On regional disciplines, see, among others: l.r. Liguria 2 maggio 1990, n. 32 [Regional Law of Liguria No. 32 of 2 May 1990]; l.r. Molise 11 aprile 1997, n. 9 e 5 maggio 2005, n. 19 [Regional Law of Molise No. 9 of 11 April 1997 and Regional Law of Molise No. 19 of 5 May 2005]; d.ass. Regione Sicilia 26 luglio 2005, n. 77 [Decree of the Regional Councillor of Sicily No. 77 of 26 July 2005]; l.r. Lombardia 23 ottobre 2008, n. 27 [Regional Law of Lombardy No. 27 of 23 October 2008]; l.r. Puglia 22 ottobre 2012, n. 30 [Regional Law of Apulia No. 30 of 22 October 2012].

[12] A. Gualdani, *La legge in materia di manifestazioni di rievocazione storica*, in Aedon, n. 3/2024.

[13] UNESCO, 12th session of the Intergovernmental Committee for the Safeguarding of the Intangible Cultural Heritage, Jeju, 7 December 2017, inscription of the Art of Neapolitan Pizzaiuolo on the Representative List of the Intangible Cultural Heritage of Humanity.

[14] UNESCO, 7th Session of the Intergovernmental Committee for the Safeguarding of the Intangible Cultural Heritage, Paris, 5 December 2012, inscription of Traditional Violin Craftsmanship in Cremona on the Representative List of the Intangible Cultural Heritage of Humanity; see also the Safeguarding Plan and related institutional materials available at unesco.it and liuteriacremonese.it.

[15] UNESCO, Sicilian Puppet Theatre (Opera dei Pupi), proclaimed a Masterpiece of the Oral and Intangible Heritage of Humanity in 2001; inscribed on the Representative List of the Intangible Cultural Heritage of Humanity in 2008.

[16] On #Carinda A.R. project - Museo internazionale delle marionette Antonio Pasqualino, see E. Sabbatini, *L'Opera dei Pupi entra nel futuro: in Sicilia i paladini rivivono tra AR, archivi digitali e installazioni immersive*, in Be Sicily Mag, 2 April 2026.

[17] International Information and Networking Centre for Intangible Cultural Heritage in the Asia-Pacific Region (ICHCAP), program for the restoration and digitization of audiovisual resources.

[18] Commission Recommendation 2011/711/EU of 27 October 2011; Commission Recommendation (EU) 2021/1970 of 10 November 2021 on a common European data space for cultural heritage.

[19] Cass. civ., sez. I, 12 marzo 2004, n. 5089 [Italian Court of Cassation, Civil Division I, Judgment No. 5089 of 12 March 2004].

[20] WIPO, Treaty on Intellectual Property, Genetic Resources and Associated Traditional Knowledge, Geneva, 24 May 2024; for Traditional Cultural Expressions, see also the work of the Intergovernmental Committee on Intellectual Property and Genetic Resources, Traditional Knowledge and Folklore.

[21] Regulation (EU) 2024/1143 of the European Parliament and of the Council of 11 April 2024.

[22] Regulation (EU) 2023/2411 of the European Parliament and of the Council of 18 October 2023.

[23] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information.

[24] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (GDPR).

[25] L. 23 settembre 2025, n. 132, Disposizioni e deleghe al Governo in materia di intelligenza artificiale [Law No. 132 of 23 September 2025, Provisions and Delegations to the Government concerning Artificial Intelligence].

[26] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (AI Act).

[27] Artt. 70-ter e 70-quater l. 22 aprile 1941, n. 633, introdotti dal d.lgs. 8 novembre 2021, n. 177, di attuazione della direttiva (UE) 2019/790 [Arts. 70-ter and 70-quater of Law No. 633 of 22 April 1941, introduced by Legislative Decree No. 177 of 8 November 2021 implementing Directive (EU) 2019/790].

[28] Corte cost., sent. 16 gennaio 2004, n. 9 [Italian Constitutional Court, Judgment No. 9 of 16 January 2004]; Corte cost., sent. 30 ottobre 2007, n. 367 [Italian Constitutional Court, Judgment No. 367 of 30 October 2007].

[29] European Court of Human Rights, decision of 29 January 2019, Ahunbay and Others v. Turkey, Austria and Germany, Application No. 6080/06.

CHAPTER 09

DANCE, VIRTUAL REALITY AND CULTURAL HERITAGE: THE DOUBLE REALITY PROJECT OF THE CENTRO COREOGRAFICO NAZIONALE / ATERBALLETTO

From the dialogue between body and place to the dialogue between body and artificial intelligence

Fabio Ferretti – *General Coordinator and Development PM, Centro Coreografico Nazionale/Aterballetto*

A National Choreographic Centre, between stage and experimentation

The Centro Coreografico Nazionale/Aterballetto is the first Centro Coreografico Nazionale established in Italy, recognised by the Ministry of Culture in 2022. It grew out of the historic Aterballetto company, founded in 1977 in Reggio Emilia — the first stable dance company in Italy outside the opera foundations (Fondazioni Liriche) — which became the Fondazione Nazionale della Danza in 2003, with the Emilia-Romagna Region and the Municipality of Reggio Emilia as founding members. Its home is the Fonderia, an early twentieth-century industrial space once used for metal casting and today a creative crucible housing rehearsal rooms, a costume workshop and offices: a place that carries in its very name and history the idea of the transformation of matter, not far removed from the transformation of the body and of image that today guides the Foundation's research projects.

Alongside the production of performances with internationally acclaimed choreographers — among others Marcos Morau, Crystal Pite, Johan Inger, Diego Tortelli, Francesca Lattuada, Fernando Melo and Philippe Kratz — CCN/Aterballetto has for years cultivated an experimental vocation towards formats, contexts and bodies that do not follow the usual conventions of age, gender and ability. Within this scope lies the work on new technologies applied to dance, developed along two complementary strands: on one hand virtual reality, explored since 2020 and now consolidated in the DANSITE platform; on the other, generative artificial intelligence, investigated together with the Sony Computer Science Laboratories in Rome through the DUETTO project. Body, image and technology thus become the material of a single line of research, aimed at moving beyond the perception of dance as a practice reserved solely for the stage.

From the stage to the headset: the birth of Virtual Dance for Real People

The lockdown caused by the pandemic imposed a sudden halt to live activities, prompting CCN/Aterballetto to rethink new ways of experiencing performance and to accelerate research that was already taking shape on the relationship between the stage and the digital world. The underlying question was, and remains, the same: is it possible to build a digital experience capable of matching, in terms of emotional engagement, that of a live performance?

CCN's response was not simply a video recording of a performance, but the design of performances conceived specifically for virtual reality, able to place the viewer — wearing a headset — at the centre of the stage space occupied by the dancers, abandoning the classic frontal relationship between audience and

stage (Figure 3). This gave rise to *Virtual Dance for Real People*, the title under which CCN/Aterballetto launched its first experiments in Cinematic VR, where viewer and dancer, though physically separated, coexist almost touching within the same virtual space. From this initial core of work, the entire strand of research now brought together under the name DANSITE would take shape in the following years.



Figure 1 - Glory Hall Exctract, Pinacoteca Tosio Martinengo, ph Umberto Favretto

Double Reality and the birth of DANSITE – Every Place Can Be a Stage

The leap in scale came with *Double Reality – il monumento e il suo doppio*, a project that won the Emilia-Romagna Region's ERDF (FESR) call for the digitisation and enhancement of cultural heritage (Digital Humanities) (Figure 2). Thanks to this regional funding, the experimentation begun with Virtual Dance for Real People developed into a genuine, permanent system: DANSITE – Every Place Can Be a Stage was born, the platform that brings together the entire repertoire of virtual reality productions created by CCN/Aterballetto at heritage sites across Emilia-Romagna.



Figure 2 - Eppur si Muove, Double Reality Brescia Musei, ph. Umberto Favretto

The project identifies several landmark sites of the region's historical and artistic heritage — Palazzo Milzetti in Faenza, the Rocchetta Mattei in Grizzana Morandi, the Reggia di Colorno, and the Parco Innovazione of the former Reggiane factories in Reggio Emilia — and stages Cinematic VR dance performances conceived site-specifically, in direct dialogue with the architecture, decoration and historical memory of each space. One example is *Il combattimento di Tancredi e Clorinda*, staged at the Reggia di Colorno: direction and visuals by Fabio Cherstich, choreography by Philippe Kratz, set to Claudio Monteverdi's composition of the same name, danced by Gador Lago Benito and Alberto Terribile. Wearing the headset, the viewer finds themselves immersed in the central scene of the performance, in an experience in which dance, art and technology combine to create a multisensory engagement otherwise impossible in a traditional museum setting. Alongside the most recent creations, the same repertoire includes the first productions made in Ferrara — *Don Juan | The Carnival Party*, choreography by Johan Inger, at the Pinacoteca Nazionale di Ferrara/Palazzo dei Diamanti — and in Modena with *Meridiana* by Diego Tortelli at the Farmacia Storica di AGO, evidence of the continuity of a path begun well before the regional funding and now brought together entirely within a single digital infrastructure.

Since 2025 the project's technology partner has been Impersive Srl, a company based in Milan, which has taken charge of the technological development and user experience design of the DANSITE platform and of new content, succeeding RE:LAB, the technology partner of the project's earlier phases.

DANSITE and the VR Corners: a widespread access infrastructure

DANSITE is not merely a digital archive, but an interactive map that shows where the virtual reality

performances were created and which heritage sites have featured in them, complete with information, direct links to the places involved and multimedia content dedicated to each production. The platform also indicates the location of the VR Corners: physical installations consisting of an information panel and headset stations, set up inside heritage sites and freely accessible to the public. By putting on the headset, visitors gain direct access to the DANSITE application, with a stated ambition to expand beyond the region's borders.

One significant technical aspect concerns the possibility of updating the application remotely: each VR Corner thus becomes an access point in constant evolution, able to host ever new content and to turn every location in the network into a permanent gateway to the region's heritage through dance. The title *Double Reality*, however, remains specifically tied to the project developed with the Fondazione Brescia Musei and the Municipality of Brescia as part of ECOMIC – Ecosistema digitale per la cultura, a national initiative promoted by the Ministry of Culture under the PNRR: here CCN/Aterballetto produced five virtual reality creations, one for each of the five museums in the Brescia network, signed by choreographers of national and international standing, also part of DANSITE, which aspires to become a national point of reference for the promotion of virtual content for heritage enhancement (Figure 1).

The resulting model partly overturns the traditional relationship between event and place: the performance is no longer a one-off occurrence tied to a single date, but stable, reactivatable content, able to reach audiences who rarely attend theatres or dance institutions, and to do so through the very channel — that of museum and monumental heritage — which those audiences already regularly attend.

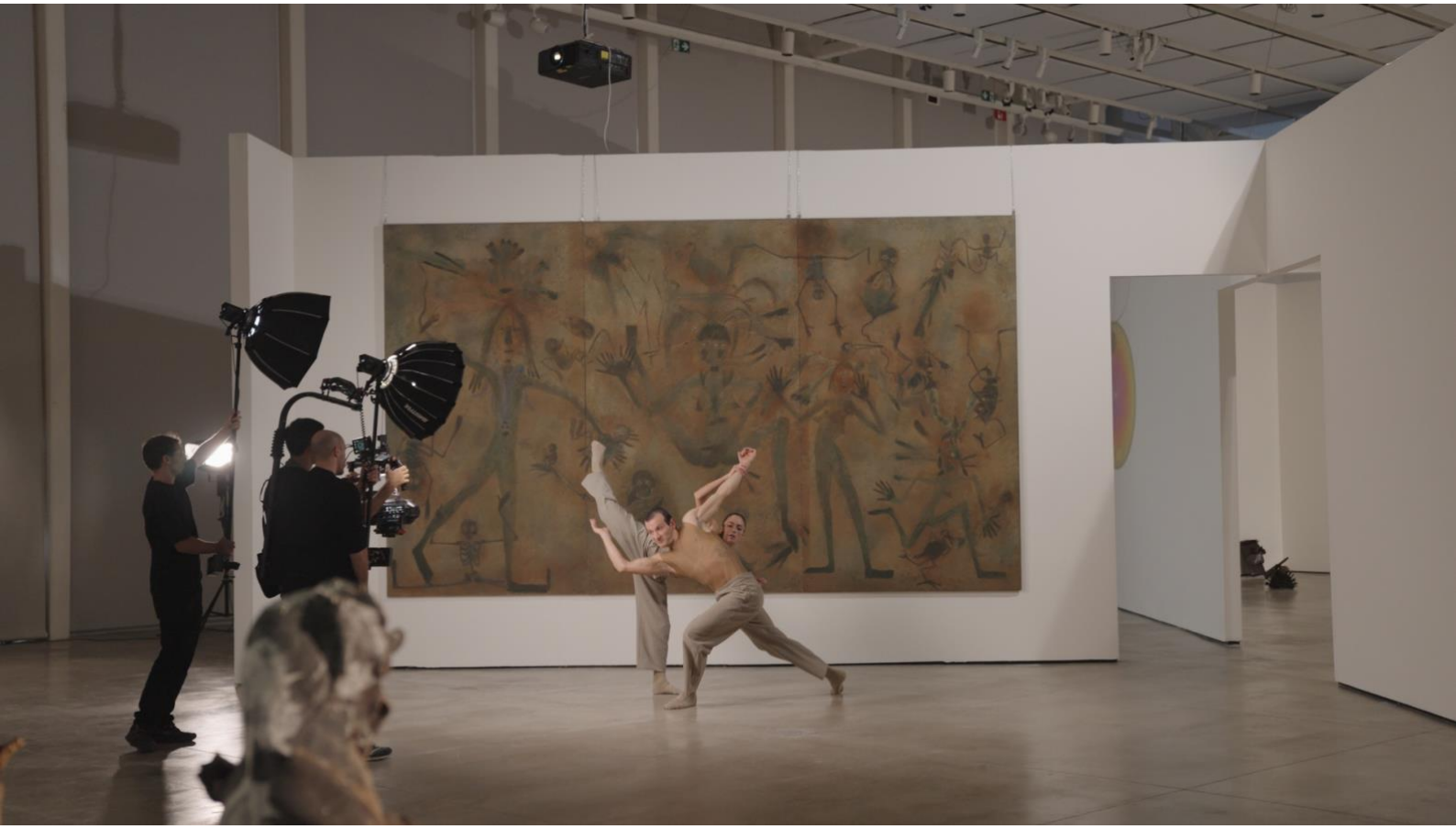


Figure 3 - An echo a frame, Fondazione Golinelli, ph. Valeria Civardi

Beyond the place: from body and monument to body and artificial intelligence

While Double Reality explores the relationship between body and place, since 2024 CCN/Aterballetto has opened a second, parallel line of research dedicated to the relationship between body and generative artificial intelligence. This gave rise to DUETTO, a project developed in collaboration with the Sony Computer Science Laboratories in Rome, with an avowedly ambitious goal: to build the first Large Movement Model, a generative artificial intelligence capable of engaging in real-time dialogue with the human body, transforming movement from a physical act into a creative process shared between human and machine.

The project has already passed the proof-of-concept stage — a working prototype demonstrated the effectiveness of the human–machine dialogue, overlaying predictive digital skeletons onto real movement — and is now on a 24-month industrialisation roadmap structured around three pillars: the construction of the generative model and proprietary dataset, led by Sony CSL; the development of an ultra-low-latency live architecture and a co-creation interface accessible even to non-technical audiences; and, entrusted to CCN/Aterballetto, artistic validation, with experimental sessions conducted with the company's dancers leading up to a final pilot performance.

For CCN/Aterballetto, DUETTO is not a demonstrative exercise, but the natural continuation of a line of thinking pursued for years on the relationship between dance and technology: the latter understood not as a tool that replaces the performer, but as an interlocutor that amplifies their possibilities. It is the same principle — AI as a kinetic partner, not as a special effect — that unites the Centre's two lines of research: the now-mature strand of virtual reality applied to places, and the strand, still undergoing industrialisation, of generative artificial intelligence applied to gesture.

Conclusions

CCN/Aterballetto's experience in the field of virtual reality, distilled in the Double Reality-DANSITE project, shows how contemporary dance can become an engine of innovation not only in language but in infrastructure, capable of producing not only works but tools — a platform, a network of VR Corners, a replicable production model — designed to last beyond the single performance. *Il monumento e il suo doppio* is not just the subtitle of a project, but the synthesis of a method: using technology to multiply, without replacing, the value of the physical encounter between audience, dancing body and place. It is the same method that, shifted from the plane of space to that of gesture, now guides the Centre's research into generative artificial intelligence.

References

Double Reality Project – il monumento e il suo doppio: winner of the Emilia-Romagna Region's ERDF (FESR) call for the digitisation and enhancement of cultural heritage (Digital Humanities); produced by Fondazione Nazionale della Danza/Aterballetto; supported by the Ministry of Culture and by European funds for the region's cultural and digital innovation. The project gave rise to the DANSITE – Every Place Can Be a Stage platform. Technology partner since 2025: Impersive Srl (Milan); technology partner in earlier phases: RE:LAB.

Double Reality Project in Brescia (co-produced with Fondazione Brescia Musei and the Municipality of Brescia, as part of ECOMIC – Ecosistema digitale per la cultura, PNRR – Ministry of Culture)

DUETTO Project: a collaboration between Sony Computer Science Laboratories – Rome and Centro Coreografico Nazionale/Aterballetto.

CHAPTER 10

DanXe: A DIGITAL ECOSYSTEM FOR BALLET HERITAGE. METHODS, TOOLS AND PERSPECTIVES

Silvia Garzarella – University of Bologna

Abstract

The cultural heritage of ballet consists of a complex ecosystem of tangible and intangible sources whose analysis requires tools capable of establishing relationships, enhancing differences, and accounting for their diachronic dimension. For a long time, in the absence of such tools, ballet studies relied primarily on qualitative approaches which, although undoubtedly valuable, were not always sufficient to grasp stratifications and evolutionary processes on a large scale. Building on these premises, this paper reflects on the potential of digital technologies through the case study of *DanXe*, an *Extended Artificial Intelligence toolbox* designed for the organisation, analysis and enhancement of dance heritage. The reflections presented here are the outcome of a doctoral research project (Department of the Arts, University of Bologna/PON Innovation) investigating the integration of traditional methodologies and digital tools for the study of ballet heritage in the digital age.

Introduction

The cultural heritage of ballet consists of a heterogeneous set of sources whose diversity, while representing a richness, also poses several challenges from both an archival and a methodological perspective [1; 2]. The coexistence of material documents, audiovisual testimonies and embodied practices requires systems for data description and organisation capable of preserving their specific characteristics while enabling consultation through the establishment of relationships between sources.

Qualitative approaches [3], which continue to provide the foundation of dance studies, make it possible to address this complexity, but may encounter certain limitations when analysis concerns large quantities of documents or requires consideration of transformations and continuities over time, both characteristic features of a heritage such as ballet, which has been transmitted predominantly through embodied practices for centuries [4; 5].

The increasing availability of digital collections and technological tools fits within this context by encouraging the exploration of opportunities arising from the possible integration of computational procedures and traditional methodologies.

Building on this premise, the present contribution proposes a case study concerning the development of an *Extended Artificial Intelligence toolbox, DanXe*, as an example of the application of digital technologies to the study of ballet heritage [6; 7; 8; 9; 10; 11; 12].

The case study is grounded in the notion of the digital ecosystem promoted by the Italian *National Plan for the Digitisation of Cultural Heritage* (PND) [13] and is based on doctoral research conducted at the Department of the Arts of the University of Bologna within the PON Innovation programme. The research was supervised by Professor Elena Cervellati and carried out with the support of the research group of the Virtual and Augmented Reality Lab (VARlab) of the University of Bologna, directed by Professor Gustavo Marfia.

The Digital Ecosystem Model and the Architecture of DanXe

Published in 2022, the Italian *National Plan for the Digitisation of Cultural Heritage* (PND) defines the digital ecosystem as an evolving network in which information, people, devices and platforms interact and influence one another [13; pp. 44–45]. Within the ecosystem, no element is considered in isolation but acquires meaning through the relationships it establishes with other elements and with the users involved in the production and sharing of content. From this perspective derives the attention that the Plan devotes to data organisation, interoperability and the development of shared infrastructures, regarded as essential for preserving and providing access to digital cultural heritage [13; pp. 29-30, 57-63].

The architecture of *DanXe* was developed to assess the applicability of this model to dance heritage. The toolbox was designed to organise the heritage related to ballet history within an *Extended Artificial Intelligence* environment in which artificial intelligence techniques are integrated into the research process and operate in support of the user's disciplinary expertise. The construction of the documentary corpus, the formulation of research questions and the interpretation of results therefore remain the responsibility of the scholar, while computational tools expand the possibilities for analysing and relating sources.

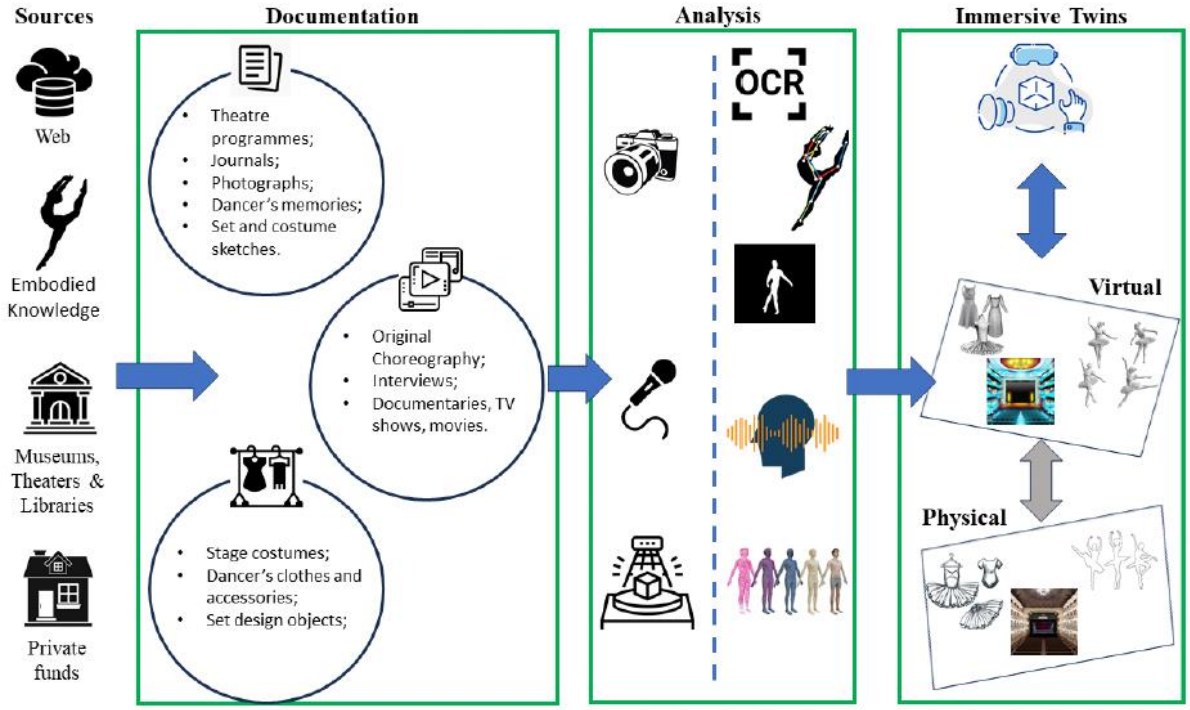


Figure 1. The architecture of the DanXe toolbox: source acquisition, analysis through computational methodologies, and heritage presentation according to the Immersive Twins paradigm.

In practical terms, this meant designing the system by taking as its starting point a recurring research situation: that of a scholar wishing to bring together within a single research pathway documents dispersed across archives, libraries, museums, personal collections and private holdings, belonging to different typologies—including librettos, theatre programmes, photographs, costume and set sketches, costumes, reviews, audiovisual recordings and oral testimonies—produced at different times and described according to non-uniform criteria. The architecture of *DanXe* seeks to reconstruct the main stages of the research process by organising them into three closely interconnected modules, within which digital tools are employed according to the

specific requirements posed by the different documentary typologies involved.

The Acquisition Module

In the first module of *DanXe*, sources from archives, museums, libraries, private collections and oral memories are digitised or acquired in the format in which they are available and organised within a common infrastructure. The system is designed to manage textual documents, images, audiovisual materials, three-dimensional models and audio recordings while preserving the specific characteristics of each documentary typology and making explicit the relationships that connect them. Digitisation constitutes the step that makes these resources available for subsequent processing, transforming the heritage under study into a searchable object [13; pp. 57-63].

The Processing Module

The second module, dedicated to analysis, operates on the corpus thus constructed. The methodologies adopted vary according to the nature of the sources: OCR and *Natural Language Processing* techniques for textual documents, *computer vision* tools for images and videos, audio analysis systems for sound recordings, and three-dimensional reconstruction procedures for objects and costumes [6]. The results produced by each area converge into a common environment where they can be related to one another, generating new information and new possibilities for exploring the heritage. Rather than suggesting a fully automated interpretation of sources, the objective of the module is to support the scholar in the most repetitive tasks and to extend the possibility of analysing documentary corpora whose size and complexity would be difficult to address through traditional procedures alone.

The Visualisation and Access Module

The third module is dedicated to the consultation of sources and the presentation of results. To this end, *DanXe* adopts the *Immersive Twins* paradigm [14], digital environments in which digitised documents and computational outputs are brought together within the same research context, making it possible to explore jointly materials that, in archival practice, are often dispersed across different institutions and collections. A stage costume, for example, can be related to the documentation concerning its production (preparatory sketches, performance photographs, audiovisual recordings, theatre programmes, and so forth); likewise, an audiovisual recording can be explored together with the data generated during the analysis, fostering an interpretation of heritage based on relationships between sources and providing scholars with a tool capable of supporting the comparative approach that has always characterised research in the performing arts.

An Open-Source Foundation for *DanXe*

For the development of *DanXe*, priority was given to open-source technologies selected according to the different documentary typologies involved. Textual documents were acquired through JaidedAI OCR, used for text extraction and subsequent indexing; images and audiovisual recordings were analysed through computer vision models, with particular reference to OpenPose and AlphaPose for recognising dancers' poses and movements, VIDAT for the assisted annotation of audiovisual content; and 3D Gaussian Splatting using StableFast3D for the three-dimensional digitisation of costumes and stage objects [15]. The presentation of results was finally developed through a WebXR platform integrated with Large Language Models (GPT-4) for the interactive exploration of the heritage. The selection of these tools was guided by the principles of interoperability and infrastructure reuse, consistently with the digital ecosystem model proposed by the PND.

Challenges and Opportunities

One of the main challenges in designing the ecosystem was identifying methodologies and tools appropriate to each documentary typology and consistent with both the nature of the sources and the objectives of the research. Digitisation involves a series of choices concerning the acquisition and processing of materials, on which the subsequent possibilities for analysis and for establishing

relationships between sources depend. The adoption of *open-source* technologies stems from this awareness and responds to the need to build a sustainable and interoperable infrastructure in which OCR, *computer vision*, three-dimensional modelling and other tools can be employed according to the characteristics of different documentary materials.

A particularly significant example for ballet heritage is the stage costume, whose fragility, often resulting from its state of conservation, makes access difficult, and whose enhancement requires reconstructing its performance history by examining its relationship with the dancer's body, the choreography, and the context for which it was created [16; 17; 18].

To address these requirements, *DanXe* employed three-dimensional reconstruction techniques, such as *Gaussian Splatting*, to produce digital models of costumes and subsequently make them accessible within the visualisation module, where they were connected to the related documentary heritage. Understood in this way, the three-dimensional model becomes part of the digital ecosystem alongside preparatory sketches, performance photographs, reviews, theatre programmes, posters and audiovisual materials, making possible an interpretation of the cultural object based on the comparison of different sources.



Figure 2. Application of Gaussian Splatting to the three-dimensional reconstruction of an object depicted in a press photograph.

In the case of the heritage of an internationally renowned dancer such as Rudol'f Nureev (1938–1993), for example, photographs published in the press made it possible to reconstruct three-dimensionally elements of costumes that are no longer accessible today. Once incorporated into the ecosystem, these models were connected with the documentation relating to the performance and contributed to reconstructing the history of the cultural object under investigation [9; 19].

The same principle can be applied to audiovisual recordings, which represent some of the most important testimonies of dance heritage.

Starting from the acquisition of the video, the processing module of *DanXe* sought to enrich the description of the source by employing *deep learning* techniques capable of automatically recognising performers, choreographic sequences, movements, stage elements and other relevant content. The extracted information can then be integrated into the metadata and queried together with the associated documentation, making it possible to relate the recording to theatre programmes, reviews, photographs, librettos and other audiovisual materials.

Organising sources within a system of relationships therefore means, on the one hand, providing users with modes of consultation that make it possible to reconstruct pathways based on people, works, places and creative processes; on the other hand, offering researchers an environment in which heterogeneous documents preserved across different institutions can be explored together. The complexity that characterises ballet heritage, distributed across tangible and intangible sources and closely connected to its performative dimension, makes this field a particularly significant testing ground for digital infrastructures dedicated to cultural heritage. The experimentation carried out through the *DanXe* toolbox suggests that engagement with the performing arts can contribute to the development of models and tools capable of addressing needs that extend beyond dance heritage alone, recognising intangible heritage as one of the most significant contexts in which future infrastructures for cultural heritage research and enhancement can be tested and refined.

References

[1] Adshead-Lansdale, Janet, Layson, June (a cura di). *Dance History. An Introduction*. New York: Routledge, 1994.

[2] Franco, Susanne, Nordera, Marina (a cura di). *Ricordanze. Memoria in movimento e coreografie della storia*. Torino: UTET Università, 2010.

[3] Brooks, Lynn Matluck, Meglin, Joellen A. *From the Editors: Preserving the Ephemeral*, in Dance Chronicle, vol. 34, no. 1, 2011, pp. 1–4.

[4] Lepecki, André. *Of the Presence of the Body*. Middletown: Wesleyan University Press, 2004.

[5] Fischer-Lichte, Erika. *Estetica del performativo*. Roma: Carocci, 2014.

[6] Stacchio, Lorenzo, Garzarella, Silvia, Cascarano, Pasquale, De Filippo, Allegra, Marfia, Gustavo, Cervellati, Elena. *DanXe: an Extended Artificial Intelligence Framework to Analyze and Promote Dance Heritage*, in «Digital Applications in Archaeology and Cultural Heritage», XXXIII, 2024, pp. 1-22.

[7] Garzarella, Silvia. *Ri-mediatizzazione del patrimonio immateriale della danza per la realizzazione di archivi digitali e prodotti di tecnologia digitale avanzata. Intorno al caso Rudol'f Nureev (1938-1993)*, tesi di dottorato, Alma Mater Studiorum – University of Bologna, Dottorato di ricerca in Arti visive, performative, mediali, XXXVII cycle, 2025.

[8] Garzarella, Silvia. *Il patrimonio culturale della danza fra materialità e immaterialità*, in Il cinquantenario della Convenzione UNESCO a tutela del patrimonio culturale (1972-2022): percorsi multidisciplinari, Alma Mater Studiorum – University of Bologna, Bologna, 2024, pp. 49-58.

[9] Stacchio, Lorenzo, Garzarella, Silvia, Cascarano, Pasquale, De Filippo, Allegra, Marfia, Gustavo, Cervellati, Elena. *Preserving and Annotating Dance Heritage Material through Deep Learning Tools: A Case Study on Rudolf Nureyev*, in CREAI 2024. Workshop on Artificial Intelligence and Creativity, CEUR Workshop Proceedings, vol. 3810, 2024, pp. 88-97.

[10] Garzarella, Silvia. *Il patrimonio culturale del balletto fra tradizione e tecnologie digitali, in Nuovi crocevia per lo spettacolo: linee di ricerca e generazioni a confronto*, Atti del convegno (Bologna, 12-13 settembre 2024), 2025, pp. 543-555.

[11] Stacchio, Lorenzo, Garzarella, Silvia, Cascarano, Pasquale, De Filippo, Allegra, Marfia, Gustavo, Cervellati, Elena. *A Large Reconstruction Model Driven Approach to Support Humans in Digitization of Dance Visual Material into 3D Environments*, in HAIC 2025 Workshop on Human-AI Collaborative Systems, CEUR Workshop Proceedings, vol. 4072, 2025, pp. 47-56.

[12] Garzarella, Silvia. *Rethinking Ballet Heritage through the Digital Ecosystem: a PND-Inspired Approach*, in «Arti dello Spettacolo / Performing Arts», XI, 2025, pp. 70-78.

[13] Ministero della Cultura. Piano nazionale di digitalizzazione del patrimonio culturale. 2022. https://digitallibrary.cultura.gov.it/il-piano/.

[14] Garzarella, Silvia, Vallasciani, Giacomo, Cascarano, Pasquale, Hajahmadi, Shirin, Cervellati, Elena, Marfia, Gustavo, *An Extended Reality Platform Powered by Large Language Models: A Case Study on Teaching Dance Costumes*, 2025: 369-375.

[15] JaidedAI OCR: https://www.jaided.ai/easyocr/; OpenPose: https://openpose.org/; AlphaPose: https://github.com/MVIG-SJTU/AlphaPose ; Vidat: https://vidat2.davidz.cn/#/ ; Stable Fast 3D: https://github.com/Stability-AI/stable-fast-3d.

[16] Barbieri, Donatella. *Costume in Performance*. London–New York: Bloomsbury, 2019.

[17] Lewis, Jane, Striezheff. Marie, *Costume: The Performing Partner*. Colorado Springs: Meriwether, 1990

[18] Bignami, Paola, *Storia del costume teatrale*. Roma: Carocci, 2005.

[19] Garzarella, Silvia, *Il cappotto di Drosselmeyer: lo Schiaccianoci di Rudol'f Nureev nell'Archivio Storico del Teatro alla Scala di Milano*, in "Fra archivio e scena. Saggi sullo spettacolo italiano del secondo Novecento". Mantova: Il Rio, 2025, pp. 123–136.

CHAPTER 11

EXPERIENCING THE ARCHIVE: DIGITAL COMMUNICATION PRACTICES FOR A SHARED INTANGIBLE HERITAGE

Elena Radaelli – University of Milan, PNRR CHANGES

Introduction

Archives of the performing arts occupy a distinctive position within cultural heritage: they are both indispensable and inherently insufficient. Indispensable, because without a trace there is no access to memory, nor any means of reconstructing what has been performed, danced, or staged before dissolving into absence. Insufficient, because the trace is structurally incapable of reproducing the event it records. This is not a contingent limitation amenable to technological resolution; it is a constitutive condition, one that Peggy Phelan identified long ago as the irreducible ontology of performance: the event has its only life in the present, and «*performance's only life is in the present, and performance's being* […] *becomes itself through disappearance*» [1]. Scripts, photographs, costume sketches, and correspondence may survive; the performance itself does not. What follows from this observation is a genuinely difficult question – not merely a logistical one – concerning how to communicate an archive of living tradition to audiences who were not present and cannot be. The "Marionette 3.0" project, developed within the PNRR CHANGES research programme at the University of Milan, takes this question as its point of departure. Working with the archive of the Teatro delle Marionette di Gianni e Cosetta Colla, the project developed a set of digital communication strategies, each calibrated to specific audiences and their cultural resources. The central argument of this paper is that audience differentiation is not a practical afterthought but the foundational methodological choice from which all technical decisions follow.

The Archive: what survives and what it means

The Colla company is among the oldest continuously active marionette theatre traditions in Europe, founded in Milan in the nineteenth century. Its archive, held at APICE, encompasses over two thousand archival units organised into ten series: scripts, correspondence, sketches, photographs and slides, posters, press clippings, seasonal programmes, relocation documents, videos, and miscellaneous materials, alongside thirty puppets. This distribution is already analytically significant: photographs and sketches together constitute approximately three quarters of the holdings. The company's history is preserved primarily through visual and material traces rather than through records of the performance itself.

This imbalance is not unique to the Colla archive. Performing arts archives generally preserve the conditions surrounding performance more readily than the event itself: preparatory materials, administrative records, visual documentation, and evidence of reception. They tell us what a production looked like before it reached the stage, how it was reviewed, and how companies interacted with theatres, institutions, and audiences. What remains largely inaccessible is the performance as lived experience. For those engaged in communicating archival heritage, this creates a particular challenge. The materials are rich and varied, but their connection to the living tradition they document is neither immediate nor self-evident. It has to be constructed through interpretation and mediation,

and through decisions about which audiences are being addressed.

Against this background, the Marionette 3.0 project identified two primary target groups: adults, ranging from academic specialists to culturally interested non-specialists, and children between six and twelve years of age, who constitute the core audience of the Colla company's performances. Distinct digital tools and communication formats were developed for each group. What connects these initiatives is not a shared medium or aesthetic approach, but a common assumption: archival materials can become meaningful and engaging beyond specialist communities when they are translated into forms that correspond to the interests, expectations, and competencies of different audiences.

For adult audiences: data visualisation and digital storytelling

A first challenge in communicating the archive to adult audiences concerns the archive's structure itself. The organisation of documents across decades, the network of correspondents, and the distribution of different material types are not simply descriptive features of the collection. They reveal patterns in the company's history that are difficult to grasp through the consultation of individual documents alone.

To make these patterns visible, the project relied extensively on data visualisation. Archival metadata extracted from the APICE management system were reorganised according to the ISAD(G) [2] standard and used to develop a suite of interactive visualisations. R was employed for exploratory analysis, while the web interface was implemented using SvelteKit, D3.js and Flourish. The resulting tools include a zoomable circle-packing visualisation of the archival hierarchy, stacked bar charts showing the distribution of document types across decades, Gantt charts representing correspondence series over time, a network graph of the company's correspondents organised by typology and period, alongside other key elements.

The rationale behind these visualisations was to make the archive available in its entirety rather than through a pre-selected set of highlights [3]. Instead of presenting users with a curated pathway, the platform allows them to navigate the collection's overall structure and to explore relationships independently. In this sense, the visualisations do not replace archival description; they provide an additional layer through which the collection can be approached and understood. The zoomable circle-packing visualisation is particularly effective in this respect. By displaying series, subseries, and archival units within a single navigable interface, it makes visible the relative proportions of different sections of the archive in ways that are difficult to achieve through conventional hierarchical lists. One pattern becomes immediately apparent: photographs and sketches account for the largest portion of the archive, whereas audiovisual materials are comparatively rare, with only sixteen video items recorded in the inventory. This same asymmetry, already noted in relation to the archive's overall composition, becomes immediately legible here through the visualisation itself, confirming its value as an analytical tool rather than merely a display device.

For non-specialist adult audiences, however, data exploration alone is insufficient. Raw structure, however elegantly visualised, does not communicate the affective and cultural significance of a theatrical tradition. This is where narrative becomes essential, and where the project developed a second register of communication operating alongside and in dialogue with the data visualisation layer. Narrative content was produced in two forms. The first consists of thematically organised scrollytelling sequences, data stories in which editorial commentary and progressive archival visualisations unfold in synchrony as the reader scrolls through the page. Scrollytelling, extensively theorised in data journalism and increasingly adopted in digital humanities contexts [4], is particularly well suited to heritage communication because it maintains, as Segel and Heer argue, a balance between «author-driven» and «reader-driven» narration [5]: thematic framing guides interpretation while the underlying interactive visualisations remain available for exploration. These data stories were developed around specific themes drawn from the archive's contents, each designed to remain accessible to readers without prior knowledge of the Colla tradition, while still maintaining analytical depth.

The second form of narrative content consists of digital storytelling pieces developed using the Shorthand platform, centred on individual puppet plays: *La Regina della Neve*, *La Sirenetta*, and *Pinocchio*.

These pieces adopt a play- and character-centred approach that translates the archive's taxonomy of series, subseries, and archival units into something closer to the experience of theatrical spectatorship: the encounter with figures that carry a history, a material presence, and a narrative trajectory. Archival photographs, costume sketches, rehearsal notes, and press reviews become the building blocks of stories that can be followed and experienced without specialist knowledge, while remaining firmly grounded in documentary evidence.

The two registers, data visualisation and digital storytelling, are not alternatives but complementary modes addressing different cognitive and interpretive needs within the same broad audience. The visualisation foregrounds analytical and exploratory engagement; the narrative foregrounds affective and interpretive engagement. Together, they articulate an approach to archival communication that takes seriously both the integrity of the archive and the diversity of its potential audiences.

For young audiences: experiential scaffolding

The challenge posed by children as an audience is not a scaled-down version of the adult challenge. It is qualitatively different: adult non-specialists can be engaged through narrative and visual exploration because they already possess the cultural frameworks – an implicit understanding of what an archive is, what theatre is, what cultural heritage means – within which those strategies can operate. Children typically do not possess those frameworks, and communicative strategies that rely on them tend to fail. At the same time, children bring capacities that adults have often partially lost: imaginative engagement, embodied learning, and a form of affective openness that precedes critical mediation. The educational laboratory developed for Marionette 3.0 was designed to work with those capacities rather than to compensate for the absence of others.

The laboratory format, developed in close collaboration with the Teatro Colla company, combines Augmented Reality, Virtual Reality, and authentic marionettes from the Colla collection in a structured sequence. Its organizing principle can be described as experiential scaffolding, where each medium prepares the encounter with the next, building toward an experience that is simultaneously:

- historical: rooted in the archive and the tradition;
- technological: mediated through contemporary digital tools;
- embodied: involving physical contact with authentic objects and actual performance practice.





Figg. 1-2: Digital storytelling session introducing the Snow Queen's play history to children; AI-generated reinterpretation of the Snow Queen puppet.

The laboratory opens with archival materials presented and contextualised by a facilitator who introduces the puppet characters and the tradition through projected digital storytelling (fig. 1). In a laboratory organised during a Colla event dedicated to *La Regina della Neve* in January 2025, children were introduced to the historical Snow Queen puppet, designed by Luigi Veronesi and preserved in the archive. No longer used on stage and replaced in contemporary productions by a live actress, the puppet became the starting point for a collective creative activity. After exploring its history and visual characteristics, participants were invited to imagine how the character might appear if the marionette were to return to the stage today. Together they discussed possible costumes, colours, and aesthetic features, and their ideas were translated into prompts used to generate new visual interpretations of the character in real time through AI image-generation tools (fig. 2). Rather than simply receiving information about an archival object, children were encouraged to engage with it as something capable of generating new meanings and new creative possibilities. The distinction between the archival record and its contemporary reinterpretation was maintained throughout as a conscious methodological choice. The generated images were not presented as replacements for the historical marionette, but as speculative interpretations grounded in an encounter with the archival object itself.





Figg. 3-4: AR 3D model of the historical Snow Queen puppet; AR 3D model of the historical Snow Queen puppet within a digital theatre environment.

This imaginative phase prepares the transition to the AR component of the laboratory. By scanning a QR code, children can visualise three-dimensional models of the marionette characters within the actual theatrical space, allowing digital figures to appear on the very stage where performances traditionally take place. The models are generated through photogrammetric acquisition of original puppets preserved in the archive and function as a form of archival reactivation rather than reconstruction. In cases where characters are no longer part of contemporary productions, the XR environment enables a temporary return of these figures to the performative space from which they have disappeared. For instance, the Snow Queen can reappear on stage in her original spatial context (fig. 3; fig. 4), not as a reenactment of a historical performance, but as a renewed encounter between archival object, theatrical space, and present audience.

The experience is directly connected to a VR application (fig. 5) designed for younger participants, in which children can manipulate a digital puppet using natural hand gestures, as if operating a real puppet. The interaction is intentionally intuitive – grasping, lifting, rotating the character in space – and does not require prior knowledge of theatre or puppetry. What is particularly significant is that this digital manipulation is not isolated from material practice, but positioned within a continuum of experience that links virtual action to physical embodiment. This continuity becomes explicit in the public deployment of the laboratory. The same VR system was presented at TourismA, and later in a more developed form during MeetMeTonight, where the presence of the puppeteers from the Teatro Colla made it possible for participants to move fluidly between virtual manipulation and the handling of authentic marionettes (fig. 6). In this configuration, the gesture learned in VR – pulling strings, orienting a body in space, responding to balance and weight – was immediately tested against the resistance and materiality of the physical puppet, producing a form of embodied comparison between digital and analogue performance.





Figg. 5-6: VR puppeteering experience; Participants interacting with virtual and physical puppets during MeetMeTonight (courtesy of Unimi - Direzione Comunicazione ed Eventi Istituzionali).

Conclusions

The Marionette 3.0 project demonstrates that the valorisation of a performing arts archive through digital tools requires, above all, a prior decision about for whom that valorisation is being undertaken. This is not a self-evident point. Much digital heritage communication proceeds as though the primary challenge were technical – the right format, the right platform, the right degree of interactivity – and treats audience as a secondary concern, something to be addressed once the product exists. The approach developed here inverts that priority: audience definition precedes format selection, and the differences between audiences are treated as substantive rather than superficial. For adult audiences, this leads to a two-register strategy: data visualisation for exploratory and analytical engagement, and digital storytelling for narrative and affective engagement. For children, it leads instead to a laboratory format that unfolds across archival encounter, speculative reinterpretation, XR spatialisation, and VR interaction, culminating in direct contact with authentic marionettes. What is crucial in both cases is that digital tools are not deployed as ends in themselves, but as transitional devices that reorganise the relationship between archival material, interpretation, and lived experience. Across these different configurations, a common pattern emerged during public engagement activities: participants rarely distinguished between digital and physical modes of engagement in the terms used by the project itself. Instead, they tended to respond to the experience as a continuous field in which interaction, observation, and interpretation were intertwined. The VR manipulation of marionettes, the emergence of characters in XR space, and the handling of physical puppets were not perceived as separate modalities, but as variations within a single experiential continuum. What both strategies ultimately share, therefore, is a refusal to treat the gap between archive and performance as a problem to be solved through simulation or substitution. The archive cannot reproduce the live event; digital media cannot replace the presence of a performer and a puppet. Yet they can reorganise the conditions under which that absence is perceived, transforming it from a loss into a productive relation. The goal is not to reconstruct theatrical experience in digital form, but to design pathways that make its irreducibility perceptible, and, in doing so, to reposition the archive as an invitation to encounter, rather than as a surrogate for what is no longer present.

Acknowledgements

The research and valorisation project Marionette 3.0 is funded by the project 0000020 CHANGES – CUP G53C22000430006, within the framework of the Italian National Recovery and Resilience Plan (PNRR), Mission 4 (Education and Research), Component 2 (From Research to Business), Investment 1.3 (Extended Partnerships), Thematic Area 5 (Humanistic culture and cultural heritage as laboratories for innovation and creativity).

The author gratefully acknowledges the coordinators of WP4 of Spoke 2 within the CHANGES PNRR programme, Francesco Tissoni and Maddalena Mazzocut-Mis, as well as Vincenzo Pernice, scientific coordinator of the Marionette 3.0 project, and the whole project team for their contribution to the development and implementation of Marionette 3.0.

Special thanks are also due to APICE and the Teatro Colla for their collaboration and continued support.

Team

Alessandra Bergamaschi (AI artist/AI imagery)

Paola Abbiati (3D scenography/spatial design)

Katia Bianco (research and valorisation)

Camilla Ippolita Donato (illustration)

Francesco Laezza (3D development/technical support)

MDVSA – Alessia Musa (3D artist/modelling)

Mariasofia Quaresima (3D scenography/spatial design)

Vincenzo Pernice (coordination and workshop design)

Elena Radaelli (data visualisation and workshop design).

References

[1] Phelan P., *Unmarked: The Politics of Performance*, Routledge, London, 1993. DOI: https://doi.org/10.4324/9780203359433

[2] International Council on Archives, *ISAD(G): General International Standard Archival Description*, 2nd edition, 2000. Available at https://www.ica.org/resource/isadg-general-international-standard-archival-description-second-edition/

[3] Whitelaw M., "Generous Interfaces for Digital Cultural Collections" in *Digital Humanities Quarterly*, vol. 9, no. 1, 2015. DOI: https://doi.org/10.63744/4pg836gkjaeb

[4] Riche N.H., Hurter C., Diakopoulos N., Carpendale S. (eds.), *Data-Driven Storytelling,* CRC Press, Boca Raton, 2018. DOI: https://doi.org/10.1201/9781315281575

[5] Segel E., Heer J., "Narrative Visualization: Telling Stories with Data" in *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, 2010, pp. 1139–1148. DOI: https://doi.org/10.1109/TVCG.2010.179

CHAPTER 12

ACKNOWLEDGEMENTS

The organizers wish to thank Digital Library, Direzione generale Digitalizzazione e comunicazione of the Italian Ministry of Culture for supporting the workshop organisation.

The organization of the workshop and the publication of these proceedings would not have been possible without the help of Roberto De Nicolò, Simona Caraceni, Daniele De Luca, Angela Lanzarini, Fulvio Cuppi, Paolo Vendruscolo.

AI, Cultural Heritage, Art and Science

Workshop proceedings – March 27, 2026

IN COLLABORATION WITH

UNIVERSITÀ DI PARMA

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

CINECA

www.cineca.it

ISBN 978-88-86037-72-3

DOI 10.82341/6fd4-5d78