FINAL PROGRAM AND PROCEEDINGS

ARCHIVING2025

Granada **24-27 June**

General Chair: Carolina Gustafsson,
Stiftelsen Föremålsvård i Kiruna (Sweden)

www.imaging.org/archiving

Sponsored by the Society for Imaging Science and Technology

imaging.org

The papers in this volume represent the program of Archiving 2025,
held June 24–27, 2025 in Granada, Spain.

Society for Imaging Science and Technology
7003 Kilworth Lane • Springfield, VA 22151 USA
703/642-9090; 703/642-9094 fax
info@imaging.org; www.imaging.org

ISBN Abstract Book: 978-0-89208-369-5
ISSN Print: 2161-8798
ISSN Online: 2168-3204
https://doi.org/10.2352/issn.2168-3204.2025.22.1.0

Cover image: Suzanne Grinnan

CONFERENCE EXHIBITORS

ARKHÊNUM
REVEALING HISTORIES

artefactual systems inc.

ChannelScience
Establish the State-of-the-Art™

IRON MOUNTAIN®
MEDIA & ARCHIVAL SERVICES

Max

NOA
CREATING ARCHIVE INNOVATIONS

PICTURAE

VerusDIGITAL
TRUE. 3D. DIGITIZATION.

WELCOME TO ARCHIVING 2025

On behalf of the organizing committee, I extend a warm welcome to all the attendees of Archiving 2025. This year we are honored to gather at the Escuela Técnica Superior de Arquitectura in Granada, a city renowned for its unique blend of history, culture, and design. Set against the backdrop of the Alhambra and centuries of architectural heritage, Granada offers an inspiring setting for our shared mission: to digitize, preserve, and make accessible the cultural heritage that defines us.

We continue this year with the theme, "Science, Sustainability, and Security," which underscores our dedication to advancing the field of Archiving through scientific innovation, while acknowledging the crucial roles that sustainability and security play in the stewardship of memory. As professionals committed to preservation, we recognize that these are not secondary concerns—they are core principles that guide our work.

These ideas are already part of our ongoing conversations, but I encourage you to take them further. Let us explore how we can integrate sustainability more deeply into our practices and build stronger, more secure systems for the future of preservation.

The Escuela Técnica Superior de Arquitectura, with its focus on design, the environment, and built heritage, is a perfect venue to reflect on how our work intersects with space, culture, and community. Granada, with its streets, its architecture, and its history, reminds us that preservation is not only about the past—it is about shaping a thoughtful, resilient, and inclusive future.

As we come together to share ideas and innovations, I invite you to fully engage in all that Archiving 2025 has to offer. Whether during formal sessions or informal conversations over coffee, every exchange is an opportunity to learn, connect, and contribute.

Thank you for being part of Archiving 2025 and welcome to Granada!

—Carolina Gustafsson, General Conference Chair, Archiving 2025

CONFERENCE SPONSOR

PICTURAE

CONFERENCE DONOR

HIP

CONFERENCE COMMITTEE

General Chair
Carolina Gustafsson, Stiftelsen Föremålsvård i Kiruna (Sweden)

Program Chair
Todd Swanson, J. Paul Getty Trust (US)

Short Course Chairs
Laura Ramsey, Metropolitan Museum of Art (US)
Eryk Bunsch, Museum of King Jan III's Palace at Wilanow (Poland)

Local Arrangements Chairs
Eva Maria Valero Benito, University of Granada (Spain)
Miguel Angel Martinez Domingo, University of Granada (Spain)

Past Chair
Robert Kastler, Museum of Modern Art (US)

AV Chair
Alexandre Leão, Universidade Federal de Minas Gerais (Brazil)

COOPERATING SOCIETIES

- American Institute for Conservation Foundation of the American Institute for Conservation (AIC)
- Council on Library and Information Resources (CLIR)
- Coalition for Networked Information (CNI)
- Digital Preservation Coalition (DPC)
- Royal Photographic Society (RPS)

PAPER REVIEWERS

Tarek A. Haila, TU Darmstadt (Germany)
Ottar A.B. Anderson, Intermunicipal Archive of Møre og Romsdal (Norway)
Hana Beckerle, Library of Congress (US)
Michael J. Bennett, University of Connecticut (US)
Peter Burns, Burns Digital Imaging (US)
Marco Buzzelli, University of Milano-Bicocca (Italy)
Antenie Carstens, National Library of South Africa (retired) (South Africa)
Elizabeth Chiang, George Eastman Museum (US)
Hilda Deborah, Norwegian University of Science and Technology (Norway)
Peter Dueker, J. Paul Getty Trust (US)
Roger Easton, Rochester Institute of Technology (US)
Susan Farnand, Rochester Institute of Technology (US)
Rebecca Frank, University of Michigan School of Information (US)
Sony George, Norwegian University of Science and Technology (Norway)
Federico Grillini, University of Copenhagen (Denmark)
Carolina Gustafsson, Stiftelsen Föremålsvård i Kiruna (Sweden)
Steffen Hankiewicz, intranda GmbH (Germany)
Lei He, Library of Congress (US)
Chris Heins, The Metropolitan Museum of Art (US)
Javier Hernández Andrés, University of Granada (Spain)
Kurt Heumiller, National Gallery of Art (US)
Meghan Hill, Library of Congress (US)
Martina Hoffmann, Schweizerische Nationalbibliothek (Switzerland)
Chrissy Huhn, University of California Berkeley Library (US)
Nora Ibrahim, Osher Map Library and Smith Center for Cartographic Education (US)
Anssi Jääskeläinen, South-Eastern Finland University of Applied Sciences (Finland)
Robert Kastler, The Museum of Modern Art (US)
Daniel Kemp, BYU (US)
Olivia Kuzio, Getty Conservation Institute (US)
Ana B. López-Baldomero, University of Granada (Spain)
Ana López-Montes, University of Granada (Spain)
Alexandre Leão, Universidade Federal de Minas Gerais (Brazil)
Hana Lukesova, University Museum of Bergen (Norway)
Vincent Wai-Yip Lum, The Chinese University of Hong Kong (Hong Kong)
Tabita Lumban Tobing, Norwegian University of Science and Technology (Norway)
Barry Lunt, Brigham Young University (US)
Giacomo Marchioro, University of Verona (Italy)
Miguel Ángel Martínez Domingo, University of Granada (Spain)
Anne Mason, National Archives and Records Administration (US)
Francisco Moronta-Montero, University of Granada (Spain)
Jeanine Nault, Smithsonian Institution (US)
Taren Ouellette, Library of Congress (US)
Thierry Paquet, University of Rouen Normandy (France)
Deepa Paulus, The Metropolitan Museum of Art (US)
Marius Pedersen, Norwegian University of Science and Technology (Norway)
Kristin Phelps, Library of Congress/US Copyright Office (US)
Alice Plutino, University of Amsterdam (the Netherlands)
Ty Popko, Walt Disney Archives (US)
Laura Margaret Ramsey, The Metropolitan Museum of Art (US)
Jamie Rogers, Florida International University (US)
Lynda Schmitz Fuhrig, Smithsonian Institution (US)
Carla Schroer, Cultural Heritage Imaging (US)
Bethany Scott, Yale University (US)
Steven Simske, Colorado State University (US)
Astrid J. Smith, Stanford University Libraries, Digital Production Group (US)
Dina Sokolova, Columbia University (US)
Markus Sebastian Storeide, Norwegian University of Science and Technology (Norway)
Michael Tetzlaff, University of Wisconsin-Stout (US)
Giorgio Trumpy, Norwegian University of Science and Technology (Norway)
Mahsa Vafaie, FIZ-Karlsruhe (Germany)
Eva Maria Valero, University of Granada (Spain)
Corine van Dongen, Koninklijke Bibliotheek (the Netherlands)
Marie Vans, Colorado State University (US)
Christoph Voges, consultant (Germany)
Charles Walbridge, Minneapolis Institute of Art (US)
David R. Wyble, Gray Sky Imaging Inc. (US)

TECHNICAL PAPERS PROGRAM
CONFERENCE SCHEDULE AND TABLE OF CONTENTS

TUESDAY 24 JUNE 2025

SHORT COURSE PROGRAM

8:45 – 10:45
SC04: Imaging Performance: Meeting Guidelines for Digital Collections
Instructors: Peter Burns, Burns Digital Imaging, and Don Williams, Image Science Associates

8:45 – 13:00
SC01: Camera Color Profiles: The Theory, Practice, and Pitfalls
Instructors: David R. Wyble, Gray Sky Imaging Inc., and Doug Peterson, Digital Transitions

11:00 – 13:00
SC05: Advanced Digital Imaging Techniques Applied to Cultural Heritage
Instructor: Miguel Ángel Martínez Domingo, University of Granada

14:15 – 16:15
SC08: OpenDICE for Imaging Quality Assessment
Instructor: Lei He, Library of Congress

SC09: Archival Negatives: A Novel Toolkit for Historically-accurate Inversions
Instructors: Alice Plutino and Luca Armellin, Imageese

SC10: Reference Assets-Bridging the Gap Between 2D and 3D Imaging
Instructors: Scott Geffert, Juan Trujillo, and Chris Heins, The Metropolitan Museum of Art

16:30 – 18:30
SC11: Advanced Concepts in Color Measurement
Instructor: David R. Wyble, Gray Sky Imaging Inc.

SC12: Can/Should AI Do That? Heritage Digitization Use Cases for AI
Instructors: Julie McVey, National Geographic Society, and Doug Peterson, Digital Transitions

SC14: Capturing Specularity with Kintsugi 3D and Camera-mounted Flash
Instructor: Charles Walbridge, Minneapolis Institute of Art

BEHIND-THE-SCENES TOURS

IS&T thanks the following institutions for opening their doors to attendees as part of the Behind-the-Scenes Tours program.

- Archivo de la Real Chancillería de Granada
- Archivo Histórico Provincial de Granada
- Department of Restoration and Conservation University of Granada
- Color Imaging Laboratory: University of Granada
- Escuela Técnica Superior de Arquitectura University of Granada
- Hospital Real
- Palacio de la Madraza
- Universitario de Cartuja

WELCOME RECEPTION
19:30 – 21:00

Join colleagues for a welcome drink and snacks at Peña la Platería, a traditional flamenco club in the heart of the Albaicín, Granada's enchanting and historic neighborhood of winding cobblestone streets. While we won't see any flamenco dancing tonight, you'll enjoy the outdoor courtyard and a lovely view of Alhambra.

Address: Placeta de Toqueros, 7

WEDNESDAY 25 JUNE 2025

WELCOME AND OPENING KEYNOTE
Session Chair: Carolina Gustafsson, Stiftelsen Föremålsvård i Kiruna (Sweden)
09:00 – 10:00

Advancing Heritage Science: European Breakthrough Initiatives Driving Digital Innovation in Cultural Heritage, *Vania Virgili, director-level research technologist, Institute for Heritage Science, National Research Council of Italy (ISPC CNR) and appointed director general, European Research Infrastructure for Heritage Science (E-RIHS ERIC) (Italy)*
The digital transformation of heritage science is essential for improving analysis, preservation and transmission of cultural assets. The European Research Infrastructure for Heritage Science is at the forefront of this transformation, aiming to integrate physical and digital access to cutting-edge techniques and resources. Through its unique catalogue of services, E-RIHS includes four research platforms. ARCHLAB provides access to physical and digital archives, including datasets, reports, and reference materials from museums and research institutions. FIXLAB grants access to large-scale research equipment for advanced diagnostics and archaeometry, offering high-resolution 2D/3D imaging and spectroscopic analyses. MOLAB is a mobile lab for in-situ, non-invasive diagnostics of immovable heritage objects. Among others, it includes advanced imaging and image processing techniques such as multispectral, spatial, and 3D imaging. On top of this, a key opportunity is the current development of E-RIHS DIGILAB. DIGILAB is designed to manage, share, and enable the creation of new knowledge from data produced by ARCHLAB, FIXLAB and MOLAB. Its conceptual model ensures seamless dataflow, linking research questions, methods, instruments, data production, analysis and interpretation, and knowledge. DIGILAB leverages the heritage digital twin concept to offer cutting-edge data-driven services. National projects in France, Italy, Poland and Slovenia contribute to its development to foster the growth of the E-RIHS community. In this context, E-RIHS strengthens the vision of a cohesive digital ecosystem for heritage science with the European Cloud for Heritage OpEn Science, a key initiative in developing the EU's Cultural Heritage Cloud, connecting institutions and professionals across Europe.

SPECTRAL IMAGING

Session Chair: Yoko Arteaga, NTNU (Norway)

10:00 – 12:00

10:00 **Exploring the HYPERDOC Database: Advancing Hyperspectral Imaging for Historical Document Analysis,** *Ana B. López-Baldomero, Francisco Moronta-Montero, Miguel A. Martínez-Domingo, Yannick Lefier, Juan Luis Nieves, Javier Hernández-Andrés, Ramón Fernández-Gualda, Anna S. Reichert, Ana López-Montes, Teresa Espejo, Javier Romero, and Eva M. Valero, University of Granada (Spain)* **1**

The HYPERDOC database is a publicly available hyperspectral imaging resource for the analysis of historical documents and mock-ups of inks and pigments. It consists of 1681 hyperspectral datacubes, containing millions of reflectance spectra, covering the VNIR (400–1000 nm) and SWIR (900–1700 nm) spectral ranges, including different ink recipes and documents from the 15th to 20th centuries, preserved in two archives in Granada, Spain. We will present the data acquisition process and structure of the database, followed by a live demonstration of its functionality, guiding participants through its use. Additionally, three applications of the database will be summarized, including document binarization, ink classification using machine learning techniques, and ink aging analysis. The HYPERDOC database facilitates the integration of advanced imaging techniques into document analysis and preservation, contributing to the non-invasive study of historical materials.

10:15 **Multi-spectral Scanning for Analogue Film Digitisation: Addressing LED Variability in Spectral Band Selection,** *Mihaela Elizabeta Balica and Giorgio Trumpy, Norwegian University of Science and Technology (Norway)* **7**

The digitisation of analogue film is critical for cultural heritage preservation, as film deteriorates over time due to environmental factors and analogue projectors are becoming obsolete. Conventional RGB scanning methods fail to fully capture the spectral complexity of film, making multispectral imaging (MSI) a feasible alternative. However, MSI faces challenges due to the limited availability of narrow-band LEDs in certain spectral regions and inherent variability in LED emissions. Aiming to minimise colour reproduction errors in film scanning, this study investigates the optimisation of LED spectral band selection and the impact of LED spectra variability. Informed by the optimised bands, multiple market-based LED sets were further evaluated using MSI capture simulations, with the 7-band and 8-band setups achieving good colour accuracy and showing rather low sensitivity to LED spectral variability. A physical multispectral capture of a film photograph demonstrated a good agreement between the capture simulation and the real results.

10:30 **Adaptive Combined Method for Material Identification in Documents of Historical Interest,** *Eva M. Valero, Francisco Moronta-Montero, Ana B. López-Baldomero, Anna S. Reichert, Miguel A. Martínez-Domingo, Rosario Blanc, and Ana López-Montes, University of Granada (Spain)* **12**

Hyperspectral imaging has been widely and consistently applied in the field of Cultural Heritage for material identification. In the specific context of historical document analysis, it is frequently supported and complemented by additional analytical techniques. In this study, we propose a straightforward method for material identification that combines adaptive direct identification—using a reference library of visible and near-infrared spectral reflectance data for pigments—with a KNN classifier applied to an extended spectral range for inks and supports. The method has demonstrated a high degree of accuracy, successfully identifying materials present in both actual historical documents and mock-ups created following medieval techniques. Its performance is illustrated through three spectral image fragments extracted from the HYPERDOC project database.

10:45 – 11:30

COFFEE BREAK / EXHIBITS OPEN / POSTERS AVAILABLE FOR VIEWING

11:30 **Pigment Identification of Ortelius' Historical Maps using Hyperspectral Imaging,** *Zealandia S. N. Fatma, Hilda Deborah, Jon Y. Hardeberg, and Eleftherios Papachristos, Norwegian University of Science and Technology (Norway)* **18**

Hyperspectral imaging (HSI) has been widely used in the conservation studies of various cultural heritage (CH) objects, e.g., paintings, murals, and handwritten historical manuscripts. In this work, HSI is used to study painted historical maps, i.e., five maps of the Scandinavia region from the Ortelius collection preserved at the National Library of Norway in Oslo. Given knowledge of their colour application and usage, HSI-based pigment identification is performed, assuming several spectral mixing theories, i.e., pure pigments, subtractive, and additive mixing models. The obtained results are discussed, showing both the pure pigment and subtractive mixing model to be suitable for pigment identification in the case of watercolour applied on paper substrate.

11:45 **Heritage Science—Spectral Sustainability and Innovative Dissemination,** *Fenella G. France, Library of Congress (US)* . **24**

While spectral imaging has now been being utilized in cultural heritage for more than 20 years, there is still a lack of uptake by heritage practitioners. While some point to cost as an issue, it appears the real concern is that of communicating effectively with interested users—conservators, curators, scholars, heritage professionals. Many people are not aware of the range of types of information and data that can be captured and made available from collections, and the potential ease of interacting with the datasets. Since spectral imaging is essentially the next element of digitization and making heritage available in the digital realm, it seems necessary for more effort to be placed on shared knowledge of the spectral capture and processing methodology, so this becomes more accessible as a tool. Setting up a new spectral imaging system, communicating and creating networks for engagement, and addressing opportunities and challenges will be discussed.

3D: CAPTURE AND PROCESSING I

Session Chair: Lien Acke, J. Paul Getty Museum (US) and University of Antwerp (Belgium)

12:00 – 12:45

12:00 **Reference Assets: Leveraging Traditional Photographic Techniques to Improve 3D Object Renditions,** *Scott Geffert, The Metropolitan Museum of Art (US)* **29**

As 3D Imaging for cultural heritage continues to evolve, it's important to step back and assess the objective as well as the subjective attributes of image quality. The delivery and interchange of 3D content today is reminiscent of the early days of the analog to digital photography transition, when practitioners struggled to maintain quality for online and print representations. Traditional 2D photographic documentation techniques have matured thanks to decades of collective photographic knowledge and the development of international standards that support global archiving and interchange. Because of this maturation, still photography techniques and existing standards play a key role in shaping 3D standards for delivery, archiving and interchange. This paper outlines specific techniques to leverage ISO-19264-1 objective image quality analysis techniques for 3D color validation, and methods to translate important aesthetic photographic camera and lighting techniques from physical studio sets to rendered 3D scenes. Creating high-fidelity still reference photography of collection objects as a benchmark to assess 3D image quality for renders and online representations has and will continue to help bridge the current gaps between 2D and 3D imaging practice. The accessible techniques outlined in this paper have

vastly improved the rendition of online 3D objects and will be presented in a companion short course.

12:15 **Willem Witsen's Coat: From Restoration to Digital Preservation,** *Frans Pegt, Rijksmuseum (the Netherlands)* **A-1**
This article explores the multidisciplinary process behind the restoration and digitisation of Willem Witsen's painter's coat—an object of both artistic and historical value. Using techniques such as photogrammetry, 3D modelling, and 360-degree photography, the project aimed to digitally preserve the fragile coat while making it accessible to both researchers and the general public. The result is a high-quality digital surrogate that supports future conservation efforts and storytelling.

12:30 **Seeing the Unseen. The Selene Project for 3D Digitization of Cultural Heritage,** *John Barrett, Bodleian Libraries (UK); Jorge Cano, Carlos Bayod, Costanza Blaskovic, and Santiago del Bosque Arias, Factum Foundation (Spain); and Ana Carrasco-Huertas, University of Granada (Spain)* **36**
Funded by the Helen Hamlyn Trust, ARCHiOx – Analysis and Recording of Cultural Heritage in Oxford, is a collaborative project which has united Oxford University's Bodleian Libraries and the Factum Foundation. The Selene Photometric Stereo Scanner was conceived and developed by the latter and has, for the last three years, been piloted at the Bodleian Library. This technology has been used to reveal near-invisible text and artwork from originals from across Oxford University's collections. Renders created with the Selene PSS, have revealed what is difficult or impossible to record through conventional photography, and have allowed for the creation of physical facsimiles. This paper serves to demonstrate how Selene recordings have assisted in the research of cultural heritage originals and natural history specimens.

EXHIBITOR PREVIEWS

12:45 – 13:10

Session Chair: Laura Ramsey, Metropolitan Museum of Art (US)

Exhibitors Arkhênum, artefactual, ChannelScience, Iron Mountain Media and Archival Services, Max, NOA, Picturae, and VerusDigital share information about their products/services in short previews.

13:10 – 14:45

LUNCH BREAK ON OWN

3D: CAPTURE AND PROCESSING II

Session Chair: Lien Acke, J. Paul Getty Museum (US) and University of Antwerp (Belgium)

14:45 – 15:15

14:45 **Specialized 3D Meshes: Variations in Mesh Structures in Heritage Databases,** *Markus Sebastian Bakken Storeide and Sony George, Norwegian University of Science and Technology (Norway)* **42**
Mass 3D digitization of heritage objects is today heavily encouraged by various institutions. This is in an effort to measure and document objects for the future, use them for visualization and dissemination, and open up for the analytic tools that are available for 3D meshes. However, the structure required of a mesh depends heavily on the application, and the data might vary significantly based on the digitizing institution, object characteristics, and acquisition workflow.

In this work, we sample 3D data stored in several major open-access databases for 3D heritage data and analyze the content. We take a close

look at sampled mesh structures by computing various graph metrics, check some integrity measures, and evaluate their possible future use. Finally, we provide an overview of the use cases and interoperability of the meshes depending on results from the mesh structure analysis.

15:00 **You Do Not Know Until It is Not Usable: Metadata and Paradata Management for Photogrammetry Preservation and Archiving,** Irina Schmid and Elizabeth H. Day, The American University in Cairo (Egypt) **48**

Photogrammetry has greatly improved the recording, preservation, and accessibility of cultural heritage in archaeology and scientific research. The increased use of 3D modeling in heritage projects brings about significant challenges, especially in terms of data management. In this context, the challenges involve ensuring that digital models are reliable, traceable, and usable. Often, these concerns are disregarded until they impede access or reuse, affecting the long-term preservation and accessibility of cultural heritage data.

THREE-MINUTE INTERACTIVE PAPER PREVIEWS I

Session Chair: Irina-Mihaela Ciortan, NTNU (Norway)

15:15 – 15:35

Surveying Imaging Workflows and Software in Cultural Heritage, *Nina Eckertz, Hilda Deborah, and Jon Y. Hardeberg, Norwegian University of Science and Technology; and Irina C. A. Sandu, Munch Museum (Norway)* **56**

This work presents insights into the imaging workflow from cultural heritage domain experts, gathered from an online survey. Non-invasive 2D imaging technology has become a cornerstone in the analysis and documentation of cultural heritage artefacts. Techniques such as hyperspectral imaging (HSI) and X-ray fluorescence (XRF) can investigate material properties, artistic processes, and conservation states. Existing analysis and visualisation tools offer functionality for specific data types but lack integration for holistic multimodal analysis. To address these limitations, we conducted a structured survey targeting researchers and practitioners in CH working with imaging technology. The survey explores their workflows, imaging technology usage, and software preferences. This study identifies key trends, challenges, and feature requirements.

Revealing Erased Words: The Application of Multispectral Imaging to the Book of Hours 50,1,1 at the Brazilian National Library, *Alexandre Oliveira Costa[1], Isamara Carvalho[1,2], Alexandre Cruz Leão[1], Kethlin Barroso[1], and Márcia Almada[1]; [1]Federal University of Minas Gerais and [2]Brazilian National Library (Brazil)* **62**

Multispectral imaging has become an essential tool for the analysis, documentation, and visualization of cultural heritage materials and objects. This study explores the application of this technique to a 15th-century illuminated manuscript held at the Brazilian National Library (Fundação Biblioteca Nacional) in Rio de Janeiro. The manuscript, currently part of ongoing doctoral research, contains erased text due to censorship through scraping. The use of multispectral imaging, incorporating eleven different wavelengths across UV, visible, and IR spectra, proved highly effective in recovering the erased words "pape" and "thoma", thus confirming the hypothesis of scholar Damião Berge regarding the lacunae and linking the codex to a 16th-century historical event.

From Invisible to Visible: Scientific Imaging Applied in Drawings by Alberto da Veiga Guignard, *Larissa Lorrane Silva Oliveira[1,2], Alexandre Cruz Leão[1], and Alexandre Oliveira Costa[1]; [1]Federal University of Minas Gerais and [2]SECULT-MG - State Secretariat of Culture and Tourism of Minas Gerais (Brazil)* **68**

Non-invasive scientific imaging is increasingly becoming an important research tool in the study of cultural heritage objects, combining sustainability and preservation in a responsible and conscious manner. The purpose of this study is to make a legibility by digital restoration of two drawings by Brazilian modernist artist Guignard, dated 1956 and 1958, respectively, one created with ink and pen nib, and the other presumably with graphite. Both drawings have likely suffered from photodegradation, with almost total loss of visibility due to fading and erasure of the drawn lines. Scientific photography techniques were employed, including IR reflectance photography, visible light photography, UV fluorescence, Multispectral Imaging, transmitted light photography, raking light photography, and Reflectance Transformation Imaging (RTI). These techniques, used in combination, yielded positive results, enhancing legibility and revealing traces and details that were no longer visible to the naked eye.

Gloss Archiving with Normal Vector, *Shinichi Inoue[1], Yoshinori Igarashi[2], Shota Tsuneyasu[1], and Yoko Mizokami[3]; [1]Tokyo Polytechnic University, [2]CHUO Precision Industrial Co., LTD., and [3]Chiba University (Japan)* **73**

This paper proposes a method of gloss archiving using normal vectors. When archiving the gloss phenomenon of a material, it is important to record not only the reflected light intensity but also the gloss unevenness. This is because the gloss unevenness greatly affects the texture of the material. However, it has been difficult to quantitatively record gloss unevenness because they are dependent on the viewing direction and lighting. Gloss unevenness on mirror surfaces are mainly caused by irregularities in the normal direction. Therefore, we came up with a solution to archive the gloss unevenness phenomenon by recording the distribution of surface normal vectors. We are currently developing a apparatus to measure the distribution of surface normal vectors. Using this surface normal data, it will be also possible to reproduce gloss unevenness images using Computer Graphics technology.

Beyond Two Scribes? Column-based Writer Identification in the 1QIsa^a Scroll, *Tabita Lumban Tobing and Patrick Bours, Norwegian University of Science and Technology (Norway)* **77**

The 1QIsa^a Scroll, one of the most significant manuscripts among the Dead Sea Scrolls, has long been the focus of debate over whether it was produced by a single hand or multiple scribes. In this study, we introduce a column-based writer-identification framework that combines unsupervised clustering, character-level verification, and cross-dataset evaluation, without assuming any fixed number of scribes. Benchmarking our hinge–feature–based approach against the widely recognized FIREMAKER dataset reveals its strengths and weaknesses. This exploratory analysis not only offers fresh insights into 1QIsa^a's scribal attribution but also underscores the need for richer or complementary features in future digital paleographic research.

15:35 – 16:15

COFFEE BREAK / EXHIBITS OPEN / POSTERS AVAILABLE FOR VIEWING

QUALITY AND GUIDELINES

Session Chair: Eryk Bunsch, Museum of King Jan III's Palace at Wilanow (Poland)

16:15 – 17:20

16:15 **Large Format Digitization with New Equipment and New Workflows in the Swiss National Library,** *Martina Hoffmann, Swiss National Library (Switzerland)* **83**

The Swiss National Library (SNL) operates a variety of different digitization projects for different kinds of materials. Besides newspapers, journals and monographies the SNL holds unique collections of writers' legacy and a great number of objects like posters, plans and drawings. Posters, plans and drawing are most of the time of a larger format than

the other objects. The digitization service has a duty to digitize them all. This paper shows the quest to find and utilize equipment large enough to digitize objects that exceed the usual sizes of books and newspapers. This is a case study of the Swiss National Library, and its vendors and equipment are named. It shows the process of getting new equipment and installing new workflows in the SNL for a specific use case of this institution. The SNL does not provide an assessment of the market or any statement on other brands or vendors. This paper is not a reinforcement of any brand or vendor but solely states the choices the SNL has made for herself. It is not to be read as advertising or recommendation of any kind.

16:30 **Survey on Mobile Phone Camera Use in Cultural Heritage Documentation,** *Leah Humenuck[1], Sony George[2], and Susan Farnand[1]; [1]Rochester Institute of Technology (US) and [2]Norwegian University of Science and Technology (Norway)* . **89**

Mobile phone cameras are imaging tools that are rapidly being adopted by various industries due to their portability and ease of use. Though not currently considered an adopted imaging tool for cultural heritage, there has been increased interest in their potential use within the field. To better understand how cultural heritage professionals considered mobile phone cameras as tools for various types of documentation, a survey was created and administered. A survey was designed and sent to cultural heritage groups involved with imaging with the goal of determining whether these types of cameras are practical imaging devices in circumstances where a studio or a DSLR may not be readily available. Initial results have shown a variety of responses and that mobile phones are being used for various types of documentation.

16:45 **Implementing a Quality Rating System for Legacy Digital Image Collections: A Case Study from the National Gallery of Art,** *Kenneth N. Fleisher, National Gallery of Art (US)* **95**

The National Gallery of Art developed a systematic approach to evaluate and categorize its extensive digital image collection spanning 20 years of technological evolution. This study addresses the challenge of inconsistent image quality resulting from varying capture technologies and methodologies over time. A four-tier rating system was created based on comprehensive analysis of capture devices, technical specifications, and workflow documentation. The system enables efficient assessment of image suitability for different applications while providing clear guidance for re-digitization decisions. The implementation includes integration with the institution's digital asset management system, offering a practical framework that other cultural heritage institutions can adapt for managing legacy digital collections while maintaining current quality standards.

17:00 **Metamorfoze Version 2.0,** *Hans van Dormolen, Independent Imaging Consultant (the Netherlands)* **99**

This year version 2.0 of the Metamorfoze Preservation Imaging Guidelines was published. Version 1.0 was published in 2012. The Metamorfoze guidelines are published by the KB, the National Library of the Netherlands.

Metamorfoze Version 2.0 is a technical and practical update of Version 1.0. And one of the three quality levels is completely rewritten for mass digitization with sheet-fed scanners.

With the technical update Metamorfoze Version 2.0 is in line with ISO 19264-1. With the practical update it is possible to use a broad range of technical targets.

The basic principle of the Metamorfoze guidelines: What you see is what you get, applies to all quality levels of the Metamorfoze guidelines.

17:15 **Closing remarks / evening on own**

THURSDAY 26 JUNE 2025

THURSDAY KEYNOTE

Session Chair: Todd Swanson, J. Paul Getty Trust (US)

9:00 – 10:00

Archiving the Invisible: Advanced Imaging to Recover Inaccessible Texts and Images in Historic Documents, *Lucía Pereira Pardo, researcher, Institute of Heritage Sciences (INCIPIT), Spanish National Research Council (CSIC) (Spain)*

Information has been accidentally or intentionally obscured in countless historic documents, photographs, and films. Specifically, legibility of texts and visibility of images may be compromised due to, on the one hand, deterioration of the support writing or image-forming materials and on the other hand, past human interventions. The massive scale of archival collections, organizations´ finite resources, severity of the damage, and the limits of current conservation methods are important obstacles to address this problem.

The goal of the project "The Museum of the Invisible – Spectral Imaging Techniques for the Digital Recovery of Deteriorated Heritage" is to use advanced imaging techniques as a possible solution to digitally unlock this inaccessible knowledge. The project has shown particularly promising results in archive collections with challenging conservation problems.

The applied methodology is comprised of a range of complementary imaging techniques, such as multiband and hyperspectral imaging in the UV-VIS-NIR range, X-Ray Fluorescence scanning, Raman imaging, and micro-CT-scanning. Image processing methods like binarization, Principal Component Analysis (PCA), or Spectral Angle Mapper (SAM) among others, are also investigated to improve the readability of the writing and the contrast of the images.

This talk provides an overview of the main results obtained so far, both on reference samples prepared in the laboratory and on historic case studies from a range of archival collections. Representative successful cases are shown, analysing the imaging techniques and processing methods that proved more adequate in each case, depending on the variety of document media and causes of information loss. Equally, unresolved challenges and further research needed to access the contents of these invaluable historic materials is discussed.

AI: GENERATING DATA

Session Chair: Julie McVey, National Geographic Society (US)

10:00 – 11:45

10:00 **JIST-first: A Building-block Approach to Character-level Writer Verification on the Great Isaiah Scrolls,** *T. Lumban Tobing and P. Bours, Norwegian University of Science and Technology (Norway)* . . . . . **see JIST 69(2)/**
DOI: 10.2352/J.ImagingSci.Technol.2025.69.2.020401

This study presents a novel character-level writer verification framework for ancient manuscripts, employing a building-block approach that integrates decision strategies across multiple token levels, including characters, words, and sentences. The proposed system utilized edge-directional and hinge features along with machine learning techniques to verify the hands that wrote the Great Isaiah Scroll. A custom dataset containing over 12,000 samples of handwritten characters from the associated scribes was used for training and testing. The framework incorporated character-specific parameter tuning, resulting in 22 separate models and demonstrated that each character has distinct features that enhance system performance. Evaluation was conducted through soft voting, comparing probability scores across different token levels, and contrasting the results with majority voting. This

approach provides a detailed method for multi-scribe verification, bridging computational and paleographic methods for historical manuscript studies.

10:15 **A Hybrid 3D Laser Scanning and Machine Learning System: Paving the Way for Autonomous Braille Digitization,** *Lei He, Library of Congress (US)* **103**

The Library of Congress has been conducting the Braille digitization to preserve its tactile braille music collection electronically. In this project we scan braille papers with a 3D laser sensor to obtain the coordinates of the recto/front and verso/back dots, which are fed into our digitization program for dots pattern recognition. Our software group the dots in different lines and further into different characters. Specifically, each dot in a character is identified according to the relative positions to its local neighbors in the same character, based on which the character is recognized and the corresponding ASCII glyph code is written into the final output file. Experiments on different collections show the robustness of our system.

10:30 – 11:15

COFFEE BREAK / EXHIBITS OPEN / POSTERS AVAILABLE FOR VIEWING

11:15 **Simple and Effective ASR for Archives,** *Anssi Jääskeläinen, South-Eastern Finland University of Applied Sciences (Finland)* **107**

Archives are traditionally identified as holders of text-based information. However, they also possess audio and video materials, which are the focus of this paper. In archival institutions, the absence of transcriptions for audio and video materials presents significant challenges. These materials often hold historical, cultural, and research value, but without transcriptions, their accessibility and usability are limited. The lack of transcriptions makes it difficult to index and search the content, hindering effective utilization. While existing ASR (Automatic Speech Recognition) technologies can assist, these may suffer from mediocre accuracy, especially with older or poor-quality materials. This work addresses the challenge by utilizing state of the art multilingual LLM (Large Language Model), simple to use UI (User Interface) and GPU (Graphics Processing Unit) ready containers to create a simple and effective multilingual transportable ASR module.

11:30 **AI-driven Metadata Extraction and Semantic Search for Audiovisual Archives,** *André Rattinger, Giacomo Alliata, Kirell Benzi, and Sarah Kenderdine, EPFL (Switzerland)* . . . **112**

ArchiveVault is a next-generation digital archiving system designed to enhance access to audiovisual collections through automated metadata extraction and advanced retrieval mechanisms. Traditional archiving methods are labor-intensive, requiring extensive manual annotation that often leads to incomplete and inconsistent metadata. ArchiveVault addresses this challenge by employing AI-based transcription, named entity recognition (NER), speaker diarization, and pose detection to extract structured metadata from audiovisual archives. This allows for rich, searchable metadata that improves retrieval precision beyond traditional keyword-based approaches.

By leveraging state-of-the-art AI techniques, ArchiveVault enables researchers, archivists, and content creators to perform semantic searches across large collections, discovering moments of interest more effectively. Our deployments in a national broadcast archive (RTS) and the Olympic Games media collection demonstrate how AI-driven processing unlocks previously inaccessible content, from spoken-word analysis to pose-based retrieval for sports footage.

AI: LEVERAGING DATA

Session Chair: Miguel Ángel Martínez Domingo, University of Granada (Spain)

11:45 – 12:30

11:45 **A White-box Machine Learning Model for Dating Archival Materials Using IR Spectroscopy,** *Hend Mahgoub[1], Patrick Layton[2], Sonja Svoljšak[3], Jasna Malešič[3], Johannes Tintner-Olifiers[2], and Matija Strlič[1]; [1]University of Ljubljana (Slovenia), [2]Academy of Fine Arts, Institute of Natural Sciences and Technology in the Arts (Austria), and [3]NUL, National and University Library of Slovenia (Slovenia)* **A-3**

This study highlights how infrared (IR) spectroscopic techniques, combined with machine learning (ML), can transform the dating of archival materials. By integrating near-infrared (NIR) and Fourier-transform infrared (FTIR) spectroscopy with ML methods, we establish correlations between spectral signatures and paper aging markers and composition. The Ancient Book Crafts (ABC) project demonstrates the practical applications of these techniques, confirming their potential to produce accurate, data-driven dating models. A dataset of 100 well-dated paper objects from the National and University Library (Slovenia) was analyzed to develop predictive models for dating. Initial results show the superior accuracy of NIR in detecting bulk aging effects, while FTIR-ER proves success as a valuable non-contact tool compared to ATR for dating and surveying archival materials. Future advancements in spectral preprocessing, model optimization, and interdisciplinary collaboration will further refine this approach, ensuring widespread applicability across cultural heritage collections and supporting evidence-based conservation strategies.

12:00 **AI-based Indexing for Secure and Efficient Archival Digitalization,** *Julia Sjöholm, The National Archives of Sweden (Sweden)* **A-5**

This extended abstract presents a full-scale production system developed by the National Archives of Sweden for large-scale digitization and AI-assisted indexing of over 54 million pages of property-related documents. The system, operational since May 2024, efficiently extracts metadata at scale while ensuring appropriate protection of sensitive archival content, maintaining efficient processing times.

The system architecture covers the entire digitization chain: physical document handling, high-speed scanning at 290 images per minute, AI processing, automated validation, and manual review of approximately 10% of the predictions. The AI-based indexing pipeline includes object detection (YOLOv9 with 0.80 precision and 0.90 recall), handwriting recognition (TrOCR with a CER of 0.0145), and automated validation, achieving an 82% efficiency gain compared to manual indexing.

Continuous model development through offline retraining using quality assurance feedback enhances performance over time. Early tests with a Donut model, integrating detection and recognition without bounding boxes, are promising and allow reuse of previously processed data.

The system maintains data sovereignty through internal processing and secure access control, with potential data-sharing collaborations being explored with the Swedish Mapping, Cadastral, and Land Registration Authority (Lantmäteriet). The architecture provides a scalable and reusable framework for public institutions aiming to combine large-scale digitization and internal network security. This document outlines the technical implementation, performance evaluation, and future development strategies.

12:15 **Enhancing Archival Transparency with AI: A RAG-based Case Study on Cinematic Heritage,** *Enea Ahmedhodzic and Andrea Mario Trentini, Università degli Studi di Milano (Italy)* . **117**

Cinematic archives preserve an invaluable heritage, yet accessing their content is often challenging due to data fragmentation, inconsistent standards, and the absence of user-friendly tools. Even when materials are available, consultation may require archival staff or specialized knowl-

edge. This paper introduces Valter, a prototype AI chatbot developed as a case study for the Film Center Sarajevo (FCS), which explores how retrieval-augmented generation (RAG) can support transparency and accessibility in under-resourced archival settings. The system uses semantic search over multilingual embeddings to retrieve relevant information and generate answers in natural language. While still under development, Valter demonstrates the potential to enhance resource discovery and metadata validation, offering a replicable approach that could inform future digital access strategies across similar institutions.

12:30 – 14:00
LUNCH ON OWN

STEWARDSHIP, KNOWLEDGE, AND CONNECTION

Session Chair: Todd Swanson, J. Paul Getty Trust (US)
14:00 – 15:15

14:00 **JIST-first: Writing for the Future. What are the Options?,** *Barry M. Lunt, Felipe Riviera, Joshua Santos, Armando Carreon, Joshua Isaacson, and Matthew R. Linford, Brigham Young University (US)* . . . . . **see JIST 69(2)/**
DOI: 10.2352/J.ImagingSci.Technol.2025.69.2.020402
An ideal archival storage system combines longevity, accessibility, low cost, high capacity, and human readability to ensure the persistence and future readability of the stored data. At Archiving 2024, the authors' research group presented a paper that summarized several efforts in this area, including magnetic tape, optical discs, hard disk drives, solid-state drives, Project Silica (a Microsoft project), DNA, and projects C-PROM, Nano Libris, and Mil Chispa (the last three being the authors' research). Each storage option offers unique advantages in each of the desirable characteristics. This paper provides information on other efforts in this area, including the work by Cerabyte, Norsam Technologies, and Group 47 DOTS, and an update on the authors' projects CPROM, Nano Libris, and Mil Chispa.

14:15 **Digital Object Authenticity: Creating a Standard Practice,** *Julie McVey, National Geographic Society (US); Doug Peterson, Digital Transitions (US); and Ottar A.B. Anderson, Intermunicipal Archive of Møre og Romsdal (Norway)* **A-8**
Galleries, Libraries, Archives, and Museums do not currently have a standardized, simple way to communicate digital surrogate quality to users through standard, shared metadata fields. Creating these shared standard fields and implementing content authenticity tools in an agreed-upon and widely adopted data model will allow for greater transparency and institutional trust. The Digital Object Authenticity Working Group (DOAWG) aims to work with existing organizations and standards commonly used in the heritage field to create a cultural heritage object authenticity standard. We believe this goal is best accomplished by bringing cultural heritage professionals together to facilitate a timely and thorough approach to establishing this standard practice in institutions across the world.

14:30 **Embodying Scholarly Annotations in the Network of Digitized Archives,** *Tsz-Kin Chau and Sarah Kenderdine, Swiss Federal Technology Institute of Lausanne (EPFL) (Switzerland)* **124**
The increasing availability of digitized archive presents new opportunities for scholarly research, yet effective reuse of these resources requires infrastructure that is interoperable with the open data. This paper presents a novel annotation platform for the scholarly research of visual material, based on a customization of the ResearchSpace platform. By directly integrating images served via the IIIF Image API (v2v3), the system ensures a high level of digital provenance and source reliability, crucial for research in history-oriented humanities.

The paper outlines the platform's technological framework, highlighting how its architecture fosters deep scholarly engagement with digitized materials and exploring the potential applications of annotation outcomes. Additionally, it discusses the challenge of streamlining the connection between researchers and digitized archives. Future improvements will focus on automating metadata integration and tackling interoperability challenges across diverse data models.

14:45 **Building Online Expert Networks to Support Continuous Learning—Case Digitisation Experts in Finland,** *Miia Kosonen, South-Eastern Finland University of Applied Sciences (Finland)* . **130**
Online networks are a highly topical issue for continuous learning and also for management policy development within the cultural heritage sector. In our project, a well-known framework for community building was adopted and tested in practice: firstly, we conducted a needs assessment and planned usability and sociability practices in parallel. We explored how to succeed in building an online expert network, relying on careful needs assessment, openness, a participatory approach and identifying lead users. Our case highlights the significance of data safety, encouraging networks to rely on platforms where the user community manages its own knowledge base. We believe our case has important implications for cultural heritage professionals internationally, but it is also insightful for other expert groups and professional sectors.

15:00 **Preservation, Security & Access Through Documentation: A Collaborative Exchange Between the Met and the National Museum Lagos,** *Chris Heins and Juan Trujillo, The Metropolitan Museum of Art (US)* **134**
In late October 2024 a team of 6 staff from the Metropolitan Museum of Art in New York travelled to Lagos Nigeria to participate in a collaborative exchange with the staff of the National Museum Lagos. While there were multiple goals based on a two-way exchange of knowledge between the institutions, this paper will mostly focus on one of the primary goals of the Met team which was to provide photographic gear and training to enable the staff of the National Museum Lagos to produce high-quality photographic documentation of their vast and important collection of West African art.

BEHIND-THE-SCENES TOURS

Tours start at 17:00 or 17:30 and end at 18:30.

CONFERENCE DINNER

20:15 – 23:30

After a taking a Behind-the-Scenes tour or exploring Granada and taking a rest or enjoying a "merienda", a late afternoon/early evening snack or tapa, join colleagues at the historic Carmen de los Chapiteles, a lovely traditional house and garden located below the Generalife, the former residence of the sultans of Alhambra. Known for centuries as the House of the "Rich Moor Abu-Jamr"—a unique, bohemian, libertine, poet, hedonist, and generous doctor who altruistically treated the poor and collected old books—the venue offers exquisite, night views of Alhambra. Cocktails are followed by a delicious Spanish dinner at a traditional dining time.

Address: Carmen de los Chapiteles,
Cam. Fuente del Avellano, 4, Centro

CLOSING KEYNOTE

9:00 – 10:00

Session Chair: Carolina Gustafsson, Stiftelsen Föremålsvård i Kiruna (Sweden)

Navigating Through a Sea of Information: Towards a More Multifaceted View of the World, *Johanna Fries Markiewicz, international coordinator, The National Archives of Sweden (Sweden)*

The invasion of Ukraine has made it evident that common knowledge about Ukraine in the "West" has been dominated by an alternative narrative that suggests that Ukraine is not a nation in its own right, with its own culture, history, and right to self-determination. This discovery of an inherited distortion in knowledge has opened up new forms of knowledge production where Ukraine is now the subject, not the object, of other narratives. The process highlights the need to diversify and represent not only Ukraine, but also many other voices and communities of Eastern Europe, the Baltic, the Caucasus, Central Asia, and others.

This shift in focus has led to an interest in researching Ukraine's history outside Ukraine, for example, in archives in Sweden and other Western countries. New sources of history are found, and new research is made possible. The digitization of archival sources is important in this process and making the historical sources freely available to the public, without restrictions on their use or distribution, is a natural standpoint, a practice closely linked to the principles of transparency, accountability, and collaboration. However, at the same time, Russia continues to use history as a tool for domination and confrontation and will, if possible, misuse the data.

Touching upon the themes of the conference—science, sustainability, and security—and with examples from a current project that the Swedish National Archives, together with Wikimedia and cultural heritage institutions in Ukraine, is working on right now, the presentation reflects upon the power of digitization as a way to give voices to histories not told, paving the way and contributing to a more multifaceted view of the world. Questions are also raised about how to secure the narratives built upon open data when there is not an academic process that reviews the writings. In the archives, single documents are contextualized by their physical position, and there is a set of material metadata that is taken account by the researchers. When publishing single documents online, for example on Wikimedia, there is a risk of losing the archival context and therefore opening it up for misuse. How is it possible to avoid disinformation and misuse and keep academic standards in the world of open data and accessibility?

THREE-MINUTE INTERACTIVE PAPER PREVIEWS II

Session Chair: Irina-Mihaela Ciortan, NTNU (Norway)

10:00 – 10:20

Tweaking Mainstream Open-source OCR Engine for Minority Languages, How To?, *Tuomo Räisänen and Anssi Jääskeläinen, South-Eastern Finland University of Applied Sciences; and Atte Föhr, National Archives of Finland (Finland)* **140**

The digitization of historical documents is vital for preserving cultural heritage, yet mainstream OCR (Optical Character Recognition) systems often fail to support minority languages due to limited training data and language-specific models. This study explores how open-source OCR frameworks can be adapted to overcome these limitations, focusing on Finnish and Swedish as case studies. We present a practical methodology for fine-tuning PaddleOCR using a combination of manually annotated and synthetically generated data, supported by high-performance computing infrastructure. Our enhanced model significantly outperforms both Tesseract and baseline PaddleOCR, particularly in recognizing handwritten and domain-specific texts. The results highlight the importance of domain adaptation, GPU acceleration, and open-source flexibility in building OCR systems tailored for under-resourced languages. This work offers a replicable blueprint for cultural institutions seeking locally deployable OCR solution.

Reconceptualizing Mass Digitization Through a Pragmatic Lens: The Move from Whole Collections to Selective Highlights, *Douglas Emery and Roxanne Peck, Hoover Institution Library & Archives (US)* **A-11**

The Hoover Institution Library & Archives (HILA) embarked on an ambitious mass digitization program, known as the Digital First Initiative (DFI) in 2019. After two years of planning, infrastructure development and hiring staff, the program launched focusing on digitizing whole archival collections and replicating the physical reading room experience online. By 2024, the DFI program had digitized three collections. Through an assessment of DFI's effectiveness, engaging staff in a pilot program to evaluate more efficient workflows along with a shift HILA's vision for digitization, the DFI program transformed from digitizing whole collections to selective highlights. These changes resulted in increased staff productivity, more diverse collections online, improved staff relationships, and clearly established priorities.

State of Open Data in Government Ministries: The Case of the United Arab Emirates, *Martin Critelli, Mehluli Masuku, and Samson Mutsagondo, Sorbonne University Abu Dhabi (United Arab Emirates)* **145**

Open data has become a multidisciplinary concept attracting different players and professionals across the globe. Given its benefits, countries are thriving to provide a conducive environment for such open data. The study examines the extent of open data deployment and presence of open data laws, policies and regulations in government ministries in the United Arab Emirates. This article analyses national and federal policies and regulations concerning open data. The study revealed the presence of a strong regulatory framework for data packaging, reusability, accessibility and the presence of open data in most government ministries. Given that most governments are still grappling with open data implementation, the UAE serves as one of the rich cases for open data deployment by government, deepening access to public data.

Digitizing Cultural Heritage: Exploiting the Gap Between Standards and Individual Approach, *David Stecker, National Gallery Prague (Czech Republic)* **150**

The digitization of cultural heritage focuses on finding a way to record reality as faithfully as possible and to achieve maximum conformity in the representation of the original object in the digital world. To this end, digitization standards and recommendations have been developed that describe best practices for creating a faithful digital representation, ideally a digital surrogate. However, reality cannot be contained in a single digital file; to better describe it we need to create several, each representing different characteristics of the original object. This brings me to the topic of the paper, I want to focus on the space that lies between the original object and its standardized representation. Therefore, at the National Gallery in Prague we are developing a more complex approach that combines standardization with individual imaging. We focus on capturing different aspects of the work using specific lighting techniques and capture methods. We are trying to document these individual approaches as a complement to the standards, thus enriching the metadata of the object. The aim is to disrupt the concept that a single digital document can replace the original, and to show that complex digitization requires a wide range of documented and repeatable approaches that are applied variably to individual objects.

MISHA3D—Three-dimensional Surface Capture with the MISHA Multispectral Imaging System, *James Ferwerda, Juilee Decker, and David Messinger, Rochester Institute of Technology (US)* **155**
In this paper we introduce MISHA3D, a set of tools that enable 3D surface capture using the MISHA multispectral imaging system. MISHA3D uses a novel multispectral photometric stereo algorithm to estimate normal, height, and RGB albedo maps as part of the standard multispectral imaging workflow. The maps can be visualized and analyzed directly, or rendered as realistic, interactive digital surrogates using standard graphics APIs. Our hope is that these tools will significantly increase the usefulness of the MISHA system for librarians, curators, and scholars studying historical and cultural heritage artifacts.

INTERACTIVE POSTER PAPER SESSION

10:20 – 11:30

Engage in meaningful conversations with the authors of the Interactive Papers over coffee and learn more about their work.

DIGITIZATION AND ACCESS

Session Chair: Ty Popko, The Walt Disney Archives (US)

11:30 – 12:15

11:30 **Providing Digital Access to the Freedmen's Bureau,** *Emily Cain, Smithsonian Transcription Center; Hollis Gentry Brown, Smithsonian Libraries and Archives; and Douglas Remley, Jill Roberts, and Kamilah Stinnett, National Museum of African American History and Culture (US)* **159**
The National Museum of African American History and Culture's Freedmen's Bureau Project is a comprehensive initiative that has provided digital access to the Freedmen's Bureau records. Previously, this important collection could only be accessed in person through the National Archives and Records Administration, with no way to search for specific people or topics. Smithsonian staff have worked with the public to index and transcribe the records to provide free full-text access to these invaluable records. To date over 600,000 pages of Freedmen's Bureau records have been collaboratively transcribed by more than 60,000 individual volunteers. This data has been made available to the public for research in the Freedmen's Bureau Search Portal. This groundbreaking search application is the result of more than a decade of data creation, processing, and cleaning; transcription; community engagement; and historical and genealogical research. The work of Smithsonian staff is ongoing and emerging technologies present exciting opportunities to expand access and continue to enable meaningful discoveries.

11:45 **A Matter of Size, Scope, and Significance: Archival Processing and Mass Digitization of the Johnson Publishing Company Archive,** *Steven D. Booth, Getty Research Institute; and Nathan Anderson, Jeanine Nault, and Luis J. Villanueva, Smithsonian Institution (US)* **165**
Acquired in 2019 by a consortium of philanthropic and cultural heritage organizations, the Johnson Publishing Company (JPC) Archive is co-owned by the Getty Research Institute (GRI) and Smithsonian National Museum of African American History and Culture (NMAAHC). Dating from 1942, when John H. and Eunice W. Johnson founded the company, to the 21st century, the JPC Archive contains over 4 million photographs of published and unpublished works documenting the Black experience, some of which were featured in JPC's 14 magazines, most notably JET and Ebony. In addition to the historically significant events and behind-the-scenes moments depicted, the Archive presents an unmatched and unique record of many facets of the life, work, and contributions of Black individuals, communities, groups, organizations, and businesses. Working collaboratively across the United States (from Los Angeles to Chicago to Washington, DC), these two large cultural heritage institutions currently co-steward this collection, with each focusing on their strengths to bring this remarkable and unique collection to the public.

12:00 **Books are Stubborn Things,** *Amy McCrory, The Ohio State University Libraries (US)* **173**
"Facts are stubborn things, and whatever may be our wishes, our inclinations, or the dictates of our passions, they cannot alter the state of facts and evidence." –John Adams
This paper addresses the complexities of book digitization. In many ways, books defy easy categories that would allow the uncomplicated workflows many people assume are possible. I will examine how this unavoidable fact impacts planning and implementation of regularized book digitization, highlighting common problems and describing how they may be resolved or avoided altogether through careful planning. The paper includes a description of how my organization used a pilot project as an opportunity to build a coordinated workflow encompassing multiple activities across several departments within a large university library.

PHOTOGRAPHIC MEDIA: PROCESSING AND ANALYSIS

Session Chair: Eva Valero Benito, University of Granada (Spain)

12:15 – 15:30

12:15 **From Negatives to Positives: Modeling the Photochemical Printing Process,** *Giorgio Trumpy, Norwegian University of Science and Technology, and Ottar A. B. Anderson, Intermunicipal archive of Møre og Romsdal (Norway)* **178**
Most analog color photographs were captured on film negatives. This study presents a scientifically validated workflow for digitizing and inverting color negatives to produce digital positives that closely emulate traditional enlarger prints. A custom imaging system with narrow spectral bands, designed to match the spectral sensitivities of photographic paper, was tested against a conventional digitization method. Final color images were computed based on the spectral densities of the paper's image-forming dyes, simulating the photochemical printing process. Results demonstrate that aligning digitization spectral bands with photographic paper characteristics improves inversion accuracy. This research lays the foundation for enhanced archival preservation of color negatives and provides a method for generating digital positives that closely match the aesthetic of original prints.

12:30 **Digital Image Processing for Identification and Classification of Historic Photoreproductive Processes Used in Architectural and Technical Drawings based on Color Analysis,** *Manto Sotiropoulou and Vasiliki Kokla, University of West Attica (Greece)* **182**
This research explores the application of image color analysis techniques to identify and classify historic photoreproductive processes—such as blueprinting, diazotype, and other early photographic reproduction methods—based on the color signatures they leave on architectural and technical drawings. The objective is to develop a systematic approach for automatically detecting the specific process used in the reproduction of these drawings, which is critical for preservation, restoration, and analysis in historical studies.

Digital microscopy is employed to examine original 20th century photoreproductions from a historical technical company's archive in Greece. The processes examined are cyanotype, both positive and negative, diazotype of black and red color of line and Gel- lithography of black and brown lines. The visual features of photoreproductions are

analyzed using computational pattern recognition techniques that emphasize the color of lines and type of printing process. The findings computational analysis are cross-referenced, and the resulting variables conclude the classification of prints, according to their colors. The results will contribute to the creation of an effective and accurate identification system for both photographic and photomechanical prints.

12:45 – 14:15
LUNCH BREAK ON OWN

14:15 **Exploring Experimental Machine Learning in Film Restoration,** *Fabio Paul Bedoya Huerta, Duplitech (Peru)* **187**
The challenges of film restoration demand versatile tools, making machine learning (ML)—through training custom models—an ideal solution. This research demonstrates that custom models effectively restore color in deteriorated films, even without direct references, and recover spatial features using techniques like gauge and analog video reference recovery. A key advantage of this approach is its ability to address restoration tasks that are difficult or impossible with traditional methods, which rely on spatial and temporal filters. While general-purpose video generation models like Runway, Sora, and Pika Labs have advanced significantly, they often fall short in film restoration due to limitations in temporal consistency, artifact generation, and lack of precise control. Custom ML models offer a solution by providing targeted restoration and overcoming the inherent limitations of conventional filtering techniques. Results from employing these local models are promising; however, developing highly specific models tailored to individual restoration scenarios is crucial for greater efficiency.

14:30 **From Dyes to Digital: A Scientific Reconstruction of Early Colors,** *Chiara Campagnari, Università degli Studi di Milano (Italy), and Alice Plutino, University of Amsterdam (the Netherlands)* . . . **191**
The growing interest in early films highlights the need for their preservation and digital restoration. Scientific methods are essential for analyzing historical coloring techniques and key characteristics for both physical conservation and digital reproduction.

This study applies colorimetric analysis to two early nitrate-based 35mm films—Voleurs de bijoux mystifiés (1906) and Satan fait la noce (1907)—to examine imbibition and au pochoir coloring methods. Spectral transmittance measurements enabled preliminary dye identification, while colorimetric data informed an accurate digital restoration. This approach aims to (1) expand research on film dyes characterization and (2) establish a material-based method for digital color restoration. By reducing handling of analogue materials, it minimizes deterioration risks, ensuring long-term preservation while maintaining accessibility for study and appreciation.

14:45 **Automatic Restoration of Historical Stereoscopic Photographs for 3D Visualization at Scale,** *Dhruva Gowda-Storz and Sarah Kenderdine, EPFL (Switzerland)* **197**
Vast archives of stereographic photographs from the 19th and 20th centuries survive in collections worldwide. While extensively digitized, these artifacts remain largely inaccessible in their intended three-dimensional form. Contemporary stereoscopic displays offer ideal platforms for experiencing these historical media, yet a significant barrier persists: the labor-intensive process of restoring deteriorated stereographs for comfortable viewing. This paper addresses this challenge through two approaches: first, establishing a comprehensive framework for manual stereograph restoration that balances historical authenticity with viewing comfort; second, presenting our ongoing development of an automated pipeline that leverages recent advances in computer vision. Our approach aims to dramatically reduce the time and expertise required for restoration, potentially enabling unprecedented access to historical stereographic archives and facilitating their reintroduction to contemporary audiences through immersive technologies.

15:00 **Current Practices for Autochrome Digitization,** *Yoko Arteaga, Irina-Mihaela Ciortan, and Giorgio Trumpy, Norwegian University of Science and Technology (Norway); and Catlin Langford, Victoria & Albert Museum (UK)* **203**
Autochromes, invented by the Lumière Brothers in 1907, consist of a glass plate, photographic emulsion, and a colour filter made of dyed potato starch granules and carbon. Due to the autochrome's fragile nature and susceptibility to fading and damage, many institutions limit the plates' exposure to light and movement. This highlights the importance of high-quality digitisation to ensure wider public access and preservation. Currently, there are no specific guidelines for digitising autochromes. To address this, a survey was conducted to understand current practices around autochrome digitisation. Additionally, imaging tests were performed to evaluate different methods and provide guidelines. The survey results and experimental findings will inform standardised digitisation approaches.

15:15 **Closing remarks**

CORPORATE MEMBERS AND PARTNERS

STRATEGIC PARTNERS

Adobe

SUSTAINING

Microsoft

xerox

SUPPORTING

Meta

DONOR

APPLIED IMAGE®

COLORADO STATE UNIVERSITY

DXOMARK

FUJIFILM

Image Engineering
MEMBER OF THE NYNOMIC GROUP

ISA
Image Science Associates

NTNU
Norwegian University of Science and Technology

ARCHIVING 2025 APPENDIX A
EXTENDED **ABSTRACTS**

Willem Witsen's Coat: From Restoration to Digital Preservation

Frans Pegt; Rijksmuseum; Amsterdam, The Netherlands

Abstract

This article explores the multidisciplinary process behind the restoration and digitisation of Willem Witsen's painter's coat—an object of both artistic and historical value. Using techniques such as photogrammetry, 3D modelling, and 360-degree photography, the project aimed to digitally preserve the fragile coat while making it accessible to both researchers and the general public. The result is a high-quality digital surrogate that supports future conservation efforts and storytelling.

1. Introduction

Willem Witsen (1860–1923) was a Dutch painter, photographer, and key figure in the literary and artistic movement known as the "Tachtigers." His paint-stained studio coat, preserved in his original workspace in the Oosterpark in Amsterdam, has become a symbol of his artistic legacy. The coat is physically fragile yet rich in meaning—with visible traces of his work and time. Since 2020, parts of the Rijksmuseum's collections have been housed in the Netherlands Collection Centre (CCNL) in Amersfoort, a shared depot with other major cultural institutions. As part of its conservation and documentation mandate, the Rijksmuseum's Imaging & Registration department was tasked with digitally documenting the Witsen coat. The coat is part of the collection of the Cultural Heritage Agency of the Netherlands (RCE) and is on permanent display at the Witsen House. The RCE's conservation department requested a 3D capture of the coat.

2. Reason for Conservation and Digitisation

Preliminary inspections revealed clear signs of degradation: moth damage, weakened fabric, and stress-induced deformations. Before the coat could be safely exhibited, it required treatment and stabilisation by a textile conservator. (I will briefly discuss this in my presentation.) At this point, it was decided to create a digital version, To see whether it is possible to monitor the jacket for research, and presentation in the future. Photogrammetry would allow for detailed recording of the coat's shape and surface without physical contact, thereby preserving its condition and enhancing accessibility. For now, this is an experiment to explore how 3D technology can be used for future object monitoring. Using Texxary Targets and Rulers, I entered scale references into the 3D model to allow for accurate measurements. (I have since used calibrated targets from CHI.)

3. Photographic and Technical Process

3.1 Preparation and Capture

The coat was photographed in a controlled studio environment using a high-resolution Canon camera (50MP), consistent lighting, and a motorised turntable (PhotoRobot). This setup ensured that every angle of the coat was clearly captured, including details such as stitching, wear patterns, and paint stains.

3.2 Photogrammetry and Modelling

The photos were imported into photogrammetry software to generate a 3D model. Challenges arose during this phase: irregularities in the coat's shape due to its deteriorated condition led to mesh distortions, especially around the collar and sleeves. These were corrected through targeted supplementary photography and careful reprocessing, as I discuss in the presentation. The 3D model can now be viewed by researchers on Sketchfab. In the future, the Rijksmuseum plans to migrate models to Voyager—a more robust and flexible viewer that will hopefully be integrated with Picturepark, the museum's digital asset management system. This integration will facilitate better long-term access within the institution—and hopefully on the website as well.

4. 360-Degree Photography and Visual Presentation

To complement the 3D model, a 360-degree photographic sequence was created using high-quality studio lighting. This method presents the object as it appears in photography, with natural shadows and highlights as seen in a photo studio. The resulting images are viewed with Micrio, a high-resolution zoomable image viewer designed for in-depth object exploration. Unlike typical 3D viewers, Micrio enables pixel-level inspection of surface details in a fixed lighting setup, making it especially suited for visual storytelling and object appreciation in educational or public contexts. This combination—Micrio for detailed visual inspection and Sketchfab (or a future viewer) for interactive exploration—demonstrates the value of using complementary platforms for different user needs. An additional benefit of 360-degree photography is that it yields a wealth of high-resolution images from many different angles.

5. Applications and Broader Impact

The Witsen Coat digital project shows how the integration of conservation and imaging technologies can benefit museums and cultural institutions. The 3D model now serves as a non-invasive analysis tool, while the 360-degree Micrio display makes the object accessible to audiences who may never see it in person. The project also lays the groundwork for broader application of imaging workflows linked to collection management systems such as Picturepark. As these tools continue to evolve, they will enable seamless access to high-quality documentation directly from both internal and public museum portals. This process offers a framework for digitising other fragile, historically rich artefacts that should not be frequently handled but deserve to be widely studied and shared.

6. Conclusion

Modern digital imaging technologies offer museums new ways to manage, understand, and share their collections with people all over the world, then think not only about interested visitors but also about having scientists watch from a distance. The preservation of Willem Witsen's coat through a combination of careful restoration, 3D modelling,

Technical Data:

Texxary Coded Target Scale Bars 100mm
Canon 5Dsr 35mm lens
Agisoft Metashape 2.2
Texxary Coded Target Scale Bars 100mm
PhotoRobot Turntable
Sketchfab
Micr.io 360-degree viewer
The coat is part of the RCE collection and is exhibited at the Witsen House in Amsterdam.

A White-Box Machine Learning Model for Dating Archival Materials Using IR Spectroscopy

Hend Mahgoub; Heritage Science Laboratory Ljubljana (HSLL), Faculty of Chemistry and Chemical Technology, University of Ljubljana; Slovenia. Patrick Layton; Academy of Fine Arts, Institute of Natural Sciences and Technology in the Arts; Austria. Sonja Svoljšak; NUL, National and University Library of Slovenia; Slovenia. Jasna Malešič; NUL, National and University Library of Slovenia; Slovenia. Johannes Tintner-Olifiers; Academy of Fine Arts, Institute of Natural Sciences and Technology in the Arts; Austria. Matija Strlič; Heritage Science Laboratory Ljubljana (HSLL), Faculty of Chemistry and Chemical Technology, University of Ljubljana; Slovenia.

Abstract

This study highlights how infrared (IR) spectroscopic techniques, combined with machine learning (ML), can transform the dating of archival materials. By integrating near-infrared (NIR) and Fourier-transform infrared (FTIR) spectroscopy with ML methods, we establish correlations between spectral signatures and paper aging markers and composition. The Ancient Book Crafts (ABC) project demonstrates the practical applications of these techniques, confirming their potential to produce accurate, data-driven dating models. A dataset of 100 well-dated paper objects from the National and University Library (Slovenia) was analyzed to develop predictive models for dating. Initial results show the superior accuracy of NIR in detecting bulk aging effects, while FTIR-ER proves success as a valuable non-contact tool compared to ATR for dating and surveying archival materials. Future advancements in spectral preprocessing, model optimization, and interdisciplinary collaboration will further refine this approach, ensuring widespread applicability across cultural heritage collections and supporting evidence-based conservation strategies.

Introduction

The application of spectroscopic techniques in cultural heritage has expanded significantly, revolutionizing the study and preservation of historical materials. Infrared (IR) spectroscopy, in particular, enables non-destructive analysis, by capturing chemical fingerprints offering insights into their chemical composition, degradation processes, and historical authenticity. When combined with machine learning (ML), spectroscopic data can be processed and interpreted at scale, uncovering hidden patterns and supporting evidence-based conservation decisions.

ML models, such as partial least squares regression (PLSR), decision trees and neural networks, have been successfully applied to spectral data of cellulose-based heritage materials, such as paper [1-3], enabling material classification, identification of degradation markers, and prediction of mechanical stability over time.

One especially valuable application is dating, where spectral signatures linked to oxidation, polymer breakdown, and other age-related chemical changes are used to estimate the age of artifacts. A well-dated representative dataset is crucial for the success of such models. IR spectroscopy is particularly effective for dating of writing support materials [1-3], thanks to its ability to detect molecular-level changes indicative of aging, such as cellulose degradation, gelatinization of parchment, and environmental impacts.

Dating archival materials is essential for understanding their historical context, authenticity, and preservation needs. Traditional methods such as watermark analysis, codicology and dendrochronology are informative but often time-consuming, subjective and require expert interpretation. IR spectroscopy provides a non-invasive, quantitative alternative which makes it a powerful tool for conservators and librarians.

This study presents part of the Ancient Book Craft (ABC) project (2022–2025, N1-0271) [4] an Austrian-Slovenian initiative that integrates IR spectroscopy and ML, supported by historical and codicological analysis, to develop robust dating methods for Medieval archival materials. The project investigates the effectiveness of various IR techniques—including Fourier-transform infrared spectroscopy in attenuated total reflectance (FTIR-ATR), external reflection (FTIR-ER), and near-infrared (NIR) spectroscopy—in capturing age-related chemical transformations, using a well-dated dataset of archival objects from the National and University Library (NUL) of Slovenia and Klosterneuburg Abbey in Austria.

The project also explores the influence of storage environments, material composition, and spectroscopic techniques on the reliability and transferability of these methods. The research is funded through the WEAVE program in collaboration with FWF (Austria) and ARIS (Slovenia) and supported by the Research Programme Group N-DAD.

Methodology

In this paper we highlight the capabilities and limitations of NIR and FTIR spectroscopy in age determination of historical paper from the 15th to mid-19th centuries using a well-dated dataset of 100 paper objects from NUL, covering a wide range of production techniques and storage conditions. Spectral data were collected using three techniques: NIR (1000–2500 nm) and FTIR (4000–400 cm^{-1}) in both ATR and ER modes (Fig. 1). A stratified sampling strategy was adopted, using a 50-year binning approach (sample size = 13 objects per bin). Thousands of measurements were collected from objects in different pages (~5 per object) and locations (~3 per page) to ensure statistical robustness.

Custom R scripts were used for spectral pre-processing and modelling. Various spectral pretreatments were explored individually, and in combination to optimize model performance

such as Standard Normal Variate (SNV), and Savitzky-Golay (SG) smoothing.

This paper focuses on the use of PLSR to develop white-box models; interpretable models whose decisions are based on input features. Models were evaluated using different cross-validation methods. Model performance was assessed using standard metrics such as RMSECV (root mean square error of cross-validation), RMSEP (prediction error) and R^2 (coefficient of determination). To understand the influence of specific spectral features on model performance, Variable Importance in Projection (VIP) plots were also analyzed, identifying the most informative wavelength regions linked to chemical markers of aging and composition.

Results and Discussion

Initial analyses reveal that NIR generally achieved the highest accuracy with a dating accuracy of ~30 years, reflecting its sensitivity to bulk aging effects over time. FTIR also provided comparable predictive performance, though with slightly lower accuracy, likely due to its surface-sensitive measurement characteristics.

PLS models were developed using varying sample sizes for each IR technique and evaluated under different cross-validation strategies. For 100-fold CV, the full dataset was used. However, a consistent underprediction of post-1650 AD samples was observed, suggesting that chronological shifts in papermaking practices may not be well captured by linear models. To address this the datasets were also divided into pre- and post- 1650 AD subsets allowing for more targeted modelling.

To better simulate real scenarios, a customized Leave-One-Book-Out (LOBO-CV) was implemented, treating all pages within a single book as a unit during testing. While this approach did not improve overall model accuracy, it helped prevent overestimation of performance due to intra-book spectral similarity, offering a more realistic assessment of predictive robustness in surveys.

Future work will focus on refining dating models through advanced variable selection techniques and exploring additional machine learning methods such as Random Forest (RF) and K-Nearest Neighbors (KNN). Cross-institutional validation using datasets from other libraries and archives (e.g., Klosterneuburg Abbey, Austria) and the evaluation of environmental influences on aging markers and prediction accuracy will be key to ensuring model robustness and transferability.

Integrating IR spectroscopy and ML offers a scalable, non-destructive approach for dating library and archival materials, making it a valuable tool for conservators, historians, and librarians.

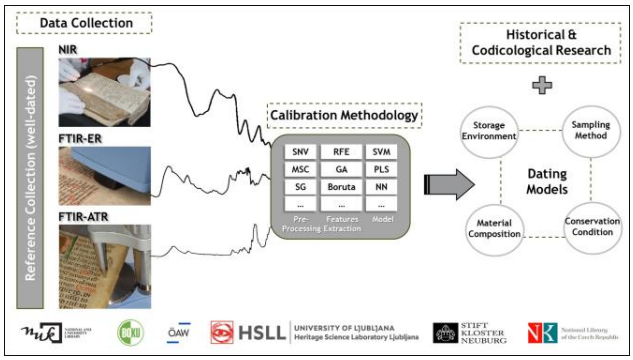

Figure 1. *Research scheme from spectral data collection using IR spectroscopic techniques to predictive dating models.*

References

[1] Trafela et al., Anal. Chem. (2007) https://doi.org/10.1021/ac070392t

[2] Coppola et al., J. Am. Chem. Soc. (2023) https://doi.org/10.1021/jacs.3c02835

[3] Nesměrák & Němcová, Anal. Lett. (2012) https://doi.org/10.1080/00032719.2011.644741

[4] ABC - Ancient Book Crafts - HSLL https://hslab.fkkt.uni-lj.si/2022/04/11/abc-ancient-book-crafts/

Author Biography

Hend Mahgoub: *Post-doctoral researcher at the Heritage Science Laboratory Ljubljana, University of Ljubljana, with a background in computer science and master's and PhD degrees in Heritage Science from UCL, UK. Her research focuses on the study of materials using analytical and spectroscopic techniques. Skilled in spectroscopy, spectral imaging, and data analysis, with strong programming skills.*

Matija Strlič: *Professor of Analytical Chemistry at Heritage Science Laboratory Ljubljana, University of Ljubljana and Honourary Professor of Heritage Science at University College London, UK.*
His research focuses on modelling of heritage materials, environments, values and decision making. He received Slovenia's Ambassador of Science Award (2015) and Zois Award (2012).

Sonja Svoljšak: *Librarian in the Early Printed Books Collection at the National and University Library in Ljubljana (Slovenia). Her main professional and scientific focus is on provenance research, reconstruction of historical collections, management of special collections, bibliographic metadata management, and development of advanced features in library catalogues.*

Jasna Malešič: *Senior advisor, conservator-restorer, and researcher in Heritage Science at the Research Department of NUL, holds a PhD in Chemistry from the University of Ljubljana. With 24 years of experience, she led the Conservation and Preservation Centre for 13 years and managed library preservation for 15 years. An IFLA representative, she contributes to preservation and emergency response guidelines. She specializes in cellulose stabilization and degradation, project management, and interdisciplinary leadership.*

Patrick Layton: *Doctoral student at the Academy of Fine Arts, Vienna, Austria, specializing in leather, parchment, paper, wood and ink for cultural heritage preservation. An experienced book conservator with a demonstrated history of working in the higher education industry. He has a bachelor's degree focused in Parks, Recreation and Leisure Studies from Brigham Young University and a Master's in Cultural Heritage Materials and Technologies from the University of the Peloponnese.*

Johannes Tintner-Olifiers: *studied Land Management and Water Use Engineering and has been a research associate since 2004 and a senior researcher at the Institute of Physics and Material Sciences (BOKU) since 2019. His research focuses on biogenic material transformation under environmental conditions, including wood aging, dating tools based on molecular decay, and newspaper degradation. Skilled in FTIR spectroscopy, thermal analysis, statistics and experimental design, he is a member of BOKU's Centre of Experimental Design.*

AI-Based Indexing for Secure and Efficient Archival Digitalization

Julia Sjöholm, the National Archives of Sweden; Stockholm, Sweden

Abstract

This extended abstract presents a full-scale production system developed by the National Archives of Sweden for large-scale digitization and AI-assisted indexing of over 54 million pages of property-related documents. The system, operational since May 2024, efficiently extracts metadata at scale while ensuring appropriate protection of sensitive archival content, maintaining efficient processing times

The system architecture covers the entire digitization chain: physical document handling, high-speed scanning at 290 images per minute, AI processing, automated validation, and manual review of approximately 10% of the predictions. The AI-based indexing pipeline includes object detection (YOLOv9 with 0.80 precision and 0.90 recall), handwriting recognition (TrOCR with a CER of 0.0145), and automated validation, achieving an 82% efficiency gain compared to manual indexing.

Continuous model development through offline retraining using quality assurance feedback enhances performance over time. Early tests with a Donut model, integrating detection and recognition without bounding boxes, are promising and allow reuse of previously processed data.

The system maintains data sovereignty through internal processing and secure access control, with potential data-sharing collaborations being explored with the Swedish Mapping, Cadastral, and Land Registration Authority (Lantmäteriet). The architecture provides a scalable and reusable framework for public institutions aiming to combine large-scale digitization and internal network security. This document outlines the technical implementation, performance evaluation, and future development strategies.

Introduction

Large-scale archival digitization requires balancing efficient metadata extraction with appropriate security measures due to the sensitive nature of historical documents. The DF project at the National Archives aims to digitize more than 54 million pages of property-related archival documents, ensuring searchable metadata for practical usage. Without effective indexing, digitized documents remain unsearchable images, greatly reducing their usability.

System Architecture

The digitization process consists of multiple integrated stages:

- Physical logistics and document handling using custom-designed carts for safe transport
- High-speed scanning with IBML Fusion HD scanners (up to 290 images/minute)
- Image post-processing, including the removal of blank reverse pages
- Format segmentation depending on usage (TIFF for long-term storage, JPEG for web presentation, grayscale TIFF for AI processing)
- An indexing pipeline consisting of:
- YOLOv9 [1] for detecting case numbers
- TrOCR Base handwritten [2] for handwriting recognition
- Binary classification model (DiT model) [3] for flagging materials for manual review post-validation
- Post-processing and metadata extraction with storage in an internal image/object repository

All processing infrastructure resides in an isolated internal network, using secure storage methods. Access is limited exclusively to personnel cleared by the Swedish Security Service (SÄPO).

Performance Evaluation

A central step in the automated AI-pipeline is a series of validation criteria to determine when an image should be flagged for manual review. For example, the system checks that the case numbers are within the specified range in the archival documentation system and number sequences. If the volume data indicates case numbers 1–500, and a prediction suggests 510, it is automatically flagged. Classification of first pages also affects subsequent indexing—if a new number is detected before the next first page, it's marked as a potential error. Sudden jumps, such as from case number 1 to 10 between two pages, are another flagging criterion.

Initial AI pipeline results:

- YOLOv9: mAP@50: 0.80, Precision: 0.80, Recall: 0.90
- TrOCR: Character Error Rate (CER): 0.0145
- Manual QA Share: approximately 10% of documents

Automated indexing with about 10% of output manually checked saves roughly 82% of the time compared to fully manual indexing by experienced production assistants.

Donut: OCR-free Document Understanding

Even though performance metrics for the initial pipeline looked strong, it was decided early on that a manual review would be conducted ahead of retraining, in preparation for scanning an older set of documents. This decision was made at the beginning of the project to prioritize launching production quickly, with the understanding that re-training could follow later. The first phase focused on scanning and indexing land registration documents from 2001 to 2008, with plans to begin processing records dated 1971 to 2000 in May 2025.

In a manual review of 3,000 images, 3.5% were identified as false negatives due to YOLO not generating predictions. This

highlights a structural limitation in bounding box-based methods. In response, the transition to the Donut model has been initiated.

Donut (Document Understanding Transformer), introduced by Kim et al. [4], offers a novel approach to document analysis by eliminating dependencies on traditional OCR and bounding boxes. It consists of a visual encoder (e.g., Swin Transformer) and a text decoder (e.g., BART) that together convert document images directly into structured text, typically in JSON format. This architecture unifies detection and recognition into a single inference step and enables reuse of previously processed data without the need for bounding box annotations.

Donut is pre-trained on synthetic document images and fine-tuned for specific tasks such as key-value extraction and classification. Its integration into the indexing pipeline is expected to improve both efficiency and robustness, especially when dealing with layout variability or degraded image quality.

A further advantage is Donut's ability to output token-level confidence scores, including for blank or low-information pages. This enhances automated validation logic within the pipeline and reduces the need for manual intervention**.**

Training data is continuously refined based on feedback from manual quality assurance. The goal is to have the Donut-based solution fully implemented by May 2025. However, it is important to emphasize that one must become familiar with the behavior of AI models to understand how to interpret confidence values in practice. It is not always clear which thresholds should trigger manual review and which can be safely accepted. Designing validation criteria is therefore a balancing act: overly strict thresholds reduce the benefits of automation, while overly indulgent criteria risk allowing low-quality data to pass through unchecked.

Model Adaptation and Continuous Learning

An interactive active learning loop is not implemented due to the material's classification level. Instead, an offline strategy with continuous retraining is used. This includes monitoring data drift and retraining models with corrected outputs. Monitoring focuses on how much output is sent to QA.

Plans are in place to fully replace YOLO and TrOCR with the Donut model. This enables reuse of all previously processed data without bounding boxes, simplifying workflows and increasing robustness.

Limited tests with Label Studio as a platform for active learning have been conducted. Since the tool currently does not meet internal security requirements, more extensive testing is ongoing in parallel projects involving non-sensitive material.

Security Framework

The entire system operates within a controlled internal network, with strict access limited to personnel holding SÄPO clearance. Data transfers are handled securely within the internal environment, and all inference runs on internal nodes. Comprehensive logging and audit systems are in place for all search and storage functions. Encryption is applied both at rest and in transit, although implementation details are confidential for operational and security reasons. As it is not possible to predict where sensitive content may appear, all images are treated as potentially sensitive.

Collaboration and Broader Application

Initial discussions with Lantmäteriet indicate that data sharing is legally feasible and and feasibility studies have been conducted in 2016 that showed economic benefits in the form of reduced processing times for both parties. Secure transfer protocols, such as the Swedish Government Secure Intranet (SGSI) or government VPN tunnels, are being investigated. For now, the digitized property documents remain internally accessible only, and Lantmäteriet continues to receive information via traditional request/response methods. The goal is to find a solution before the project concludes in 2026 or 2027 if further funding is granted for the project.

The system architecture and methodology are designed to be reusable in other digitization projects within the National Archives. Other public institutions and archives can adapt the method by:

- Integrating AI directly into scanning workflows
- Hosting training and inference internally to maintain data sovereignty
- Adjusting validation criteria based on metadata and requirement levels

This allows resources to shift from metadata work to increased scanning capacity. It's worth noting that models are material-specific, and each document type requires its own model fine-tuning.

Conclusion

The AI-based indexing system at the National Archives of Sweden demonstrates that it is possible to successfully combine large-scale digitization with appropriate security requirements. The current implementation processes hundreds of thousands of documents monthly with 90% fully automated indexing, representing an 82% efficiency gain over manual methods. This has allowed the Archives to accelerate their digitization timeline while maintaining protection for handling potentially sensitive materials.

The ongoing transition to the Donut model is expected to significantly evolve the system's capabilities. By eliminating the dependency on traditional bounding box methods, Donut's unified detection and recognition approach promises to simplify workflows and potentially enhance model training efficiency. The model could potentially reducing manual review workloads.

Other institutions can adapt this framework by implementing specific components based on their needs:

1. Technical adaptation: Organizations can adopt the containerized Kubernetes workflow approach using commodity GPU hardware, avoiding cloud-based solutions when processing sensitive materials. The documented YOLOv9 and Donut architecture can be repurposed without specialized hardware requirements.
2. Process adaptation: The validation logic framework can be customized with domain-specific rules. For example, an institution digitizing demographic records could implement validation checks based on ID numbers rather than case numbers.
3. Security scaling: The multi-tiered security approach can be scaled according to classification requirements - from fully air-gapped systems for highly sensitive material to network-isolated but internally connected systems for less sensitive content.
4. Workforce integration: The 10/90 QA ratio provides a reference point for staff resource allocation, with specialized roles for security clearance, model training, and validation that can be adapted to existing archival workflows.

Looking ahead, the focus remains on completing the Donut-based solution by May 2025 and developing secure APIs for inter-agency access to digitized materials, with specific emphasis on the collaboration with Lantmäteriet.

By sharing this experience, the National Archives of Sweden contributes to the broader discussion on how AI can transform archival practices while respecting the unique security considerations that historical documents often require.

References

[1] I.-H. Y. H.-Y. M. L. Chien-Yao Wang, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," *ArXiv,* p. 18, 29 Feb 2024.

[2] T. L. J. C. L. C. Y. L. D. F. C. Z. Z. L. F. W. Minghao Li, "TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models," *ArXiv,* p. 10, 6 Sep 2022.

[3] Y. X. T. L. L. C. C. Z. F. W. Junlong Li, "DiT: Self-supervised Pre-training for Document Image Transformer," *arXiv,* p. 10, 19 july 2022.

[4] T. H. M. Y. J. N. J. P. J. Y. W. H. S. Y. D. H. S. P. Geewook Kim, "OCR-free Document Understanding Transformer," *ArXiv,* p. 29, 6 October 2022.

Author Biography

Julia Sjöholm is a project manager at The Swedish National Archives, with a background as an archivist and recent training in AI development with a focus on machine learning from a vocational university.

Their work in the DF project involves exploring AI-driven indexing to improve the efficiency of archival digitalization and metadata generation.

Digital Object Authenticity: Creating a Standard Practice

Julie McVey, National Geographic Society, Washington, DC, USA; Doug Peterson, Digital Transitions, New York City, NY, USA; Ottar A.B. Anderson, Intermunicipal archive of Møre og Romsdal, Ålesund, Norway

Abstract

Galleries, Libraries, Archives, and Museums do not currently have a standardized, simple way to communicate digital surrogate quality to users through standard, shared metadata fields. Creating these shared standard fields and implementing content authenticity tools in an agreed-upon and widely adopted data model will allow for greater transparency and institutional trust. The Digital Object Authenticity Working Group (DOAWG) aims to work with existing organizations and standards commonly used in the heritage field to create a cultural heritage object authenticity standard. We believe this goal is best accomplished by bringing cultural heritage professionals together to facilitate a timely and thorough approach to establishing this standard practice in institutions across the world.

Motivation

Galleries, libraries, archives, and museums (GLAM) play an important societal role in stewarding shared cultural objects and histories, and for the past several decades have worked toward increasing access and practicing preservation through the foundation of high-quality digital imaging programs. As our abilities to create accurate and complete digital surrogates of physical objects have increased due to advances in lighting and camera technology, the professionalization of the cultural heritage imaging field and its standards have developed further as well. For more than a decade, imaging professionals have been working together to create a shared understanding of image quality and what makes an acceptable and authentic digital object for various types of uses as a digital surrogate of the original. Additionally, the field has also long worked on standardizing metadata and advocated for best practices with regards to interoperability, sharing data, and increasing the usability of digitized collections through digital asset management systems (DAMS) and discovery platforms.

As adoption of shared technical and accessibility standards increases and more collections are made discoverable online, institutions are also considering questions of content authenticity and verification. Beginning in 2021, the Coalition for Content Provenance and Authenticity (C2PA) [1] has been working toward developing an open standard that addresses the need for authenticating and verifying digital content. These efforts, led by several large technology companies with a major focus on born-digital content in the journalistic fields of photography and videography, have spurred a discussion in the cultural heritage imaging field around what authenticity means for our work. Many institutions are surfacing use cases around authenticity not only for born-digital content, but around digital surrogates that are created to represent the valuable physical objects held in the public trust by galleries, libraries, archives, and museums worldwide.

Problem

In addition to establishing and protecting more traditional provenance metadata such as creator, date, geographic location, and applied edits, cultural heritage imaging has a significant interest in demonstrating nature and quality of the imaging process. If, for instance, a researcher accesses a collection of maps in an archive's online discovery interface to establish evidence for a tribal land claim, they do not currently know if the digital surrogates they are accessing are of the highest quality and therefore show all of the possible information contained in the physical object. They also do not have a guaranteed way to verify important metadata fields such as creator, original object date, source institution, or others if the image is downloaded and used separately from the institution's discovery system. While C2PA is actively working to address the provenance and veracity of the latter types of information, there is no agreed-upon way to share or to verify image quality information using common embedded metadata fields.

This gap has been discussed in various professional gatherings, and beginning in 2024, a small group of three cultural heritage professionals began meeting to discuss content authenticity and demonstrating image quality in embedded metadata. This interdisciplinary group brings perspectives from the imaging equipment and software industry, photography and imaging workflows, and digital asset management and discovery. In analyzing the problem of relaying and securing relevant metadata, the gap we have identified is not one of interest or expertise, but in widespread agreement about best practices, workflows, and platforms. The Digital Object Authenticity Working Group (DOAWG) [2] has therefore formed to coordinate efforts across disciplines to determine what additional metadata should be available, how those fields should be structured, and how the cultural heritage imaging community across GLAM institutions authenticates and verifies digital object quality.

Approach

Cultural heritage institutions already have a plethora of standards used to apply metadata to records and their related digital objects: IPTC metadata [3] is embedded to more securely keep track of provenance, IIIF manifests are created for interoperability purposes, and most institutions

implement a form of core descriptive metadata that provides a controlled list of types and formats to inform the user about what kind of object surrogate they are viewing. The C2PA technical specification is currently the most well-developed set of open-source tools and standards for establishing provenance data for digital objects, and is in the process of adoption by the ISO [4]. Companies such as Adobe, Capture One, Microsoft, and others are beginning to integrate these tools into their editing software suites. Because these softwares are widely used by the cultural heritage imaging community, we expect C2PA to emerge as the leading solution to implementing content authenticity standards for the majority of institutions. The C2PA group is also in conversations with IPTC about representing certain pieces of IPTC-standard image metadata within the C2PA framework, thus making these organizations a sensible entry point for working to facilitate widespread adoption of newly standardized metadata fields pertaining to digital objects.

To create a workable solution, DOAWG aims to work with existing organizations and existing standards toward the goal of interoperable display of visual assets and their metadata, to standardize embedded metadata in visual objects, to simplify the relay of image quality and quality control processes, and to underlie these fields with authenticity and provenance security measures. Throughout 2024, DOAWG members worked to establish a baseline understanding of authenticity needs within the heritage community by introducing and participating in conversations about C2PA technologies and how they might relate to common digitization practices and workflows.

In May of 2024 Ottar A.B. Anderson presented at the 2and3D Photography conference at the Rijksmuseum in Amsterdam with a preliminary data model for representing imaging process metadata using IPTC fields [5]. Anderson also began communication with the IPTC Photo Metadata Working Group to formally propose some adaptations in the IPTC-standard specific for the GLAM institutions. The meeting successfully established a common understanding of how GLAM teams could more accurately use existing metadata fields with proposed changes and set the stage for development of new fields to contain imaging process metadata.

Early in September the meeting was successfully carried out and established a strong common understanding between the two working groups. The IPTC NewsCodes 2004 Q3 update [6] was published by the end of September resulting in several of the proposed adaptations being implemented and with this bringing the established professional relationship between the two working groups even closer.

In September of 2024, Julie McVey and Doug Peterson, with significant contributions from Anderson, co-authored an introductory white paper [7] detailing definitions of authenticity and provenance for cultural heritage digital objects and connecting them to the authenticity framework that the Content Authenticity Initiative advocates. McVey and Anderson also presented at the DT Heritage Roundtable, along with Santiago Lyon of Adobe, to communicate the current state of the field and the tools that are currently developed and deployed in various workflow softwares [8][9][10].

Throughout the end of 2024 and into early 2025, DOAWG has continued to communicate with the IPTC Photo Metadata Group, the International Image Interoperability Framework (IIIF) consortium, and various C2PA members to refine understanding of how imaging process metadata fields and C2PA security measures can address the needs of various types of GLAM institutions and their use cases. This work continues, and in June 2025 Anderson and McVey will present to the IIIF annual meeting in Leeds, UK to propose integration of their proposed IPTC fields and C2PA measures into the IIIF display and delivery platform.

Results

Our work to date represents the initial phase of a multi-step process addressing digital object authenticity in cultural heritage contexts. Through stakeholder interviews and technical review with heritage professionals, we discovered significant variation in how institutions conceptualize and implement quality assurance in their digitization workflows. These conversations highlighted that while most institutions maintain internal quality standards, there exists no standardized method for communicating these practices to end users.

We then confirmed this gap by engaging with technology stakeholders including C2PA, IPTC, IIIF, and Digital Transitions. These discussions verified that no comprehensive solution currently exists but clarified that an effective solution could be built leveraging existing technologies and extending metadata frameworks and standards for image quality such as ISO 19264 and FADGI.

Conclusion

With this understanding, the next step is to open lines of communication with the broader cultural heritage community to seek input on what an ideal solution should encompass. By facilitating dialogue between diverse stakeholders, we aim to gather comprehensive requirements that address the unique needs of galleries, libraries, archives, and museums.

Our ultimate goal is to develop and propose a specific solution to technology stakeholders that effectively meets the heritage community's needs for demonstrating and verifying image quality while maintaining compatibility with existing authenticity frameworks and workflow practices.

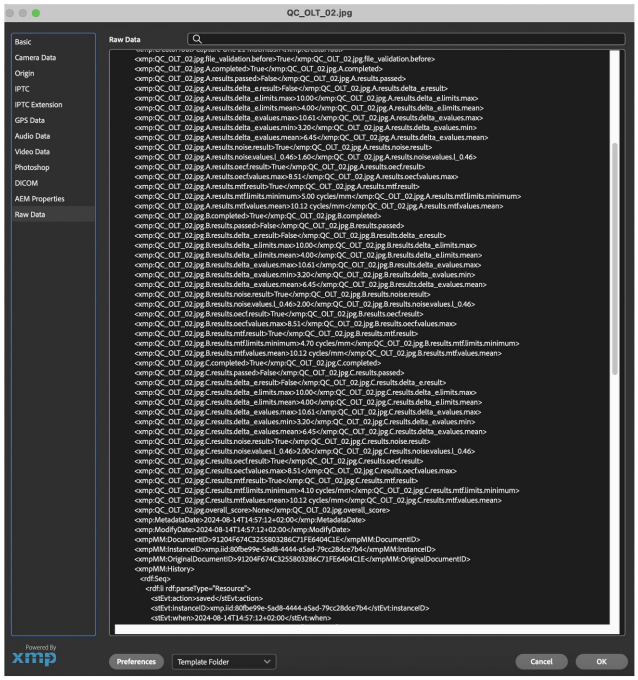

***Figure 1**. Image quality metadata stored in the Raw Data field using XMP format, screen shot from Adobe Photoshop* [11]

References

[1] Coalition for Content Provenance and Authenticity, "C2PA Specifications." c2pa.org. https://c2pa.org/specifications/specifications/1.3/explainer/Explainer.html (accessed Feb. 20, 2025).

[2] Digital Object Authenticity Working Group, "About DOAWG." doawg.org. https://www.doawg.org/ (accessed Feb. 27, 2025).

[3] International Press Telecommunications Committee, "IPTC Photo Metadata Standard." iptc.org. https://iptc.org/standards/photo-metadata/iptc-standard/ (accessed Feb. 20, 2025).

[4] ISO TC 171, "Authenticity of information—Content Credentials." iso.org. https://www.iso.org/standard/90726.html?browse=tc (accessed Feb. 20, 2025).

[5] O.A.B. Anderson, May 30, 2024. "Paradigm shift in the technical metadata structure?" Presented at 2and3D Photography 2024, Amsterdam, Netherlands. [Online] https://2and3dmagazine.rijksmuseum.nl/2024/2and3d-photography-2024#Downloads.

[6] IPTC NewsCodes Working Group, "NewsCodes 2024 Q3 release including Media Topics and Digital Source Type Updates." iptc.org. https://iptc.org/news/newscodes-2024-q3-release-including-media-topics-and-digital-source-type-updates/ (accessed Feb. 20, 2024).

[7] J.M. McVey and D.E. Peterson, "Content Authenticity Initiative for Heritage—A Primer." white paper, https://docs.google.com/document/d/1fbDx633NXaALMD5HWlw_PsiE4jeiLnxxcKvSJmuBtII/edit?tab=t.0#heading=h.tjp3eb37d4xy (accessed Feb. 15, 2025).

[8] S. Lyon, *The Content Authenticity Initiative*. (Oct 2024). Accessed: Feb. 15, 2025. [Streaming video]. Available: https://vimeo.com/1027400413.

[9] J.M. McVey, *What is Authenticity in Cultural Heritage?* (Oct. 2024). Accessed: Feb. 15, 2025. [Streaming video]. Available: https://vimeo.com/1027410877.

[10] O.A.B. Anderson, *Authenticity via Image Quality*. (Oct. 2024). Accessed: Feb 15, 2025. [Streaming video]. Available: https://vimeo.com/1027696708.

[11] M.W. Holtmon, J.F. Karlsmoen, M.F. Krog, "Automated Quality Assurance of Digitization in the Digital Archives." Bachelor thesis, Dept of Computer Science, NTNU, 2022. [Online]. Available: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3002862?show=full.

Author Biography

Julie McVey is Director of Digital Archives for the National Geographic Society's Special Collections. She oversees the digital preservation program and contributes expertise in digital curation, metadata standards, technology innovations, and collections accessibility and outreach. She holds an MA in History and MLIS from the University of Maryland, College Park.

Doug Peterson is Co-Owner and Head of R+D at Digital Transitions. He oversees DT's published technical guidelines and Digitization Certification training series. He is a member of the ISO and sits on TC 42, which works on digitization standards. He holds a BS in Commercial Photography from Ohio University.

Ottar A.B. Anderson is Head of Photography at Intermunicipal archive of Møre og Romsdal in Norway (IKAMR). Former photography background from the Royal Norwegian Air Force and Kodak Q-lab photo technician. He is a member of the ISO and sits on TC 42, which works on digitization standards.

Reconceptualizing Mass Digitization Through a Pragmatic Lens: The Move From Whole Collections to Selective Highlights

Douglas Emery, Roxanne Peck; Hoover Institution Library & Archives, Stanford University; Stanford, CA

Abstract

The Hoover Institution Library & Archives (HILA) embarked on an ambitious mass digitization program, known as the Digital First Initiative (DFI) in 2019. After two years of planning, infrastructure development and hiring staff, the program launched focusing on digitizing whole archival collections and replicating the physical reading room experience online. By 2024, the DFI program had digitized three collections. Through an assessment of DFI's effectiveness, engaging staff in a pilot program to evaluate more efficient workflows along with a shift HILA's vision for digitization, the DFI program transformed from digitizing whole collections to selective highlights. These changes resulted in increased staff productivity, more diverse collections online, improved staff relationships, and clearly established priorities.

Digital First Initiative (DFI)

In early 2021, after more than two years of planning, staff at The Hoover Institution Library & Archives (HILA) began digitizing its first collection as part of the Digital First Initiative (DFI). The DFI program was conceived of as an ambitious plan to digitize entire collections and make them freely accessible for researchers anywhere in the world through the creation of a preeminent virtual reading room. The multi-year process leading up to the DFI launch included building physical and digital infrastructure, recruiting and retaining photographers, conservation staff, archivists and digital specialists to support the newly created in-house mass digitization program (DFI) that would enable newly digitized physical collections to be digitally preserved and made available online.

Assessing DFI

By 2024 we had completed our third collection under the DFI program. This comes out to ~600 archival boxes digitized and ~500,000 digital images created by a staff of 3 full time photographers. There had been no preset plan to assess the program at a specific date. However, as part of a broad review by senior leadership, DFI strategies were reevaluated, and they concluded it was taking too long to get material digitized and online. Another issue found in the review by senior leadership was the inclusion of multiple decision makers in the process. It was hard for people to reach consensus given the perceived competing priorities between units. Additionally, after reviewing 2 years of data, they found the program intermittently experienced drops of up to 50% in monthly imaging output in addition to delays from the point of imaging to the item being made available to researchers online for up to 39 weeks. This was understandable because the organization had never done mass digitization before and had no firsthand experience about managing large scale rapid capture digitization at this scale.

Senior leadership looked to have a mass digitization program with consistent and sustainable output given the limited equipment and staffing resources. With a drop in imaging productivity and a lag in the time content is available for researchers already identified, a key part of our post DFI digitization planning was to define realistic and sustainable metrics across the digitization workflow from beginning to end. The directive to set up metrics came from senior leadership, but it was up to each unit to decide them. To create a sustainable mass digitization program, it was also critical for senior leadership to define the most important priorities for the program: speed and access. Setting clear priorities would help to serve as a guiding north star so when competing priorities emerged, the priorities would allow staff to have difficult conversations and make tough decisions.

Testing faster procedures

The review by senior leadership and a renewed focus on speed and access served as a springboard to create a pilot program in spring 2024 that sought to find efficiencies to our work by encouraging simplification for each step of the digitization process across all units. Each unit (Conservation, Description, Digital Imaging) identified different approaches to test that could streamline the workflow. The spring pilot program evaluated ten different procedures across all units. Procedures tested by Conservation and Description units included minimizing conservation repairs, minimal archival processing and no longer atomizing folders.

Digital Imaging testing focused on three key areas: FADGI specifications, elimination of unnecessary steps, and setting metrics. We had started DFI by following FADGI 4-star specifications. We were creating complete digital surrogates of every 2D item, including blank backs of photocopies. We also included an object level target in every final image, which slowed down imaging for every capture because it meant the photographer had to both center each item to the target and make it parallel. We had three QC processes. And because HILA had never engaged in mass digitization before, no baseline existed for daily throughput, making it difficult to gauge the productivity of the photographers.

Establishing metrics

After six months of testing and discussion each unit implemented changes. In Digital Imaging, we stopped imaging blank backs, we stopped including the object level target in the final images, we changed our imaging specifications from FADGI 4-star to FADGI 3-star. And we established a requirement of 800 captures per day. This led to a doubling of imaging production. Photographers fill out a form to record their work in detail. This allows us to track metrics and performance. If performance falls below expectations, then corrective action is taken to help the photographer get back on track.

To create a baseline for photographers, we took the average daily capture of the highest performing photographer and the averages of the other two photographers to arrive at a mean of

800 captures per day. The Digital Imaging manager led live demonstrations on how to efficiently work to achieve higher productivity and to set expectations. Setting expectations around a consistent, steady, and continuous approach to imaging and QC is the most important part to achieving sustainable rapid imaging. We set up a process of doing capture and QC for 30 minutes at a time with 5-minute breaks in between. Proper ergonomics is critical as well for maintaining productivity and is greatly improved with adjustable height tables throughout the lab and Hag Capisco chairs designed for dynamic movement.

Selective highlights

As the pilot program changes were integrated into the mass digitization procedures, HILA's Director made a notable change in his vision for HILA's mass digitization programming. We were going to stop imaging whole collections and to selectively digitize only the material that has significant research value as decided by our curators and for which intellectual property rights were transferred from donors to HILA. During the first three years of DFI, HILA went through two legal cases while not related to the mass digitization workflow nevertheless led to a reversal in our approach to whole collection digitization. We aimed to streamline our mass digitization processes while also addressing copyright challenges. This was and continues to be complex, as archival collections often include copyrighted materials like newspaper clippings, personal correspondence, and legal documents. Copyright considerations were unavoidable and had to be integrated into every stage of the digitization process—without becoming burdensome—from curators and archivists to post-imaging workflows.

At a surface level, this pivot to selective digitization could be viewed as a slow high-touch workflow that is the antithesis of efficiency and not a true mass digitization operation. At our institution, we define mass digitization by consistent output volume—producing the same number of images in the same amount of time, but from many collections rather than a few. A key concept that will help ensure success is shifting the mindset to digitize the highlights now and digitize additional content, if needed, later. This approach makes more content available sooner, rather than waiting years for a collection to be completely digitized and published online. Selective digitization also allowed us to develop a flexible workflow that brought multiple benefits to the organization; improved communication and working relationships with curators, staff adapting by reimagining efficiency within the constraints of copyright, increased capacity to diversify our digital holdings by building a large corpus of material from many collections instead of just a few, establishing metrics, and reduced legal risks - all of which serve our primary goal of making more material available online for researchers.

Conclusion

When DFI launched—with a dual focus on digitizing entire collections and creating a virtual reading room, it prioritized rigid, formal processes and replicating the in-person experience over providing fast access to collections for researchers. The creation of a virtual reading room is still an elusive goal that is currently on pause. That initial focus of DFI on formality and capturing everything led to a digitization workflow not suited for rapid digitization. We were able to successfully transition our mass digitization program through the combination of a review of DFI's processes and effectiveness and creating an environment that supported a culture of experimentation from staff in the 2024 pilot program. Mass digitization programming is currently curator driven and is primarily focused on digitizing selective highlights. However, HILA will still consider digitizing whole collections, at curator's request, especially those that are in the public domain or where copyright for all collection items was transferred to HILA by the donor. HILA's transition to selective digitization in conjunction with implementing the lessons learned as part of our spring 2024 pilot program, we are now delivering content faster because we are prioritizing selective, high research value content over whole collections. This means greater digital collection diversity is available sooner rather than later.

Author Biography

Douglas Emery became the Head of Digital Imaging at the Hoover Institution Library & Archives in 2023 where he oversees a team of photographers and manages multiple imaging workstreams. Previously, he had extensive experience with mass digitization projects, fine art photography and working as a producer for artist Taryn Simon.

Roxanne Peck has over 20 years of experience in academic libraries including UCLA, Penn State, UC San Diego and the University of Connecticut. Currently she is the Assistant Director, Hoover Institution Library & Archives where she oversees the operational units which include Digital Imaging, Description, Preservation, Engagement and Research Services

NOTES

AUTHOR INDEX

[1] JIST 69(2)[1]:
DOI: 10.2352/J.ImagingSci.Technol.2025.69.2.020401

[2] JIST 69(2)[2]:
DOI: 10.2352/J.ImagingSci.Technol.2025.69.2.020402

Sponsored by the Society for Imaging Science and Technology
7003 Kilworth Lane • Springfield, VA 22151 USA