FINAL **PROGRAM** AND PROCEEDINGS

ARCHIVING2019
Digitization, Preservation, and Access

May 14-17, 2019 • Lisbon, Portugal

General Chair: Lukas Rosenthaler, University of Basel (Switzerland)

www.imaging.org/archiving

Sponsored by the Society for Imaging Science and Technology

Manuscripts are reproduced from PDFs as submitted and approved by authors; no editorial changes have been made.

Printed in Portugal.

Cover image courtesy of Peter Burns.

WELCOME TO ARCHIVING 2019

On behalf of the program's organizing committee, I'm happy to welcome you to Lisbon, a beautiful city of great cultural importance—past and present. This is the perfect place for a conference that is about the long-term preservation of the testimonies of the past, using the most modern methods of the digital age to safe guard them. As one of the oldest cities in the world—inhabitants date to Neolithic period—Lisbon itself is a wonderful symbol of the link between the past and the future. This 16th edition of the Archiving Conference offers an engaging program that will give you many new ideas and the energy to explore them.

A long-standing tradition of this event, the short courses on Tuesday May 14th offer practical introductions and further training in modern methods for capturing images, processing image data, providing access to images and their metadata, and optimal workflow organization.

The opening of the conference on Wednesday May 15th starts with a panel on the digitization of Portuguese photography collections followed by a session on standards and guidelines, an extremely important topic. The keynote given by Jonas Palm—who has attended all 15 previous Archiving Conferences, since the very first one in San Antonio, Texas in 2004—addresses how the field has changed during the past 15 years. The rest of the day is devoted to the important topics of standards, guidelines, and efficient and reliable workflows.

Our second day opens with a keynote from David Taubmann, one of the fathers of the JPEG2000 file format. Due to the IIIF standard, the JPEG2000 is becoming more and more popular in the archiving community, but it is basically still shrouded in mystery. I am sure that this keynote will reveal the secrets of JPEG2000 to a large extent. The following technical sessions are devoted to archival model, workflows, and preservation formats and access. Last but not least, the wonderful Behind-the-Scenes Tours offer the opportunity to visit local cultural heritage institutions and learn about their collections, workflows, and mandates. These are always a highlight of the Conference.

We end the week with a keynote from Jean-Yves Le Meur, the Digital Memory project leader of CERN, the largest international scientific research infrastructure of the world. CERN is known to produce several petabytes of data within fractions of a second, which brings long-term archiving to a different scale. And CERN also deals with images . . . a highlight of this talk.

The day continues with the interactive poster session—another highlight of the conference—where we get to meet authors, discuss new ideas, make new connections, and socialize with fellow archivists and researchers. The morning concludes with a session about best practices, case studies, and lessons learned; I believe we all can learn a lot from the experiences of our colleagues. The conference concludes with a half day focus on access, new technologies, and workflows. Ultimately, preservation only makes sense if in the end our assets can be used and re-used by others. Access, capture, and processing is changing dramatically with new technologies; this session gives us clues as to where the journey might take us.

It's an exciting program that Archiving 2019 has to offer and I'm looking forward to meeting you in wonderful Lisbon!

—*Lukas Rosenthaler, General Conference Chair, Archiving 2019*

CONFERENCE COMMITTEE

General Chair
Lukas Rosenthaler, University of Basel (Switzerland)

Program Chair
Jeanine Nault, Smithsonian Institution (US)

Short Course Chairs
Walther Hasselo, Heritage Leiden (the Netherlands)
David Wyble, Avian Rochester, LLC (US)

AV Chairs
Alexandre Leão, Federal University of Minas Gerais (Brazil)
Jérémie Gerhardt, IRYStec Software Inc. (Canada)

Technical Program Committee
Michael J.Bennett, University of Connecticut (US)
Martina Hoffmann, consultant (the Netherlands)
Erik Landsberg, Cultural Heritage Digitization Consulting (US)
Christoph Voges, Hochschule für ange-wandte Wissenschaft und Kunst (HAWK), and consultant (Germany)

Steering Committee
Ulla Bøgvad Kejser, Det Kongelige Bibliotek/The Royal Library (Denmark)
Peter Burns, Burns Digital Imaging (US)
Suzanne E. Grinnan, IS&T (US)
Lukas Rosenthaler, University of Basel (Switzerland)
David Walls, US Government Publishing Office (US)
Don Williams, Image Science Associates (US)

PAPER REVIEWERS

Nora Armani, SR Socially Relevant™ Film Festival (US)
Kit Arrington, Library of Congress (US)
Philippe Bayle, i2s (France)
Michael J. Bennett, University of Connecticut (US)
Jon Blundell, Smithsonian Institution (US)
Peter Burns, Burns Digital Imaging (US)
Antenie Carstens, National Library of South Africa (retired) (South Africa)
Anne Cherif, i2s (France)
Michael Conyers, The Arc/k Project (US)
Terrance D'Ambrosio, NEDCC (US)
Shannon Davis, Washington University in St. Louis (US)
Daniel Dennehy, Minneapolis Institute of Art (US)
Roger Easton, Rochester Institute of Technology (US)
Chris Edwards, J. Paul Getty Trust (US)
Susan Farnand, Rochester Institute of Technology (US)
Peter Fornaro, University of Basel (Switzerland)
Rebecca Frank, University of Michigan (US)
Jérémie Gerhardt, IRYStec Software Inc. (Canada)
Steffen Hankiewicz, intranda GmbH (Germany)
Walther Hasselo, Heritage Leiden (the Netherlands)
Lei He, Library of Congress (US)
Kurt Heumiller, The Museum of Modern Art (US)
Martina Hoffmann, consultant (the Netherlands)
Michael Horsley, National Archives and Records Administration (US)
Anssi Jääskeläinen, South-Eastern Finland University of Applied Sciences (Finland)

Piet Janssens, Royal Library of Belgium (Belgium)
Robert Kastler, The Museum of Modern Art (US)
Ulla Bøgvad Kejser, Det Kongelige Bibliotek/The Royal Library (Denmark)
Julia Kim, Library of Congress (US)
Erik Landsberg, Cultural Heritage Digitization Consulting (US)
Rimvydas Laužikas, Vilnius University (Lithuania)
Bertrand Lavedrine, CNRS (France)
Christopher Lee, University of North Carolina (US)
Aliza Leventhal, consultant (US)
Julia Lovett, University of Rhode Island (US)
Anne Mason, National Archives and Records Administration (US)
Phil Michel, Library of Congress (US)
Brian Miller, Full Circle Fine Art Services (US)
Dharitri Misra, National Library of Medicine (US)
Kate Murray, Library of Congress (US)
Taren Ouellette, Library of Congress (US)
Jonas Palm, Riksarkivet/Swedish National Archives (Sweden)
Hugo Quisbert, ArkivIT (Sweden)
John Rees, National Library of Medicine (US)
Thomas Rieger, Library of Congress (US)
Lynda Schmitz-Fuhrig, Smithsonian Institution Archives (USA)
David Schuster, Binghamton University (US)
Bethany Scott, University of Houston Libraries (US)
Kate Sherwood, Smithsonian Institution (US)
Steven Simske, Colorado State University (US)
Dina Sokolova, Columbia University (US)
David Stover, Virginia Museum of Fine Arts (US)
Larry Telford, FamilySearch International (US)
Corine van Dongen, Koninklijke Bibliotheek (the Netherlands)
Hans van Dormolen, Hans van Dormolen Imaging & Preservation Imaging (HIP) (the Netherlands)
Bruno Vandermeulen, KU Leuven Libraries (Belgium)
Marie Vans, HP Inc. (US)
Christoph Voges, Hochschule für ange-wandte Wissenschaft und Kunst (HAWK), and consultant (Germany)
David Walls, US Government Publishing Office (US)
Doretha Williams, National Museum of African American History & Culture (US)
Meghan Wilson, Library of Congress (US)
David Wyble, Avian Rochester, LLC (US)

COOPERATING SOCIETIES

- American Institute for Conservation Foundation of the American Institute for Conservation (AIC)
- ALCTS Association for Library Collections & Technical Services
- Coalition for Networked Information (CNI)
- Digital Library Federation
- Digital Preservation Coalition (DPC)
- IOP/Printing & Graphics Science Group
- ISCC — Inter-Society Color Council
- Museum Computer Network (MCN)
- The Royal Photographic Society

SOCIETY FOR IMAGING SCIENCE AND TECHNOLOGY

TECHNICAL PAPERS PROGRAM
CONFERENCE SCHEDULE AND TABLE OF CONTENTS*

TUESDAY 14 MAY 2019

SHORT COURSE PROGRAM

8:45 – 10:45 (2 hours)
SC1A: Introduction to Color Measurement for Archiving
Instructor: David R. Wyble
SC1C: JHOVE 101: Open Source File Format Validation
Instructors: Carl Wilson and Martin Wrigley
SC1D: Introduction to IIIF
Instructor: Peter Fornaro

8:45 – 13:00 (4 hours)
SC1B: Introduction to Digital Image Processing
Instructor: Christoph Voges

11:00 – 13:00 (2 hours)
SC2A: Advanced Concepts in Color Measurement
Instructor: David R. Wyble
SC2B: Spectral Imaging and Technical Aspects
Instructors: Fenella G. France and Meghan Wilson
SC2C: Quality Assurance Workflows for Digitization Projects
Instructor: Martina Hoffman

14:15 – 16:15 (2 hours)
SC3A: Scanner & Camera Imaging Performance: Ten Commandments
Instructors: Peter Burns and Don Williams

14:15 – 17:15 (3 hours)
SC3B: Spectral Image Processing
Instructors: Fenella G. France and Meghan Wilson

14:15 – 18:30 (4 hours)
SC3C: End-to-End Digitization Workflow: Goobi-to-Go for Newbies
Instructors: Steffen Hankiewicz and Jan Vonde
SC4B: Data Compression Formats for Digital Archiving
Instructor: Christoph Voges

16:30 – 18:30 (2 hours)
SC4A: Introduction to Color Management for Cultural Image Capture
Instructors: Don Williams and Peter Burns

17:30 – 18:30 (1 hours)
W4D: Workshop Echoes
Instructors: Walther Hasselo and Olav Kwakman

ARCHIVING 2019 WELCOME RECEPTION
18:30 – 20:00
Casa do Lago, Campo Grande 294
Join colleagues following the short course program. Casa do Lago is an 8- to 10-minute walk from Edificio U, Universidade Lusófona.

WEDNESDAY 15 MAY 2019

PHOTOGRAPHIC ARCHIVING IN PORTUGAL
Session Chairs: Lukas Rosenthaler, University of Basel (Switzerland) and Jeanine Nault, Smithsonian Institution (US)
9:00 – 10:30

9:00 **Welcome Remarks**
IS&T and Archiving 2019 leadership
Silvestre Lacerda, director-general of the Directorate-General for Books, Archives, and Libraries (DG/LAB), and director of the Torre do Tombo National Archive

9:10 **Implementing Photographic Preservation Standards in Portugal**
Antonio Manuel Pinto Morais, president of IPF ISO Technical Committee - Portugal CT 174 Photography

9:20 **Portuguese Projects on Photography Collections Digitization Panel**
Moderator: Luis Pavão, director, LUPA, Luis Pavão Lda
Panelists:
- Carla Marisa da Silva Barro, Digitalização e Reprodução at DGLAB / Portuguese Center for Photography (CPF), will present "Digitization: One tool for photographic collections" and discuss the important role of digitization in the access and management of photographic collections and their experience.
- Marco Daniel Duarte, head of the Fatima Shrine's Research Department, and André Melícias, coordinator of the Fatima Shrine's Archives and Library on "Digitizing the visual memory of a global Shrine"
- Paulo Leitão, head of information systems, Calouste Gulbenkian Foundation Art Library
- Isabel Corda, head, and Luis Pavão, photography conservator, Arquivo Municipal de Lisboa: Arquivo Fotográfico, will present "AML photography digitization, report on a 27-year project", reviewing the evolution of processes since 1992: image format and dimension, image quality evaluation, database upgrade, and the delivery of images to the public.
- Dina Noite, head of the preservation, conservation, and restoration department at the Arquivo Regional e Biblioteca Pública da Madeira, will discuss how their small photography collection of 10,000 images in 2016 grew to a collection of around 2 million, and how that profoundly changed their work flow and a rethinking of the digitization method.

10:30 – 11:10 Morning Coffee Break / Exhibit Open / Interactive (Poster) Papers Available for Viewing

---

* Page numbers indicate where the paper is found in the "bound" version of the conference proceedings on the accompanying USB stick.

PRESERVATION/ARCHIVING: STANDARDS AND GUIDELINES

Session Chair: Hugo Quisbert, ArkivIT (Sweden)
11:10 – 12:25

11:10 **Preserving the European Union's Digital Publications Heritage: Lessons Learnt on our Journey to Making the Past Accessible for the Future,** *Els Breedstraet, Publications Office of the European Union (Luxembourg)* **1**

The Publications Office of the EU is managing a digital preservation service on behalf of the EU institutions. Its vision and strategy on this service are described in a Digital Preservation Plan. This plan also details the processes, methodologies and tools applied, and documents the scope of the repository.

This short paper describes the establishment of a trustworthy digital preservation service and will showcase some challenges faced and issues remaining to be solved.

11:35 **Archiving AV Materials FAIR: An Oral History Collection in the Repository DANS-EASY,** *Eliane Fankhauser, DANS (the Netherlands)* **4**

It is crucial for both the research community and the public that AV materials are archived in a FAIR manner. This paper is about AV materials contained in DANS's EASY repository, and their depositing and archiving according to the FAIR principles. Using an Oral History collection in EASY as use case, it is explored to what extent the FAIR principles are followed and where their implementation is challenging. Moreover, tools currently developed to help assess the FAIRness of datasets before or after deposition are introduced.

12:00 **Metamorfoze Preservation Imaging Guidelines, Version 2.0,** *Hans van Dormolen, Hans van Dormolen Imaging & Preservation Imaging (HIP) (the Netherlands)* **9**

Since its publication in 2012, the Metamorfoze guidelines are used and adopted worldwide by the cultural heritage community and by camera & scanner manufactures. In 2017, ISO/TS 19264-1 was published. This ISO technical specification is based on the unification of the Metamorfoze and FADGI guidelines and on new technical insights. ISO/TS 19264-1 provides universally agreed upon terms, units and methodology for cultural heritage imaging.

Version 2.0 of the Metamorfoze Preservation Imaging Guidelines is under construction now. And a draft version is published.

TWO-MINUTE EXHIBITOR PROFILES

Session Chair: Christoph Voges, Hochschule für angewandte Wissenschaft und Kunst (HAWK), and consultant (Germany)
12:25 – 12:40

Archiving 2019 exhibitors FilmFabriek, i2S, Image Engineering GmbH & Co. KG, Image Science Associates, RFS, and Zeutschel GmbH share information about their products and services in these 2-minute previews.

GROUP LUNCH
12:40 – 14:00

OPENING DAY KEYNOTE

Session Chair: Lukas Rosenthaler, University of Basel (Switzerland)
14:00 – 14:50

14:00 **How the Market Changed—And the Lives of Photographs,** *Jonas Palm, director of preservation strategies, Riksarkivet/National Archives (Sweden)* . . . . *****

Through history we have been obsessed with creating images of what we see. It has been a way to convey information, ideas, or just something beautiful. The birth of photography changed that market completely. It paved the way for major changes in society, partly by showing information more vividly than just text and art images, and partly through the possibility of spreading this information widely within a relatively short time. Today images are made available within seconds around the globe.

The way photographers chose subjects and motifs depended upon the camera technology and the chemistry of the photographic medium. Even though analog photography rapidly spread around the world digital imaging, the latest incarnation of photography, compared with analog photography, truly brought photography to the masses. The development of photography has defined the ways of how to use the medium, a medium which has become a major part of our cultural heritage.

The lives of photographs have changed accordingly from portraits on mantelpieces to political powerful and explosive content. Some have rested for a long time as they were originally, and others were changed – the content manipulated. Preservation of analog and digital photography is as important as ever to secure the original content.

To fully understand a media like photography it is good to take a look at why, when, and how things evolved. When the Society for Imaging Science and Technology's Archiving Conferences started in 2004 in San Antonio, Texas, photography was in the middle of a major change from analog to digital. This is reflected on the program themes through the following years. New aspects on preservation emerged as well as on digital photography.

This presentation looks at the history of photography that had an impact. It is also intended as a reminder of what photographs are, not only physical objects to be preserved. It is easy to concentrate on preservation issues, which may be exciting, but we should remember that many times the information they carry is much more exciting.

DIGITIZATION: LARGE SCALE MASS DIGITIZATION AND WORKFLOW MANAGEMENT SYSTEMS

Session Chair: Martina Hoffmann, consultant (the Netherlands)
14:50 – 17:35

14:50 **Total Digital Access to the League of Nations Archives: Digitization, Digitalization, and Analog Concerns,** *Colin M. Wells, United Nations Office at Geneva (Switzerland)* **12**

The terms digitization and digitalization have often been used interchangeably, confounded, and misunderstood. Case studies on archives digitization, which can be understood to be the transformation of analog resources to digital formats, are plentiful, but there appears to be less discussion of digitization as but one aspect of the digitalization of archives, here meant to include the broader digital transformation of business models and services. The example of the Total Digital Access to the League of Nations Archives project can be used to examine how comprehensive digitalization, including digitization alongside traditional analog concerns, have shaped the definition of project workflows and operations within the more limited sphere of its digitization operations.

* No paper associated with talk; abstract only.

SOCIETY FOR IMAGING SCIENCE AND TECHNOLOGY

15:15 Managing the Digitisation Chain: Practical Tools for Process Management and Multi-faceted QA, *Nele Gabriëls, Dirk Kinnaes, Diederik Lanoye, Bruno Vandermeulen, and Mark Verbrugge, KU Leuven Libraries (Belgium)* **17**

The complexity of the digitisation process in a decentralised environment requires profound workflow and risk management, with attention for solid communication and for QA across the digitisation chain. In the absence of an all-in solution, ongoing projects at KU Leuven Libraries have served as a virtual lab environment to gradually create and test specific tools and standardise workflows since 2009. The resulting set of step-specific QA tools and processes form an integral part of a standardised yet modular digitisation workflow. The modular setup provides flexibility when developing project-specific workflows. Detailed tracking of the complete digitisation process for each individual project through a workflow management system allows for shared communication as well as for overall high-level risk management and multi-faceted QA in a time-efficient manner.

15:40 Interactive Paper 2-minute preview—Bringing the Humanities and Engineering Together through Multi-disciplinary Senior Design Team Projects, *Susan Farnand, Ian Kurtz, Reese Salen, Matthew Nygren, and Maher Abdelkawi, Rochester Institute of Technology (US)*

Please note: This author is not able to present during the Interactive Paper Session on Friday. She will stand by her poster to discuss its content during the Wednesday afternoon coffee break. See paper abstract on page viii.

15:42 – 16:15 Afternoon Coffee Break / Exhibit Open / Interactive (Poster) Papers Available for Viewing

16:15 Collaborative Capture: Leveraging Minimal Studio Space for Maximum Output, *Margaret C. McKee and Adam Neese, The Menil Collection (US)* **21**

In early 2018, the Menil Collection received financial support via a generous grant from the Houston Endowment to spearhead a new digitization initiative. Through tactical cross-departmental collaboration, the Menil has been able to strengthen institution-wide support for imaging. This initiative has allowed us to hire and train staff members, equip an industry-standard photography studio, and develop institution-wide workflows that support the long-term values in our commitment to taking extraordinary care for works of art and the central role of research and scholarship in the collection.

IS&T would like to thank the following Behind-the-Scenes Tour providers

Arquivo Municipal de Lisboa | Fotográfico
Conservation, Scanning, and Storage Departments

Arquivo Nacional da Torre do Tombo
Services of Conservation and Digitization

LUPA, Luis Pavão Lda
Conservation and Digitization

Calouste Gulbenkian Foundation
Art Library and Archives

Cinemateca Portuguesa–Museu do Cinema

Cineric Portugal

National Library of Portugal

Managed by Digital Asset Manager Margaret McKee and technically managed by Conservation Imaging Specialist Adam Neese, this project has resulted in over 1,500 objects being digitized to date with additional images of every object being captured for condition documentation and scholarly research. This paper outlines workflows established for collaborative capture at the Menil and discusses case studies in which the workflows have been utilized.

16:40 Digitizing and Preserving the Tuol Sleng Genocide Museum Archives: Stories of Compromises in a Challenging Environment, *Alexis Lecoq, Digital Divide Data (Cambodia); Jacqueline Vincent, Brechin Imaging / The Brechin Group Inc (Canada); Frederick Zarndt, Digital Divide Data (US); and Pheaktra Song, Tuol Sleng Genocide Museum (Cambodia)* . . **25**

Stories of mass digitization projects and the preservation of those physical materials to be digitized often bring to mind well-equipped set ups in clean dust-free, air controlled rooms with highly skilled staff. One particularly thinks of this for museums and archives located in temperate-climate developed countries. But what about stories of digitization projects with limited resources in less than optimum environments, how does one envision those?

The Archives of Tuol Sleng Genocide Museum (TSGM) is the largest and most complete record of Khmer Rouge actions during the Democratic Kampuchea (DK) regime. The Archives contain forced "confessions", many extracted under torture; biographies of prisoners, guards, and officials; photographs; original negatives; and other paper-based bound materials. Many items are very fragile and were not created with long-term preservation in mind so when handled for research they are at risk. The first Archive team did not record the provenance of the documents so research and recreating history is a challenge. In addition these records are not easily accessible for most Cambodians. For these reasons, and because the climate in Phnom Penh is hot and humid preservation is much harder than in more temperate climates, UNESCO and the Korea International Cooperation Agency (KOICA) have provided funds to digitize the Museum's archives. With these funds and under direction from UNESCO experts, along with the support of the Ministry of Culture and Fine Arts (MoCFA), Digital Divide Data and Brechin Imaging are collaborating to preserve and digitize the Archives and to train the Tuol Sleng staff in digitization and preservation.

Once ingested into the Project's database and crowd-sourcing website, which is currently being developed, these records will help illuminate the stories of people caught in this dark period of Cambodian history for the next generations.

17:05 Methodology and Tools for Quality Verification and Measurement Interpretation in a Digital Cinema Environment, *Miloslav Novák, University of Hradec Kralove, and Karel Fliegel, Czech Technical University in Prague (Czech Republic)* **31**

In this paper, we present a set of methodologies for Digital Cinema projector calibration, quality verification and evaluation by application of the international visual test SMPTE Digital Projection Verifier (DProve) as well as related SMPTE and ISO standards. We demonstrate modified SMPTE guidelines for faster, simpler and more precise measurements in practical conditions and introduce several tools for illustrative quality evaluation. The measurement interpretation is based on both, absolute tolerances defined by the SMPTE standards, as well as relative assessment calculated by the CIEDE2000 color difference formula. The latter approach is intended to better understand the importance of the color and tonality error for viewer's perception in the specific cinema low light levels. On the examples of selected D-Cinema projector and display measurements, we introduce possible strengths and weaknesses of various projection technologies, used in Czech and Slovak commercial cinema or review rooms, which differs by the specific manufacturer's technology or type of light source.

17:30 Closing Remarks; day ends at 17:35 / Evening on Own

THURSDAY 16 MAY 2019

WELCOME REMARKS AND THURSDAY KEYNOTE

Session Chair: Jeanine Nault, Smithsonian Institution (USA)
9:00 – 10:00

9:00 **Remarks**

9:10 **The JPEG2000 Suite of Standards: Capabilities and New Opportunities,** *David Taubman, deputy head school (reserach), School of Electrical Engineering and Telecommunications, UNSW Sydney, and director, Kakadu Software Pty. Ltd. (Australia)* . . . *****

JPEG 2000 is a family of standards that provide efficient and highly-scalable coding of imagery, capable of handling content with extreme resolutions and bit-depths; providing lossy and lossless compression in a unified framework; and producing codestreams that are highly-accessible based on region-of-interest and resolution-of-interest. Very recently, the family has been extended with a new High Throughput algorithm (JPEG 2000 Part-15) that dramatically lowers the computational burden of encoding and decoding JPEG 2000 content, while offering truly information lossless transcodability to/from all existing JPEG 2000 content.

While JPEG 2000 is widely adopted for the archival of still and moving imagery, the capabilities of JPEG 2000 are largely under-appreciated. The purpose of this presentation is to increase awareness of what JPEG 2000 can do, focusing on capabilities that could be of great benefit in archival applications. These capabilities include efficient interactive browsing of large media over low bandwidth networks, coding of imagery with rich color information (e.g., hyperspectral content), and coding of radiometric and raw camera data, as well as the embedding and incremental dissemination of metadata. The presentation also introduces the new Part-15 algorithm for high throughput and shows how it can benefit complex systems that are based on JPEG 2000.

PRESERVATION/ARCHIVING: ARCHIVAL MODELS AND WORKFLOWS

Session Chair: Steffen Hankiewicz, intranda GmbH (Germany)
10:00 – 12:20

10:00 **Preserving Irreplaceable National Digital Cultural Heritage in the Arctic World Archive,** *Morten Thorkildsen, Norwegian National Museum; Jahn Fredrik Sjøvik, formerly with Norwegian National Museum; and Bendik Bryde, Piql (Norway)* **39**

The Norwegian National Museum and Piql have collaborated on preserving some of the most valuable art history Norway has to offer. The project is the start of a global initiative aiming to protect some of the most valuable cultural heritage the world has to offer in a digital format. Safely secured on an Arctic island, the Arctic World Archive safeguards a copy of irreplaceable documents, drawings, sound pieces, pictures and movies, even databases or other types for format with significant heritage value.

10:25 – 11:05 Morning Coffee Break / Exhibit Open / Interactive (Poster) Papers Available for Viewing

11:05 **A Blueprint for Preserving Virtual World Cultural Heritage Using Preservica & Custom Metadata Schema,** *Marie Vans, HP Inc., and Patricia Franks, San José State University (US)* **42**

Faced with the possibility of running out of storage space and losing important born-digital cultural heritage artifacts, we are embarking on a project to design a process for preserving objects generated by our virtual world-based group so that these issues will not become critical in the future. Because of the high-turnover nature of student-based groups, preservation of objects associated with annual and learning events is essential for continuity in the community. We present our blueprint for preserving these artifacts using a combination of an established archiving system together with a customized metadata schema that reflects the specific needs of VCARA and an easy to use interface for browsing collections.

11:30 **Preservation of Evolving Complex Information Objects,** *Ivan Subotic and Lukas Rosenthaler, University of Basel (Switzerland)* **47**

Having trustworthiness as the driver, the long-term preservation of evolving complex information objects from a RDFbased Virtual Research Environment (VRE) has to ensure the integrity, authenticity, and provenance of the research data it encompasses. Besides the known difficulties, preservation of evolving complex information objects from a VRE provide additional challenges, as not only the objects created inside the VRE but also the VRE as such with its ontologies describing the structure of the digital objects, and additionally any referenced bitstream data, can evolve and change over time. This change over time needs to be captured in such a way, so that not only each object can be recreated to any version from its past, but also its context, namely all surrounding and connected digital objects, and correspondingly also their context and so on. Further, we propose to store all fixity information of the digital objects themselves and also of the provenance to a public blockchain, where it would serve as a single source of truth, which all users could trust.

11:55 **Tools for Presenting Multimedia Performance Documentation Using 3D Visualisations,** *Roman Berka[1], Bohuš Ziskal[2], and Zdenek Trávnícek[1]; [1]Czech Technical University in Prague and [2]CESNET (Czech Republic)* **51**

Within the project dedicated to the phenomenon of Laterna magika, tools and techniques for multimedia performances presentation and preservation were investigated, developed and applied. In order to present performances in their complexity and with respect to their historical context, a suitable data model had to be identified to build links between various documentation components with respect to existing standards and methods used by memory institutions. Being fragmented and incomplete, the existing documentation components were collected and digitized, the selected scenes were rebuilt in the form of 3D models. The actor's movement on stage was reconstructed and recorded using the motion capture system. For content management and presentation purposes, various tools were developed allowing for visualisations in VR environment and viewing/annotation of performance documentation components.

GROUP LUNCH
12:20 – 13:30

* No paper associated with talk; abstract only.

ACCESS: FORMATS FOR PRESERVATION AND ACCESS

Session Chair: Erik Landsberg, Cultural Heritage Digitization Consulting (US)

13:30 – 14:45

13:30 **Standardized Reflection Transformation Imaging (RTI) for Documentation and Research,** *Peter Fornaro, University of Basel, and Andrea Bianco, TRUVIS AG (Switzerland)* **57**

Reflection Transformation Imaging (RTI) is a powerful tool to capture the surface of an object for later examination. We brought RTI to the next level regarding reproducibility, simplicity and image quality. Besides the possibility to capture the surface topography, we can measure and reproduce materials composed of complex reflectance behaviours, with high colour and gloss fidelity. Full-featured cross-platform software for the creation and viewing of such enhanced RTIs has been developed and commercialized. The application allows simple interaction and comparison of RTI datasets.

However, the crucial feature of RTIs is their interactivity and the possibility to individualize viewing configurations for specific surface examinations. In the context of image interoperability, it is essential to be able to describe such viewing configurations to ensure reproducibility, e.g. in case of collaborative work.

The trend of linked open data shows that RTIs as well must follow open access rules. The success of standardized access to image data is demonstrated well by the international image interoperability framework (IIIF).

In this paper, we address two issues: The importance of standardized web-access to RTI data as well as an approach to store and disseminate specific RTI viewing configurations.

13:55 **ObjectVR Fashion: The Drexel Digital Museum Project,** *Kathi Martin, Nick Jushchyshyn, and Claire King, Drexel University (US)* **61**

"As cultural institutions continue to disseminate their holdings, now through digital technologies for reproduction, a challenge to humanities collections looking to digital curation and dissemination of their holdings is to make the quality and significance of the virtual representation meet or exceed that of the original object (Martin et all 2016 p.270)." The Drexel Digital Museum (DDM) has been experimenting with new media production for exhibition, preservation, and publication of historic fashion since 1999. This presentation traces the successes and challenges of our search to bring the highest quality virtual experience of dress to the largest and most inclusive audience through the evolution, and disappearance, of cutting edge technology.

14:20 **Toward a "Digital Noah's Archive" (DNA),** *Richard J. Solomon[1,2], Melitte Buchman[3], Eric Rosenthal[2,4], Jonathan M. Smith[1], and Clark Johnson[2]; [1]University of Pennsylvania, [2]Creative Technology, LLC, [3]Envision-Imagery LLC, and [4]New York University (US)* **66**

Visions for the future range from utopian to dystopian, with at least one instance suggesting that human progress will take forms none of us can imagine. Lacking any Delphic oracle, it seems prudent to prepare for the worst even while hoping for the best. In this paper we propose applying a novel, archival, silver-based medium to preserve the data of our global heritage for the long-term future. This media, Write Once Read Forever (WORF), is currently at the point of productization. WORF is distinguished by a stable, dense, data media requiring minimum maintenance and no continuous energy input for centuries-long and potentially millennia storage. As there are many risks to our civilization, our proposal is that multiple digital copies of a "Digital Noah's Archive" (or a DNA Ark)—encompassing our historical and current knowledge base—be placed in outer space away from Earth's fraught and currently unsustainable environment, as well as copies on our planet. Space, however, presents challenges of its own, so WORF media are currently being tested by NASA on the International Space Station to confirm its resistance to hostile space environments including ionizing radiation. These proposed, multiple, comprehensive Arks containing WORF data would expand on the sparse information about human existence on Earth as preserved in NASA's Pioneer and Voyager analog plaques launched to outer space in the 1970s.

Papers end at 14:45

BEHIND-THE-SCENES TOURS

15:30 – 18:00

Separate registration required. Visit the registration desk for details and to confirm your tour.

CONFERENCE RECEPTION AND DINNER

19:30 – 23:00

Casa do Alentejo, Rua Portas de Santo Antão 58, 1150-268 Lisboa

Cocktail reception followed by sit-down dinner.

FRIDAY 17 MAY 2019

CLOSING KEYNOTE AND AWARD PRESENTATIONS

Session Chair: Lukas Rosenthaler, University of Basel (Switzerland)

9:00 – 10:05

9:00 **Welcome and IS&T Awards**

9:15 **CERN's Digital Memory: When Patrimony Data Meets Scientific Data,** *Jean-Yves Le Meur, Digital Memory project leader, CERN (Switzerland)* . . . . *****

Since 1954, the European Organization for Nuclear Physics (CERN) has developed and exploited accelerator technologies to run fundamental research on high-energy particles. Punctuated by the discovery of new particles and the invention of the World Wide Web, the last 65 years have seen an explosion in the amount of data captured, whether scientific datasets or multimedia documentary. Both are now considered an essential component of the Institute's heritage, and they must be preserved for future generations.

Thus, as they are digitized, nearly 450,000 photographs, 4,000 hours of audiovisual material, and 6,000 hours of sound recordings from the past are now being added to the data acquired digitally. This massive digitization gives access to content full of surprises, and to new projects of enhancement through art, cinema, or virtual reality.

This exciting content is gaining a new audience at the same time that it is subject to the new risks of digital obsolescence. By managing it on the same infrastructure as the physics data, the multimedia heritage benefits from common developments. In particular the two content types could share a preservation platform, under construction, conforming to the OAIS model and based on open source digital repository software Invenio and a preservation software such as Archivematica.

* No paper associated with talk; abstract only.

INTERACTIVE (POSTER) PAPER PREVIEWS

Session Chair: Jeanine Nault, Smithsonian Institution (US)
10:05 – 10:35

Each author discusses their work for 2 minutes. The previews are followed by the Interactive (Poster) Paper Session/Coffee Break, where authors stand by their posters and discuss their work with attendees.

Every Engineering student at the Rochester Institute of Technology (RIT) must complete a senior capstone project as a condition of graduation. Most students fulfill this requirement in the form of a two-semester, multi-disciplinary senior design (MSD) team project. These projects are designed to provide the students opportunities to apply their classroom training in a collaborative environment. The multi-disciplinary teams bring together students with varied skill sets in projects that require them to assess customer needs, establish engineering requirements, benchmark existing solutions, evaluate possible concepts, and apply engineering practices to design, build, test, and document a working prototype device or process. See the MSD program website for further information.

The projects undertaken within the MSD program cover a broad range of applications, with sponsors including the National Science Foundation, Wegmans, Boeing, and ABVI/Goodwill. In recent years, projects undertaken in this program have included several efforts in conjunction with cultural heritage institutions. These include a streamlined manufacturing process for micro test targets for Image Science Associates; an update of the Image Permanence Institute's data logger, which is used to monitor the temperature and relative humidity in archives and libraries; streamlining of the George Eastman Museum's 2D artifact digitization process; and construction of period-appropriate replicas of an 18th century printing press and 16th century reading wheel for the Cary Graphic Arts Collection at RIT and the Rossell Hope Robbins Library at the University of Rochester. This paper features the Reading Wheel project to illustrate what the projects involve and how they can be used to bring the worlds of the humanities and engineering together.

Creating Artificial Ground-truth Data for Document Image Page

We propose a framework that can be used to create artificial ground-truth data for document images. The resulting data can then be used to train machine-learning systems to perform page segmentation tasks. The main focus of this system is on images of historical documents. The framework creates document images with headlines of differing sizes, multiple column layouts, pictures and decorative elements. To improve the resemblance with historical document images, a set of backgrounds is created manually by extracting background textures from real historical documents. The fading and curling typical of old manuscripts are also simulated.

Experiments with a neural network – trained on data generated using the proposed framework and applied to real-world images – show promising results with robust segmentation of text and non-text image areas.

Preprocessing Pipeline for Italian Cultural Heritage Multimedia

Preprocessing is an important task and a fundamental step in Information Retrieval, Text Mining, Natural Language Processing (NLP). While datasets in the English language can rely on well-established tools and methods for text preprocessing, the situation for the Italian language is more nuanced, due to a sum of factors, not least that fewer experiments and studies were made, and algorithms developed. Here we present an experimentation, a work in progress whose purpose is to define a pipeline able to preprocess texts. The different steps of the pipeline have been implemented and tested individually on Cultural Heritage datasets. The results obtained have been evaluated in the context of unsupervised automatic keyword extraction algorithms, such as RAKE or TextRank.

Content Analysis & Anonymization Made Simple,

Imagine the world of possibilities where all information would be easily findable and usable without the burden of copyrights, privacy issues or sensitive information. We at the Digitalia research center at the South-Eastern Finland University of Applied Sciences have elaborated on this issue further. Copyright issues we of course cannot affect due to legislation, but in our recent paper, we discussed the hidden information that could be unleashed by using a simple ocr with content analysis. Now we will go further by enhancing the content detection and including an anonymization in our service repertoire. We will also focus on making these services more easily usable by the interested community with the development of REST API.

Automatic Metadata (Entity) Extraction and Workflow Efficiency: Real

In a changing digitization landscape, automatic metadata extraction is becoming more important than ever before. At the same time, requirements regarding the extracted metadata are becoming more demanding as well: we are no longer interested in just extracting some data, we want to extract and identify entities.

Large leaps have been made on different aspects of metadata extraction, however integrated and effective workflows successfully and efficiently applying metadata extraction to real collections in a market environment are still rare.

This paper describes the research, the principles applied and the implementation of just such a workflow.

Bibliotheca Philadelphiensis: Collaborative Digitization and Data

In 2016, the University of Pennsylvania Libraries, along with fourteen partnering institutions in the Philadelphia area, was awarded a grant from the Digitizing Hidden Special Collections and Archives initiative of the Council on Library and Information Resources (CLIR) to produce the United States' largest regional online collection of medieval manuscripts. For the Bibliotheca Philadelphiensis project, otherwise known as BiblioPhilly, partnering institutions had thirty months to digitize more than 160,000 pages from 450 European medieval and early modern manuscripts. According to the terms of the grant, the digitized manuscripts had to be made available in the public domain via a searchable digital interface, be easily downloadable at high resolution, and accompanied by both expertly compiled descriptive metadata and unique physical collation models that help researchers to date manuscripts, understand how codices were disassembled and reconstructed in different periods, recombine fragments, and much more. While the BiblioPhilly project required intensive data capture from photographers and catalogers along with specific, time-sensitive, and particularly careful handling conditions, the process functioned smoothly through project management and a cooperative spirit among colleagues. As a result of these efforts, researchers may now creatively interact with the materiality of a manuscript in a digital environment in a manner that would be impossible with the physical manuscript itself. In a manner new to the field, BiblioPhilly enables researchers to become not just assessors of, but participants in, a long history of manuscript repurposing, reconstruction, and transformation.

SOCIETY FOR IMAGING SCIENCE AND TECHNOLOGY

How Accurate with Auto Settings Can be the Smartphone Camera for Cultural Heritage Color Reproduction?, *Alexandre Cruz Leão, Federal University of Minas Gerais (Brazil), and Stephen Westland, University of Leeds (UK)* **98**
Accurate Color Reproduction is very important for Cultural Heritage purposes, mainly for documentation. Normally the digital images are more saturated and contrasted than the human eye can see it and became necessary to recover the right color information. This research will present the color difference, in terms of CIELAB $\Delta E_{ab}*$, by four different smartphones with Auto Settings, with and without neutral color on the image provided by color targets (Colorchecker passport and Kodak Q-13 grayscale). To verify the color accuracy, some color samples were produced by a professional print and fineart paper. The colors printed were: Red, Green, Blue, Cyan, Magenta and Yellow. To perform the color analysis and color correction were used four methods: image without color target, with color target, linearization by Colorchecker (greyscale patches) and Color Profile. The result shows the accuracy in terms of color reproduction, the best, the worst, and the average for each smartphone for all six printed color, as well as the best methodology for each smartphone.

Involuntary Pseudos. The Untransposed Stereos in the Digital Archives,
Victor Flores, Universidade Lusófona de Humanidades e Tecnologias (Portugal) **103**
Digitizing stereo photographs can bring back to museums and archives some of the challenges that early photographers had to deal with. When digitizing the original negative plates in wet collodion or in gelatin silver, the protocol followed for stereos is apparently the same as for mono photography. As a result, archives are presenting to their public and researchers files with untransposed left and right images of each stereo pair. In other words, wrong or 'pseudo views' are being uploaded daily in institutional websites and databases, adding an extra layer of difficulty for the presentation and viewing of some historical gems of western visual culture. This problem was first noticed during a research project dedicated to Portuguese stereo photography, which has catalogued thirty one collections of stereo photography, consisting of mainly negatives.

Development of a Multi-Disciplinary Database of Cuneiform Tablets — An Improvement of 3D Models and Data Re-use,
Jaroslav Valach, Academy of Sciences of the Czech Republic, and Petra Štefcová and Ladislav Polák, National Museum (Czech Republic) . . **107**
This contribution describes a digitized cuneiform tablet collection based on a complex database architecture that integrates several scientific domains of data obtained from various types of analyses. The system enables creating 3D enhanced digital models, coupling data from science, and humanities to pose complex queries. Attention is also paid to the possibilities of re-use of the acquired data.

Colorimetric Scan, Display, and Print for Archiving based on the Ergonomic International Standard ISO 9241-306:2018 at Work Places, *Klaus Richter, Berlin University of Technology (Germany)* . . . **111**
ISO 9241-306:2018 shows colorimetric methods for output optimization of displays and projectors at work places. The optimization by equal spacing of colour series, visibility, and readability is intended. There are input linearization methods for scanners and photography and output linearization methods for displays, printers, and offset print. By a start output of a "digital" ISO-test chart with 729 colours (9x9x9 rgb* values) for example the loop "ISO-standard file -> ISO-print -> ISO-scan -> ISOfile" is closed and the rgb* colour data of the original ISO-file are approximately reproduced at the end of the loop. For any hue there is a linear relation in both directions between the rgb* and the CIELAB LCh* data. The closed loop and the linear relations are important properties for archiving.

The Challenges of Archiving Audio Visual Heritages in Northern Nigeria, *Musa Salih Muhammad, Ahmadu Bello University (Nigeria)* **113**
This paper attempts to showcase the challenges militating against the establishment, funding and promotion of audiovisual archiving collections held by different government institutions in Northern Nigeria. The rationale of such attempt rests on the observation that although audiovisual materials constitute vital information resource in such institutions, they are often not adequately managed after their creation. The paper also identifies the major constraints and limitations of audiovisual materials found in the public sectors, media houses, and the challenges in their management. This serves as an important mechanism for raising general awareness on the audiovisual materials and on the need to preserve and safeguard such materials for future generations, as a means of making them accessible to the public. The paper shows that many of the problems related to the management of such materials in the country emanate from, such as national policy on preservation and conservation, lack of trained personnel in the management of audiovisual materials, technological challenges, inadequate storage facilities, and the perception by users that audiovisual materials are not as valuable as records in other formats. The paper proffer strategies for dealing with the challenges of preserving and improving the use of audiovisual materials in Nigeria.

Functional Applications of Text Analytics Systems, *Steven J. Simske, Colorado State University, and Marie Vans, HP Inc. (US)* **116**
Text analytics can provide a wide breadth of valuable information, including summarization, clustering, classification, and categorization to enable better functional interaction with the text. This includes improved search, translation, optimization, and learning. In this paper, we describe advanced analytical approaches used to enable improved utility of the text documents and information later. This adds value to the preservation of the information and provides new access points to the information. We emphasize the role of functional approaches to testing and configuration of these systems, with the view that the primary role of archiving is to make the content as re-usable, re-purposeable, and discoverable as possible.

Digital Archive Use in Physical Education and Sports Culture,
Yung-Cheng Hsieh and Tzu-Han Chen, National Taiwan University of Arts (Taiwan) **120**
Sports and physical education have flourished in Taiwan, with many splendid achievements over the years. However, the lack of an appropriate preservation method has resulted in dispersal of the relevant artifacts into private possession. This study conducted a literature review and in-depth interviews aimed at digitalizing archives of these valuable artifacts from critical figures in sports and physical education. These digital archives are contained in multimedia formats in a digital database, which was successively developed both online and offline to demonstrate the diversity of sports culture. The digitalization and online communication of the memory and spirit of Taiwanese physical education culture, and the incorporation of these cultural assets in daily life to foster a cultural memory inheritance constituted the main objective of this study.

A Current Practically-oriented Manual for Digitizing Photographic Negative Collections and Producing Analogue Safety Copies,
Carolin Pommert (Germany) **125**
The paper aims to show the gap between theoretical conservation guidelines for handling and generating safety copies of photographic negative material and the practice of professional staff in large and small collections. By translating the discipline specific guidelines into a practically oriented manual, the importance of saving photographic negative heritage could be guaranteed in the future. Collection management should be able to understand guidelines, information and

practical research results, and transfer it into working processes in their collections for bringing them save into the future.

Challenges in the Cloud of Personal Archives, *Hugo Quisbert, ArkivIT (Sweden).*
In this paper we address some challenges in the development of personal archives. Personal archives can be seen as a collection of archival holdings connected to one specific individual during its lifetime. The collection can be the result of personal activities in private life or part of activities within an organisation. In order to develop easy-to-use and low-cost archival services for individuals, any company need to carry out a pre-scanning over the challenges that the company may confront during development. Some of the challenges we could identify in advance where connected to the following areas: Characteristics of personal archiving services, the user, underlying technical infrastructure, portability, security, pricing, and the market itself.

INTERACTIVE PAPER (POSTER) SESSION AND COFFEE BREAK
Meet the authors listed above and learn about their papers.
10:35 – 11:25

MANAGEMENT AND PARTNERSHIPS/ COLLABORATIONS: BEST PRACTICES, LESSONS LEARNED, AND CASE STUDIES
Session Chair: Anne Mason, National Archives and Records Administration (US)
11:25 – 12:40

11:25 **Prioritizing Black Digital Narratives from Process to Preservation,** *Leah Jones, Hannah Scruggs, Kamilah Stinnett, and Doretha K. Williams, National Museum of African American History and Culture (US)*
An outreach endeavor that centers African American experiences at the core of its mission, the Robert F. Smith Fund project is one of the major public programs for National Museum of African American History and Culture. Because many collections capturing African American history are not accessible to the general public, there is knowledge, culture, and lived experiences lost in boxes and unreadable recorded formats. The Smith Fund consists of multiple components entrenched in the idea the preserving and providing access to black narratives is only achievable by engaging black communities at multiple levels. The Community Curation Program works with individuals to provide digital preservation services to ensure that family history is safeguarded for generations to come. The Freedmen's Bureau Transcription Project opens the public to lives of the recently emancipated. Lastly, the Family History Center allows museum attendees to excavate their own family histories by offering access to databases such as Ancestry.com.

11:50 **Digital Archiving Technologies, Practices and Ethical Guidelines in Crowd-sourced and Community-based Efforts in Culturally Endangered Societies,** *Brian Pope, Scott Purdy, and Michael Conyers, The Arc/k Project (US)*
The inclusion of citizens as part of the cultural heritage community is a recipe full of promise (with a pinch of peril). In an era in which technology allows for lay people with very little training to collect immense amounts of in-the-field data pertinent to cultural heritage archiving, it behooves institutions to actively encourage and mature such activation by enthusiastically instructing and guiding this large and varied segment of the community through best practices. Although using data from individuals with little to no working knowledge of the intricacies of digital preservation is not feasible, it is equally unreasonable to dismiss such a massive and equipped population as completely unworthy. The

Arc/k Project's work in conjunction with citizens and communities across the globe demonstrates that a middle ground can achieve verifiable results with limited resources and in areas that are most in need of preservation.

12:15 **Crowdsourcing the Smithsonian: Developing and Maintaining the Smithsonian Transcription Center and the Digital "Volunpeer" Community,** *Caitlin Haynes[1,2], Janet B. Abrams[2], and Michael Schall[1]; [1]Quotient, Inc. and [2]Smithsonian Institution (US)*
The Smithsonian Institution has digitized millions of its archival, library, and museum collections over the past decade, but this work is only the first step in providing and improving collection accessibility and use. To further increase the discoverability and engagement of their digitized materials, the Smithsonian developed the Smithsonian Transcription Center—a pan-institutional crowdsourcing project that enlists the public (anywhere in the world) in transcribing and reviewing digitized field notes, diaries, letters, specimen catalog cards, and more. This paper and presentation will discuss the development, growth, and achievements of the Transcription Center, along with the ways in which the lessons learned by Smithsonian staff can inform others interested in creating their own crowdsourcing platforms.

GROUP LUNCH
12:40 – 13:45

ACCESS: INTEGRATION OF LINKED OPEN [USABLE] DATA (LOD/LOUD), OPEN SOURCE SOLUTIONS, AND APIS
Session Chair: Walther Hasselo, Heritage Leiden (the Netherlands)
13:45 – 15:25

13:45 **Digitizing, Archiving...and then? Ideas about the Usability of a Digital Archive,** *André Kilchenmann, Flavie Laurens, and Lukas Rosenthaler, Data and Service Center for the Humanities DaSCH (Switzerland)*
Digitizing is everywhere. Archives and libraries bring their holdings into the digital world. New sources are created exclusively digitally. The archive is and will become digital. The process of digitization has been going on for more than twenty years and has increased considerably since the appearance of the first smartphone. It is now possible to be online, read, and generate digital content around the clock. And the amount of data is enormous.

At the Data and Service Center for Humanities, we are confronted with this amount of data in the research field, and we try to bring the different data models and different media types into a uniform but flexible system. The difficulty is not the data storage anymore but the presentation and usage of the data. The aim is not to archive data, but to keep them alive. Availability and usability are playing an important role. Right now, we are two front-end developers building our infrastructure web applications and think about new possibilities to bring the data to the users.

14:10 **Linked Open and Annotated Science and Heritage Data,** *Fenella G. France and Andrew Forsberg, Library of Congress (US)*
The challenges involved in assuring the longevity and validity of heritage science data, and access to that data, require that all scientific data terminology and experimental procedures are not proprietary and have a common meaning across heritage institutions. While the temptation to create a new set of thesauri and definitions is great, it merely exacerbates the "siloed" impact that tends to separate rather than

aggregate colleagues and data. Using IIIF and the Mirador viewer to integrate scientific and scholarly data about heritage objects, it became apparent when attempting to create a cohesive structure that terms in common use amongst one group of users were not necessarily familiar to the others. Therefore, easily accessed but rigorous controls on terms needed to be put into effect, with preference deliberately given to reusing existing resources.

14:35 **Cluster-based Unsupervised Automatic Keyphrases Extraction Algorithms: Experimentations on Cultural Heritage Datasets,** *Maria Teresa Artese and Isabella Gagliardi, CNR (Italy)* . . . **156**
Automatic keyword extraction is the process of identifying key terms and key phrases from documents that can appropriately represent the subject of the documents. We present here a work-in-progress, an experimentation done on unsupervised keyword extraction, with the aim of automatically associating scored keyphrases to texts, using (standard or innovative) cluster based methods, and integrating word embedding to enhance semantic relatedness of keyphrases.

In the paper we present the datasets used, the state-of-the-art for unsupervised automatic extraction algorithms, based on cluster methods, and we describe in details the algorithms implemented and preliminary results obtained. The results obtained are discussed, commented, and compared with those obtained, in previous experimentations, using TextRank, RAKE, and Tf-idf.

15:00 **Technical Challenges and Approaches to Build an Open Ecosystem of Heterogeneous Heritage Collections,** *Ricard de la Vega, Natalia Torres, and Albert Martínez, Consorci de Serveis Universitaris de Catalunya (CSUC) (Spain)* **161**
Empowering Communities with a Heritage Open Ecosystem (ECHOES) is a project that intends to provide an open-source and modular architecture to gather different digital contents related to European heritage.

When a wide amount of heterogeneous data of collections, disciplines, and countries are joined, different kinds of technical challenges must be considered. In this article are detailed these challenges and the approaches that have been used, with more or less success, to solve it.

15:25 – 15:50 Afternoon Coffee Break

DIGITIZATION: NEW DEVELOPMENTS IN TECHNOLOGIES AND WORKFLOWS
Session Chair: Miloslav Novák, University of Hradec Kralove (Czech Republic)
15:50 – 17:05

15:50 **High Resolution Film Scanning Reconstruction by Image Stitching and Intensity Correction,** *Lei He, Library of Congress (US)* **167**
Testing at the Library of Congress and other institutions has determined that there can be greater than 3000ppi of real information in original film negatives and transparencies. This exceeds the real capture resolution of virtually all digitization systems which capture the entire film in a single image. The limitations of optics, sensors, and light restrict the effective resolution achievable in a single capture to substantially less than the information available in the original materials.

The aim of this project was to develop a system to combine multiple image segments acquired at very high resolutions to create a single merged image that effectively contains the total information available in the original film. There are several commercially available solutions which can be used to perform this task, however, none of those tested produced optimal results. The project was specifically focused at multi-

* No paper associated with talk; abstract only.

segment monochrome capture of large format photographic negatives from the LOC Prints and Photographs (P&P) Division, however the process developed is applicable to a wide range of applications.

16:15 **The Digitalization of Analogue Stereo Photographs and the Creation of the Digital Stereo Archive,** *Rodrigo Peixoto, Filipe Costa Luz, and Jorge Oliveira, Universidade Lusófona de Humanidades e Tecnologias (Portugal)* **171**
This paper analyzes the problems arising from the remediation of the relief effect in the transition from analog to digital of stereo photography. One of the main problems in this conversion is the portability of the awe effect that constitutes an important part of the experience when viewing a stereo pair. This image conversion process, necessary to the creation and dissemination of digital files of 19th century stereoscopic photography, is not linear. The digital stereoscopic projection cards present a number of difficulties for a proper consistency reproduction of the relief effect. Through the study comparison of different viewing apparatus (both digital and analogue including 3D and VR) of a specific stereo image, we will present important results achieved with a sample of 134 participants that were exposed to these devices and propose a guideline manual for the digital stereo archive.

For the developing of this work it has been crucial the research done by both authors for the Stereo Visual Culture project (supported by the FCT Foundation ref. PTDC / IVCCOM /5223/2012), the stereopsis analysis made with the research center HEI-Lab (Digital Human-Environment Interaction Lab) and through the particular case study based on the VR application developed for the recreation of XIX century Carlos Relvas' studio, in exhibition at Museu Nacional de Arte Contemporânea do Chiado (Lisbon) from November 2018 until February 2019.

16:40 **ISO Standards for Monitoring Image Quality,** *Dietmar Wueller, Image Engineering GmbH & Co. KG (Germany)* . . . . *****
The presentation will provide an update on the current ISO standard procedures for monitoring the image quality of digitization systems for cultural heritage. The talk will cover three ISO standards: ISO/TS 19264-1:2017 Photography – Archiving systems – Image quality analysis – Part 1: Reflective originals, ISO/TR 19263-1:2017 Photography -- Archiving systems – Part 1: Best practices for digital image capture of cultural heritage material, and ISO 19262:2015 Photography – Archiving Systems – Vocabulary.

OBSOLETE MEDIA AWARD FOR BEST INTERACTIVE PAPER PRESENTATION AND CLOSING REMARKS
17:05 – 17:15

IS&T would like to extend a special thank you to
Arquivo Nacional da Torre do Tombo,
Universidade Lusófona, and
their staff for their support of Archiving 2019.

AUTHOR INDEX

* No paper associated with talk; abstract only.

IS&T BOARD OF DIRECTORS: JULY 2018 – JUNE 2019

SOCIETY FOR IMAGING SCIENCE AND TECHNOLOGY

IS&T CORPORATE MEMBERS

SUSTAINING CORPORATE MEMBERS

Adobe

hp

Qualcomm

xerox

SUPPORTING CORPORATE MEMBERS

FUJIFILM

KONICA MINOLTA

Lexmark

DONOR CORPORATE MEMBERS

CLARIANT

DxOMARK
IMAGE LABS

JPO
JAPAN PATENT OFFICE

RICOH

Felix Schoeller Group
Best Performing Papers. Worldwide.

TRek Has Joined

Advanced Energy

Preserving the European Union's digital publications heritage: lessons learnt on our journey to making the past accessible for the future

Els Breedstraet; Publications Office of the European Union; Luxembourg, Luxembourg.

Abstract

The Publications Office of the EU is managing a digital preservation service on behalf of the EU institutions. Its vision and strategy on this service are described in a Digital Preservation Plan. This plan also details the processes, methodologies and tools applied, and documents the scope of the repository.

This short paper describes the establishment of a trustworthy digital preservation service and will showcase some challenges faced and issues remaining to be solved.

Short paper

Introduction

The Publications Office of the European Union (OP), based in Luxembourg, is an interinstitutional office whose task it is to publish the publications of the institutions of the European Union. Its core activities are producing and disseminating legal and general publications in a variety of paper and electronic formats. It manages a range of websites to provide citizens, governments and businesses with digital access to official information and data from the EU. These sites include EUR-Lex, the EU Open Data Portal, EU Publications, TED (Tenders Electronic Daily) and CORDIS. OP is also mandated to centrally preserve the EU digital publications heritage [1].

The Publications Office's digital preservation service

OP started its digital preservation activities already at the end of the 1980s. The first three versions of OP's digital archive, ARCDON [2] (1987-1995), EUDOR [3] V1 (1996-2001) and EUDOR V2 (2002-2016), were file-based document management systems. The first version, ARCDON, contained digitized copies of the Official Journals. In the next version, EUDOR V1, digitized copies of the other publications identified by OP were added. The content was further completed in the third version of the repository, EUDOR V2. Each version added something significant to its predecessor. For instance, EUDOR V2 was key to the creation of a digital archive of all EU publications dating from 1952. However common to all these versions was that, files were stored without any associated metadata except for their filenames. These filenames followed strict naming conventions reflecting the content of each document.

With the entry into force of Council Regulation 216/2013 on the electronic publication of the Official Journal of the European Union [4] came a clear obligation to be able to preserve the electronic edition of the Official Journal that is considered as authentic and produces legal effects.

In response to this and in light of the growing importance of digital long-term preservation in general, OP decided that the next version of the archival repository, EUDOR V3 (from 2017 onwards) had to become a proper trustworthy long-term digital archival repository. This new version needed to fulfill requirements such as enhanced security, rich metadata, a means to check fixity, improved workflows and audit trails. Although remaining the data owner, OP opted to outsource the operational aspects relating to the repository as a managed service. In practice this means that an external service provider takes care of the implementation and provision of the service: developing and maintaining (IT) systems and infrastructure, taking care of and following up ingestions of digital objects in the repository, drafting and updating documentation, etc. OP however continues to provide (technical) specifications, follows up on daily operations and administratively manages the contract.

To guarantee authenticity and trustworthiness, the choice was made to base the long-term digital archival repository on following standards: ISO 14721:2012 (Open Archival Information System - OAIS) [5] as reference model for the implementation of the repository and ISO 16363:2012 [6] to verify its trustworthiness.

Digital Preservation Plan: introduction

A key instrument for OP's long-term digital preservation service is the Digital Preservation Plan. It defines and documents the vision and strategy of the long-term digital preservation service that OP manages on behalf of the EU institutions. This policy document is an official commitment made by OP to provide the service but it is at the same time a prerequisite for the audit and certification of the trustworthiness of the long-term digital archival repository. The Digital Preservation Plan details the processes, methodologies and tools applied, and documents the scope of the repository in terms of the content covered. It also covers governance, funding, service level and quality expectations. The text is supported by a series of annexes, containing for instance details on the ingestion workflow (based on OAIS [5]), risk management, fixity, representation and provenance information, formats, and details of the collections preserved. The plan will serve as a guideline to cope with technological and organisational issues and will help to build a "preservation culture" inside the organisation.

Digital Preservation Plan: drafting journey

Once the decision to establish a trustworthy long-term digital preservation service was taken, OP's long-term preservation team started working on a key element of that service, a detailed Digital Preservation Plan.

The first step on the drafting journey was to draw up a skeleton for the plan. This was done with the help of an

external consultant. The chapters defined in the skeleton were divided among the core members of the team. This small drafting team consisted of three colleagues: an archivist, an IT expert and a collection expert. Each of them worked on the chapters closest to their area of expertise. This work led to a preliminary draft.

In a second phase, the preliminary draft was discussed during bi-lateral meetings with experts on long-term preservation and/or archiving, representing each EU institution (i.e. OP's stakeholders). Stakeholders approved the vision and strategy outlined in the summer of 2017. The draft was refined, enriched and enhanced with the feedback collected during the meetings. This eventually resulted in a first version of the plan.

As a next step, the document was sent to OP's Management Committee, the highest governance instance that represents OP's stakeholders, who formally adopted the first version at the end of 2018.

The plan is currently being gradually phased in. An important step in its implementation has been putting into practice the governance structure. A milestone in this process was the establishment of the Steering Committee, which had its first meeting in March 2019. This committee consists of experts in the field of long-term preservation. They are nominated by their institution and can take decisions on its behalf. The Committee will supervise the activities of OP in the field of long-term preservation and will, in particular, be responsible for overseeing the implementation and governance of the Digital Preservation Plan. It will also act as a knowledge-sharing platform.

Digital Preservation Plan: issues, and how to turn them into opportunities

During the drafting of the Digital Preservation Plan, the team has encountered several stumbling blocks. By seeking to resolve these issues, valuable lessons have been learnt and in most cases the team has been able to turn the initial obstacles into opportunities. Some examples are set out below.

One of the most important issues is that policies are being defined while operations run. OP has its mandate to fulfil, so the show must go on. The migration from the previous archival repository (EUDOR V2) to EUDOR V3 started more or less at the same time as the drafting of the Digital Preservation Plan. This meant that the implementation of the system started, without yet having the supporting policies in place. Consequently, the operational archival practice and the policy outlined in the Digital Preservation Plan need to be aligned. In order to achieve this, the long-term preservation team has to analyse and review the practices in place. It is not easy to remedy issues in an archival system in which thousands of digital objects are ingested on a daily basis. However, being obliged to do this thorough analysis allows the team to gain a deep insight in what is going on in the "black box" of the long-term digital archival repository and to take informed remedial decisions. The lessons thus learnt will be extremely valuable for the future.

A second issue to take into account is the fact that OP's long-term digital archival repository does not stand alone. As explained by L. Bountouri et al. [7] "before its archiving in the long-term digital archival repository, a digital object passes through two other information systems. The first system deals with the reception of metadata and digital content. During the reception workflow, OP receives digital objects and metadata, while at the same time takes action on both of them, such as generating new formats of a digital object and transforming the metadata to richer structures." The second system ensures their storage and dissemination. In this system, data curation actions are taking place on the digital content and its respective metadata,. This situation leads to constraints, but also offers advantages. Content and metadata arriving in the long-term digital archival repository have gone through prior validations and are therefore usually already well structured.

Another stumbling block encountered was the lack of descriptive metadata in the archive. In the old repository only very minimalistic descriptive metadata accompanied the digital files. However, a trustworthy long-term digital archive requires many well-defined and well-structured metadata. Choices and policies regarding metadata have to be reflected in the Digital Preservation Plan. This means that metadata schemas have to be defined and documented and subsequently, metadata have to be generated and added to the files. Fortunately, some metadata could be produced and added in an automatic way. On the upside, the initial lack of metadata allows OP to carefully think about metadata schemas, especially while defining newly added metadata.

As a final example, the lack of time and experience can be mentioned. As outlined above, the drafting core team contained only three colleagues. They had to take on the writing in addition to their other daily responsibilities. This inevitably had an impact on the speed of advancement. As drafting progressed, it became clear that the different backgrounds of the core drafting team were a recipe for success. The different perspectives led to insights that would not have been achieved if they all had the same profile. The impact of the lack of direct experience was also mitigated by the availability of input from external consultants who supported the process and by feedback from experts of other institutions.

Digital Preservation Plan: challenges ahead

Maintaining a digital preservation service is a continuing journey. Several issues still remain to be resolved. These include:

1) Provenance metadata: In a trustworthy archive, it has to be possible to track back every event that may have produced a change of any type in the life cycle of a digital object. This has to be encoded in provenance metadata. Moreover, as explained in the OAIS standard [5], provenance metadata can support the authenticity of a digital object. Today, the provenance metadata provided by the producer are only to a limited extent present in the long-term digital archival repository. It is planned to model them further in the near future, using the PREMIS Data Dictionary [8] and to add them to the existing metadata. This is particularly important in OP's context, because (as explained above) a digital object passes through other information systems, before arriving in the long-term digital archival repository.

2) File formats: There are many outstanding questions regarding file formats. A major one is how we can in a reliable way identify and validate file formats received by the archive. Other important questions are: Should the archive accept all file formats; or can producers be required to submit specific ones?; What level of preservation [9] will be guaranteed for which format?

3) Establishing a robust skills and technologies portfolio: Despite high workload and limited human resources, the OP's long-term preservation team will have to find a way to acquire and maintain the necessary skills and knowledge to be capable to run a long-term digital preservation service that respects the highest standards. Technology watch has to be done on a permanent basis to stay up to date in the quickly changing field. As the team is small, a balance will have to be sought between acquiring expertise within the team on one hand and seeking help, advice and assistance outside the team, e.g. from colleagues working in related services (IT, data modelling, etc.) and/or external service providers and organisations (e.g. the Digital Preservation Coalition).

4) ISO 16363-certification [6] as a trustworthy repository: Obtaining this certificate is an ambitious objective. Many requirements need to be fulfilled and the procedure is demanding. Therefore, OP will aim for a step-by-step approach: first launching the certification of the part of repository containing the electronic edition of the Official Journal, then moving to other parts at a later stage. The reasoning behind this approach is that Regulation 216/2013 [2] stipulates that the digital version of the Official Journal produces legal value and that OP "…shall be responsible for: (a) publishing it and guaranteeing its authenticity; …(e) preserving and archiving the electronic files and handling them in line with future technological developments.". Consequently, it is crucial to preserve the Official Journal correctly in the long-term and be able to ascertain its authenticity.

Conclusion and next steps

The Digital Preservation Plan will serve as a guide further down the road to a trustworthy long-term digital archive for the EU institutions. It will need to be continuously developed and kept up to date in line with new insights and changing realities (technological, organizational, etc.).

Next steps will be to tackle the challenges listed above and in particular to complete the supporting annexes.

It is hoped that the lessons learnt on this journey will serve as inspiration and guidance for others already, or about to embark, on a similar adventure.

References

[1] European Union. "Decision of the European Parliament, the Council, the Commission, the Court of Justice, the Court of Auditors, the European Economic and Social Committee and the Committee of the Regions of 26 June 2009 on the organisation and operation of the Publications Office of the European Union.", OJ L 168, 30.6.2009, p. 41-47. Available at: http://data.europa.eu/eli/dec/2009/496/oj [Accessed 06.10.2018]
[2] ARCDON stands for ARChives sur Disque Optique Numérique
[3] EUDOR stands for EUropean DOcument Repository.
[4] European Union. "Council Regulation (EU) No. 216/2013 of 7 March 2013 on the electronic publication of the Official Journal of the European Union. Official Journal of the European Union.", OJ L 698, 13.3.2013, p. 7-9. Available at: http://data.europa.eu/eli/reg/2013/216/oj [Accessed 19.03.2019]. Amended by: European Union. "Council Regulation (EU) 2018/2056 of 6 December 2018 amending Regulation (EU) No 216/2013 on the electronic publication of the Official Journal of the European Union", OJ L 329, 27.12.2018, p. 1–2. Available at: http://data.europa.eu/eli/reg/2018/2056/oj [Accessed 19.03.2019].
[5] International Organization for Standardization: ISO 14721, 2012: space data and information transfer systems, open archival information system (OAIS), reference model. International Organization for Standardization (2012). Available at: https://www.iso.org/standard/57284.html. [Accessed 19.10.2018]
[6] International Organization for Standardization: ISO 16363, 2012: space data and information transfer systems, audit and certification of trustworthy digital repositories (2012). Available at: https://www.iso.org/standard/56510.html. [Accessed 19.10.2018]
[7] Bountouri L., Gratz P., Sanmartin F. Digital Preservation: How to Be Trustworthy. In: M. Ioannides (Ed.): ITN-DCH 2017, LNCS 10605, pp. 364–374, 2018.
[8] PREMIS Editorial Committee: PREMIS data dictionary for preservation metadata, November 2015. Available at: http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf. [Accessed 08.03.2019]
[9] See for example Megan Phillips, Jefferson Bailey, Andrea Goethals, and Trevor Owens, "The NDSA Levels of Digital Preservation: An Explanation and Uses," In: Proceedings of the Archiving (IS&T) Conference (2013). Available at: https://ndsa.org/documents/NDSA_Levels_Archiving_2013.pdf [Accessed 08.03.2019]

Author Biography

At, the Publications Office of the EU, Els Breedstraet is heading the archiving and long-term preservation service. She holds a degree in library and documentation science and has large experience as information manager.
https://www.linkedin.com/in/elsbreedstraet/