IS&T Archiving Conference (ARCHIVING 2020)

Digitization, Preservation, and Access

Online 7 – 21 May 2020

ISBN: 978-1-7138-1332-3

Printed from e-media with permission by:

Curran Associates, Inc. 57 Morehouse Lane Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2020) by Society for Imaging Science & Technology All rights reserved.

Printed with permission by Curran Associates, Inc. (2020)

For permission requests, please contact Society for Imaging Science & Technology at the address below.

Society for Imaging Science & Technology 7003 Kilworth Lane Springfield, Virginia 22151 USA

Phone: 703-642-9090 Fax: 703-642-9094

info@imaging.org

Additional copies of this publication are available from:

Curran Associates, Inc. 57 Morehouse Lane Red Hook, NY 12571 USA Phone: 845-758-0400

Fax: 845-758-2633

Email: curran@proceedings.com Web: www.proceedings.com

TECHNICAL PAPERS PROGRAM CONFERENCE SCHEDULE AND TABLE OF CONTENTS

MONDAY 18 MAY 2020

SESSION I

Session Chair: Jeanine Nault, Smithsonian Institution (US)

10:00 - 11:00 EDT

10:00 WELCOME REMARKS

10:05 PRESERVATION KEYNOTE

The Ever-changing Work that is Digital Preservation,

Since the 1960s, digital preservation has transformed from a secondary activity at a select few cultural heritage organizations to a vital international effort with its own best practices, standards, and community. This keynote presentation and paper presents an overview of the changing scope of digital preservation, issues, and strategies for digital preservation in the cultural heritage community.

11:00 - 11:25 EDT e-Coffee Break / Exhibit Open

Join other attendees in one of the technical topic discussion rooms, the coffee break area, or the exhibit hall.

SESSION II

Session Chair: Steffen Hankiewicz, intranda GmbH (Germany)

11:30 - 12:30 EDT

11:30 PRESERVATION

Digital Preservation has evolved from an early-stage field based heavily on research and the sharing of information to a nascent industry based on practical activity. In this transition there is a risk that the vital activity of sharing information and expertise declines in favor of the day-to-day practicalities of caring for content. This work explores how the Preservation Action Registries (PAR) Initiative can not only help to bridge the gap, but in doing so, create new opportunities that can help make automated digital preservation a practical reality even for non-expert users by describing a proof-of-principle demonstration of the automated application of Digital Preservation Policy, and subsequent changes to that policy.

11:50 DIGITIZATION

Digital imaging, as an archival practice, is not a "solved problem" for the cultural heritage community. As Google, publishers, and other content providers digitize and deliver resources at scale, there is an increasingly pressing demand from users to digitize the rich resources in library special collections, archival institutions, and the vast array of invaluable content in private collections. This paper introduces a research and learning initiative (Dig4E-Digitization for Everybody) designed to bridge the knowledge gap that presently exists between well-established or

MONDAY KEYNOTE SPONSOR



emergent international standards derived from imaging science, on the one hand, and local practices for digital reformatting of archival resources. The paper describes the rationale for the education and training initiative and summarizes the intellectual structure and the technical platform of an innovative sequence of self-paced online resources that can be adapted for a variety of audiences.

12:10 DIGITIZATION

Access to Collections: Challenges of Physical and Digital— Assessing Digitization Decisions, Fenella France and Andrew

12:30 - 12:55 EDT e-Coffee Break / Exhibit Open

Join other attendees in one of the technical topic discussion rooms, the coffee break area, or the exhibit hall.

SESSION III

Session Chair: David Walls, US Gorvernment Publishing Office (US)

13:00 - 13:50 EDT

13:00 EXHIBITOR PREVIEWS

Archiving 2020 exhibitors Image Engineering, FilmFabriek, Colorburst, and Arkhênum share information about their products and services in these 2-minute previews.

13:10 PRESERVATION

Expanding the Scope of Digital Collection Development for Heritage Preservation: The Case of the Odin Oyen Collection,

David Mindel, University of Wisconsin-La Crosse (US) 22

With the potential of digitization as an alternative form of donation, institutions must more often weigh the importance of physically owning an

object versus only digitally capturing that object. This postcustodial approach may run counter to some library donation prerequisites and traditional collection development policies, instead focusing on the cultural and intellectual benefits that such a compromise may bring. This approach is not without potential obstacles, including issues of copyright, ownership, and reproductions. However, this paper aims to reinforce the benefits of the post-custodial model through the evolution of the Odin Oyen digital collection—a collection comprised of physical materials owned by public and private entities reunited through digitization for the purpose of preserving local and cultural heritage.

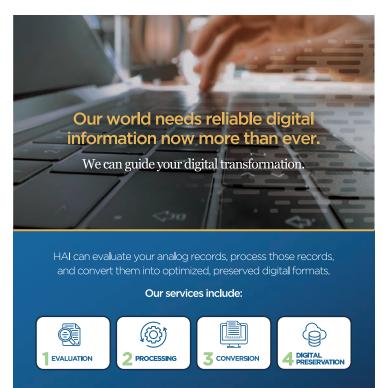
13:30 ADVANCED IMAGING

To Predict the Lightfastness of Prints on Foil Applying Artificial Neural Network, Mahasweta Mandal and Swati Bandyopadhyay, Jadavpur University (India)

The lightfastness of prints is an extremely important characteristic for assessing their print stability. The fastness properties of prints can be described in terms of print durability and image stability. Good lightfastness assures the good print stability after long use. This study has focused to describe the lightfastness of printed foil samples due to long time exposure. It may also be used for the authenticity or validity of the product. Moreover, any kind of deterioration in package print quality will affect the sale value of the product adversely. Little work has been done to study the fastness properties of printed films and foils. In this work, blister foils printed in the gravure printing process are taken as the sample as it has extensive usage in food and medicine packaging. The samples are exposed in artificial lightfastness tester BGD 865/A Bench Xenon Test Chamber (B-SUN) for assessing the light fastness of cyan, magenta, yellow and black ink on the foil. The spectral curves and colorimetric values of prints are measured in the ocean optics spectroradiometer (DH2000BAL) before and after exposure. An Artificial Neural Network model (ANN) is proposed to predict the fading rate of the printed foil. The optimal model gives excellent prediction with the minimum mean square error (MSE) for each color and a correlation coefficient of 0.98-0.99.

13:50 DAY ENDS

Feel free to continue the discussions. Join other attendees in one of the technical topic discussion rooms, the coffee break area, or the exhibit hall



Contact HAI today for a no-cost, initial consultation on the digital transformation of your historical and operational records.

Laura Starr

Business Development Manager Istarr@historyassociates.com 240.514.0926



www.historyassociates.com

TUESDAY 19 MAY 2020

SESSION IV

Session Chair: Walther Hasselo, Heritage Leiden (the Netherlands)

10:00 - 11:00 EDT

10:00 ADVANCED IMAGING KEYNOTE

Spectral Archives: Obstacles and Opportunities, Roy Berns, Rochester Institute of Technology (US); abstract only

Optical radiation can be readily separated into individual wavelengths. A material's spectral properties, whether emitted, transmitted, or reflected, is fundamental, defining the material unambiguously. Artwork reproductions having the identical spectral properties to the original art will match the original in color under all illuminating conditions and for all observers. The reproduction's appearance mimics the original, useful for lighting decisions. The spectral data can be used for authentication, conservation treatments, and as a component of technical examination. Given the availability of spectral imaging systems and the fundamental nature and utility of spectral data, why are image archives of cultural heritage overwhelmingly RGB? Why are studio cameras only RGB? There must be obstacles preventing spectral imaging from entering the studio. What are they? Multi-spectral and hyper-spectral imaging seem exclusive to academics and conservation scientists having imaging expertise. Are there reasons why studio photographers are excluded? These and similar questions are the subject of this presentation, along with a review of the principles and applications of spectral imaging.

11:00 - 11:25 EDT e-Coffee Break / Exhibit Open

Join other attendees in one of the technical topic discussion rooms, the coffee break area, or the exhibit hall.

SESSION V

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

11:30 - 12:30 EDT

11:30 EXHIBITOR PREVIEWS

Archiving 2020 exhibitors Creekside, HAI, nextScan, and WG5 share information about their products and services in these 2-minute previews.

11:40 ADVANCED IMAGING

Integrating Advanced Imaging of Ancient Manuscripts,

Archives, libraries, and commercial firms are utilizing new advanced imaging methods for research into cultural heritage objects. New technical systems, including the latest multispectral (MSI) and x-ray fluorescence (XRF) imaging systems and higher resolution cameras raise major challenges for not only the integration of new technologies, but also the ability to store, manage, and access large amounts of data in archives and libraries. Recent advanced imaging of ancient Syriac palimpsests (parchment manuscripts with hidden texts embedded within them) demonstrated an approach that utilized multiple imaging techniques and integration and analysis of data from multiple sources. Three palimpsest imaging projects (Archimedes Palimpsest, Syriac Galen Palimpsest, HMML Palimpsest) supported research with a range of

advanced imaging techniques with MSI and XRF, requiring implementation and standardization of new digitization and data management practices for the integration, preservation, and sharing of advanced image data.

12:00 DIGITIZATION

FOCAL TALK Refining the Theory-to-Practice Path for FADGI Still

The still imaging portion of FADGI continues to be a living document that has evolved from its theoretical digital imaging principles of a decade ago into adaptations for the realities of day-to-day cultural heritage workflows. While the initial document was a bit disjointed, the 2016 version is a solid major improvement and has proven very useful in gauging digital imaging goodness. With coaching, encouragement, and focused attention to detail many users, even the unschooled, have achieved 3-star compliance, sometimes with high-speed sheet-fed document scanners. 4-star levels are not far behind.

This is a testimony to an improved digital image literacy for the cultural heritage sector that the authors articulated at the beginning of the last decade. This objective and science based literacy has certainly evolved and continues to do so. It is fair to say that no other imaging sector has such comprehensive objective imaging guidelines as those of FADGI, especially in the context of high volume imaging workflows. While initial efforts focused on single instance device benchmarking, future work will concentrate on performance consistency over the long term. Image digitization for cultural heritage will take on a decidedly industrial tone.

With practice, we continue to learn and refine the practical application of FADGI guidelines in the preservation of meaningful information. Like rocks in a farm field, every year new issues and errors with current practices surface that were previously hidden from view. Some are incidental, others need short term resolution. The goal of this paper is to highlight these and make proposals for easier, less costly, and less frustrating ways to improve imaging goodness through the FADGI guidelines.

12:30 - 12:55 EDT e-Coffee Break / Exhibit Open

Join other attendees in one of the technical topic discussion rooms, the coffee break area, or the exhibit hall.

SESSION VI

Session Chair: Peter Burns, Burns Digital Imaging (US)

13:00 - 14:00 EDT

Three 10-minute papers followed by 30 minutes of group discussion

13:00 DIGITIZATION

When we talk about digitization processes, it is necessary to understand that they can be done in countless different ways, qualities and techniques, from cell phones to high resolution captures or more complex systems such as multispectral and three-dimensional digitization. These differences have an impact on the amount and detail of information captured by each technique, the human and material resources necessary for each technique, maintenance, and long-term care for the digital surrogates

WEDNESDAY 20 MAY 2020

generated. The desire for digitizing cultural heritage materials must be linked to programs for the preservation of the materials being digitized and the preservation of all the digital files generated by digitization. The institution needs to find the balance between the resulting quality and quantity of the materials that will be digitized and what is possible to sustain in the long term through these processes. When we want the great, this can be the enemy of the good. The good can be done in the best way and be great. It is also possible to work with projects ranging from good to great according to decisions and selections made by the institution on how to deal with digital preservation, digitization, access and preservation in the long term. We want the best for our institutions and collections, and we aim for efficient dissemination programs using the materials generated by digitization. The purpose of this text is to help us think about our wishes for digitization and dissemination within this universe of pos-

13:10 ADVANCED IMAGING

Evaluating the Application of ISO 19264 Color Validation Techniques for 3D Imaging, Chris Heins and Scott Geffert, The Metropolitan Museum of Art (US); abstract only

The focus of our work is to develop and apply an ISO19264 compliant test method for objective color analysis of 3D renders. In recent years 3D imaging has grown in popularity, accessibility, and utility in the world of cultural heritage. As we develop workflow and standards, one area of major concern is color fidelity. Following ISO 19264 scene referred color calibration standards, we can validate the color quality of our 2D photography. If we follow the same protocols for the source photos used to generate 3D model textures, does the calibrated color survive all the way through to the model and on the platforms where the model is viewed/used? Is there a quantifiable benefit to applying scene referred color calibration to source photographs for photogrammetry? Does the choice of source image color space affect the outcome?

13:20 PRESERVATION

Waste Not, Want Not: Assessing the Environmental Sustainability of the University of Houston's Digital Preservation Program, Bethany Scott, University of Houston Libraries, and

The University of Houston Libraries previously had no data surrounding the environmental sustainability of its digital preservation program. We set out to gather this data and package it in a way that can be communicated easily to stakeholders such as Libraries administration. Additionally, we explore ways that the digital preservation program could become more environmentally sustainable in the future, and we provide actionable recommendations that other digital preservationists can quickly and easily implement to reduce the carbon footprint of their organization's digital preservation program.

13:30 GROUP Q&A

Join other attendees for a discussion of the three papers presented in this session.

14:00 DAY ENDS

Feel free to continue the discussions. Join other attendees in one of the technical topic discussion rooms, the coffee break area, or the exhibit hall.

SESSION VII

Session Chair: Walther Hasselo, Heritage Leiden (the Netherlands)

10:00 - 11:00 EDT

10:00 DIGITIZATION

FOCAL TALK Smithsonian 3D Pipeline, Vincent Rossi and Jonathan Blundell, Smithsonian Institution (US); extended abstract A-1*

The Smithsonian is developing a suite of open source tools to produce, manage, and deliver 3D assets from the Institution's collections at an ambitious scale. The beta versions (and eventual full releases) of these tools are being made freely available as open source projects to other museums, learning institutions, and commercial entities worldwide.

10:30 DIGITIZATION

FOCAL TALK Automated 3D Mass Digitization for the GLAM

The European Cultural Heritage Strategy for the 21st century has led to an increased demand for fast, efficient, and faithful 3D digitization technologies for cultural heritage artefacts. Yet, unlike the digital acquisition of cultural goods in 2D which is widely used and automated today, 3D digitization often still requires significant manual intervention, time, and money. To overcome this, the authors have developed CultLab3D, the world's first fully automatic 3D mass digitization technology for collections of three-dimensional objects. 3D scanning robots such as the CultArm3D-P are specifically designed to automate the entire 3D digitization process thus allowing to capture and archive objects on a large-scale and produce highly accurate photo-realistic representations.

11:00 - 11:25 EDT e-Coffee Break / Exhibit Open

Join other attendees in one of the technical topic discussion rooms, the coffee break area, or the exhibit hall.

SESSION VIII

Session Chair: Fenella G. France, Library of Congress (US)

11:30 - 12:20 EDT

11:30 ADVANCED IMAGING

Conventional color imaging has three channels—R, G, and B. In multispectral imaging within the visible spectrum, the number of channels increases in order to improve color accuracy and estimate spectral reflectance factor. Image quality criteria important in multispectral imaging include colorimetric accuracy, sharpness, registration, and low noise. The color transformation matrix, connecting camera signals with CIE tristimulus values, affects color accuracy and the visibility of image noise and misregistration when the multiple channels are combined to a color-managed image. When the final goal is a color-accurate image for one set of illuminating and viewing conditions, the color transformation is often derived directly using nonlinear optimization minimizing the average color difference between spectrophotometer- and camera-based

^{*}See Appendix, after page 105.

colorimetric coordinates. Optimization requires starting values and least squares minimizing spectral or tristimulus RMS error is typically used. Although it is effective for achieving convergence, the optimized matrix can result in a large reduction in image quality caused by noise propagation via the color transformation matrix. These concepts are reviewed.

11:50 DIGITIZATION

FOCAL TALK Automating 35mm Photographic Film Digitization: X-Y Table Capture System Design and Assessment,

Michael Bennett, University of Connecticut Library (US) 68 35mm still image formats are some of the most abundant photographic film types in cultural heritage collections. However, their special handling needs coupled with high resolution digital capture requirements have traditionally posed logistical constraints with regard to the formats' digitization at scale. Through the use of a programmable X-Y table camera capture system, both slide and strip 35mm photographic film can be digitized in an automated fashion following Federal Agencies Digitization Guidelines (FADGI).

12:20 - 12:55 EDT e-Coffee Break / Exhibit Open

Join other attendees in one of the technical topic discussion rooms, the coffee break area, or the exhibit hall.

SESSION IX

Session Chair: Ulla Bøgvad Kejser, Det Kongelige Bibliotek/The Royal Library (Denmark)

13:00 - 14:00 EDT

Three 10-minute papers followed by 30 minutes of group discussion

13:00 DIGITIZATION

Machine Learning and IIIF are popular topics today when it comes to digitisation projects and digital humanities. But are these really practical topics or just buzzwords? Are these rather exclusive technologies of some elite cultural and research institutions? Or can everyday digitisation projects with less exquisite materials really benefit from such technologies?

The example of the community around the open source software Goobi shows what the reality of numerous digitisation projects really looks like. What is no longer just theory and can be used in everyday life without having to develop software yourself? And what added value can actually be expected here?

13:10 ACCESS

Digitization projects of analog photographic collections are still growing in number, and therefore such assets of images become bigger continuously. Also, there is a strong trend towards open data and interfaces to access and reuse the image resources (FAIR data). To be able to search and find images in a repository, metadata of a certain depth must be existing. Typically, indexing and valorization, done by experts that know

the (photographic) collections, is necessary to achieve such meta-information. There are various metadata standards based on different concepts for the description of collections. Some, like ISAD(G), are more related to the physical structure of archives, others, like CIDOC-CRM, take into account the content of the images in detail. Enhancing the depth of indexing increases the time necessary drastically. It is also a task that is not easily scalable because specific content related knowledge is necessary. With the assistance of artificial intelligence, historic photographic collections could potentially be enhanced with metadata semi-automatically. For the successful application of machine learning, it is essential to have robust training sets.

In the presented paper, we show our observations in monitoring participants indexing historic collections of photographs. In the observations of workshops of people working with photographic heritage, it was monitored how single photographs but also image groups are described. Based on that knowledge, machine learning components can be trained and optimized for that particular type of source material. The demonstrated approach has the potential to support the work of valorization substantially. In addition, the approach has, to some extent, the potential to preserve the fundamental structures of knowledge of contemporary witnesses.

13:20 DIGITIZATION

The Digital Production Lab in John C. Hodges Library at the University of Tennessee, like many university imaging studios, has long relied on a seasonal and temporary labor force drawn from our student body. Improving human-computer interaction in our production workflows should reduce training time and errors while increasing throughput and worker confidence.

In this paper, we present our on-going efforts adapting Elgato's Stream Deck XL hardware to control a computer running Apple's macOS operating system and Phase One's Capture One Pro software using Python and AppleScript code. The paper outlines our custom code linking these parts together, describes how this relatively inexpensive input device streamlines our digitization process, and includes ideas for future application.

13:30 GROUP Q&A

Join other attendees for a discussion of the three papers presented in this session.

14:00 DAY ENDS

Feel free to continue the discussions. Join other attendees in one of the technical topic discussion rooms, the coffee break area, or the exhibit hall.

THURSDAY 21 MAY 2020

SESSION X

Session Chair: Jeanine Nault, Smithsonian Institution (US)

10:00 - 11:00 EDT

10:00 ACCESS KEYNOTE

Mind the Gap: Shifting the Gender Balance Online with Cultural Collections, Effic Kapsalis, Smithsonian Institution Provost Office (US); abstract only

US citizens view cultural organizations as trusted resources in a societal landscape where factual information is debated. At the same time, reflecting the increasingly diverse US populations is key to the future sustainability of US cultural organizations. The Smithsonian American Women's History Initiative (womenshistory.si.edu), launched in 2017, declared that it will be a resource for a more complete and diverse American history. A nearly 175-year-old organization like the Smithsonian does have major gaps in representation throughout its long history. This keynote highlights the tools and resources the Smithsonian is deploying to rapidly increase representation across its diverse digital ecosystem with its more than 155 multidisciplinary collections.

11:00 - 11:25 EDT e-Coffee Break / Exhibit Open

Join other attendees in one of the technical topic discussion rooms, the coffee break area, or the exhibit hall.

SESSION XI

Session Chair: Martina Hoffmann, Martina Hoffmann Consulting (the Netherlands)

11:30 - 12:40 EDT

11:30 ACCESS

Mapping Oral Histories: Augmenting Digital Audio Collections with GIS, Virginia Dressler, Kent State University (US) 88

The paper will focus on a project at Kent State University using a local oral history digital collection. The project displays the potential of how the application of an additional layer of geospatial information into an existing digital collection can improve user access and provide alternate methods to browse material (geographically). Transcriptions from the May 4 oral history collection at Kent State University were analyzed and tagged at any point there was a mention of one of the location points of interest. A new website was created where oral histories could be browsed using a historical map from the time period (spring 1970). This paper will outline the project and provide some initial steps for other institutions to begin such a project.

11:50 ACCESS

Linked Open Data Prototype of the Historical Archive of the European Commission, Mariana Damova, Mozaika Ltd.

In the age of the WWW consistent efforts have been made to make information from the archives available for the general public to facilitate access to the wealth of documentary history for research, consultation, or education purposes. Linked Open Data (LOD) provide a well suited framework to expose archival content to the general public while enriching it with content from other sources. This paper describes the creation of the first of its kind linked open data prototype to access data from the Historical Archive of the European Commission (HAS), carried out within ISA² programme of the European Commission. We present the designed ontology based on ISAD(G), ISAAR(CPF) and RIC-CM models and the business processes of HAS, and the created knowledge base from a sample of HAS data, re-using authority lists from the

Publication Office and EuroVoc and allowing querying via SPARQL endpoint.

12:10 DIGITIZATION

200 Manuscripts in 200 Days: High Throughput Digitization of the Advocates' Manuscript Collection, Gavin Willshaw, National Library of Scotland (UK)

The National Library of Scotland is pioneering a new approach to the digitization of medieval manuscripts, using the Book2Net Dragon system to image over 200 fragile and unique volumes in just one year. This paper provides an overview of the project to date, highlighting the balance the Library has been able to reach between maintaining quality and high throughput while minimizing damage made to the material. It discusses why the Library purchased the Dragon system, summarizes its key features, and provides an honest assessment of the Library's experience to date. Furthermore, the paper outlines how this new work strand has been integrated into the wider digitization program and the strategic aim to have a third of collections in digital format by 2025.

12:30 CLOSING REMARKS

e-RECEPTION

12:40 - 14:00 EDT

Join other attendees in the coffee break area for an e-reception featuring home concerts by members of our community.

RECORDED TALK FOR VIEWING

ACCESS

In the 1960's Peter J. Scott and colleagues at the now National Archives of Australia developed a new way of documenting records known as the Australian 'Series' System. Adopted by public records institutions in Australia and New Zealand, and selectively around the world, this approach forms the basis of the National Archives Commonwealth Record Series (CRS) system. In 2018 following views expressed that digital records pose a serious challenge to traditional ways of contextualization it was decided to review the CRS system in this respect. This paper looks at the process of that review and the eventual development of an enhanced model merging concepts from PREMIS with the CRS to enable a more flexible approach of documenting records in all forms.