IS&T Archiving Conference (ARCHIVING 2017)

Riga, Latvia
15 – 18 May 2017

ISBN: 978-1-5108-5607-3

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.

Printed by Curran Associates, Inc. (2018)

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 MAY 15, 2017

ARCHIVING 2017 SHORT COURSE PROGRAM

8:00 – 12:00 (4 hours)

ArchSC01: Computational Photography Techniques for Cultural Heritage Documentation and Archiving: Reflectance Transformation Imaging (RTI) and Photogrammetry
Instructor: Carla Schroer, Cultural Heritage Imaging

ArchSC02: Scanner & Camera Imaging Performance: Ten Commandments
Instructors: Don Williams, Image Science Associates, and Peter Burns, Burns Digital Imaging

ArchSC04: File Formats for Preservation
Instructors: Benjamin Yousefi, National Archives of Sweden, and Bert Lemmens, PACKED

13:30 – 15:30 (2 hours)

ArchSC05: Spectral Imaging—Digital Spectral Capture
Instructor: Fenella G. France, Library of Congress

ArchSC06: Fundamentals of Color Measurement
Instructor: David R. Wyble, Avian Rochester, LLC

ArchSC07: Digital Collection Development
Instructor: John Sarnowski, ResCarta Foundation

ArchSC08: PDF/A Challenges and Validation Tools
Instructors: Carl Wilson, Open Preservation Foundation, and Boris Doubrov, Dual Lab

15:45 – 17:45 (2 hours)

ArchSC09: Spectral Imaging—Digital Spectral Image Processing
Instructor: Fenella G. France, US Library of Congress

ArchSC10: Quality Assurance Workflows for Digitization Projects
Instructor: Martina Hoffmann, National Library of the Netherlands (KB)

ArchSC11: Introducing the Open Source Software Suite Kitodo: More than just a Workflow Tracking Tool
Instructor: Frank Ulrich Weber, Zeutschel GmbH

ARCHIVING 2017 WELCOME RECEPTION

17:45 – 19:30
Latvian Museum of Photography
Mārstaļu 8, Riga Central District
Join colleagues following the short course program for drinks, snacks, a tour of the museum, and the photography exhibit "Sparks."

TUESDAY MAY 16, 2017

9:00 – 10:25

WELCOME AND OPENING PANEL

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

9:00 **Welcome**

9:25 **The State-of-the-Art of Archiving in the Baltics**
Moderator: Uldis Zarins, director of development, National Library of Latvia
Panelists:
Rimvydas Laužikas, professor, Vilnius University (Lithuania)
Raivo Ruusalepp, director of development, National Library of Estonia
Arturs Zogla, head of digital library, National Library of Latvia

Memory institutions of the Baltic States have undergone a rapid paradigm shift during the past two decades, shedding the last remnants of old Soviet approaches, embracing the digital shift, and fundamentally changing the way they perceive and carry out their public service mission. This journey, however, has been all but easy. On one hand they have benefited greatly from increased access to the existing best practice, know-how, and tools, as well as financing opportunities. On the other hand, practical progress has been somewhat hindered by limited resources to implement the change. During the panel discussion, the participants reflect on the biggest challenges and achievements, present the current state of digital archiving in the Baltic States, and provide an outline of future developments. We delve into such aspects of archiving work as addressing the expectations of different groups of stakeholders—including funders and the general public—collaboration among different sectors and organizations, and the ways to address skill gaps.

10:25 – 11:05

COFFEE BREAK / EXHIBITS OPEN / PREVIEW POSTERS

Conference Center Lobby, level -1

11:05 – 12:20

ASSET AND COLLECTION MANAGEMENT I

Session Chair: Kari Smith, MIT Libraries, Institute Archives and Special Collections (USA)

11:05 **My Precious Information—How to Preserve It?,** *Anssi Jääskeläinen, Miia Kosonen, and Liisa Uosukainen, South-Eastern Finland University of Applied Sciences (Finland)* **1**

Do you think your information remains safe inside a cloud? Do you have another truly trustworthy place where you can store all your precious information? These questions lead us to the basic problem behind this paper: None of the official instances are interested in materials possessed by average Joes and Janes. You will have to be politically or otherwise important person to get your personal life story into official digital repositories. We at the Digitalia (Research Center on Digital Information Management) at South-Eastern Finland University of Applied Sciences, believe that there is a strong need for a digital preservation service that would give ordinary citizens the right to decide what to do

with their personal information. It is not right that common folks must rely on cloud drives with dubious terms and conditions or unreliable portable or optical devices to store their precious digital information. This article describes an initiative of a low cost full-scale digital archive solution that will be available to common people.

11:30 **TIFF in Archives: A Survey about Existing Files in Memory Institutions,** *Peter Fornaro, Lukas Rosenthaler, and Erwin Zbinden, University of Basel, and Martin Kaiser, KOST-CECO (Switzerland)* **6**

One of the most widespread formats used to represent high quality image data is the TIFF format. TIFF is a well-known, established, flexible, adaptable file format for handling images and data within a single file. The flexibility of TIFF allows for many different variants and can also include metadata, which follows other format definitions such as IPTC-data, EXIF-data or ICC-data for color transformation. Therefore TIFF is a complex file format that can be problematic for the use in archives, even though it is still the most common option for most GLAM institutions.

The aim of the TI/A initiative was to find a proper subset of tags for the use of TIFF in archival environments. To select proper features in such a recommendation, it was necessary to analyses existing files first. In this paper we present the results of two surveys that have been done in this context:

A) The analysis of about 4 million TIFF files stored as digital assets in memory institutions. The files represent a large variety of TIFF formats, regarding e.g. compression schemes, quantization depth, and date of creation. B) A survey about the number, use, and relevance of digital files in archives, museums, and libraries. The survey was done in the context of an ongoing project of the Swiss government to find a sustainable strategy for archiving digital cultural heritage objects.

11:55 **Archiving Websites Containing Streaming Media,** *Howard Besser, New York University (USA)* **11**

The software most North Americans use to archive websites is notably deficient in capturing streaming media. This paper reports on a project to re-architect the Heritrix web crawler in a newer approach to archiving websites. The project focuses on web content produced by contemporary young composers, and also explores developing relationships with these creators that address other web archiving issues such as copyright and high quality capture

12:20 – 12:40
EXHIBITOR PROFILES

Session Chair: Don Williams, Image Science Associates (USA)

12:40 – 14:00
GROUP LUNCH

Restaurant Klīversala, level 1

14:00 – 15:15
ADVANCED IMAGING I

Session Chair: Peter Fornaro, University of Basel (Switzerland)

14:00 **From Closed Testaments to Books: Virtual X-Ray Reading as an Alternate Digitization Technology for Fragile Documents,** *Fauzia Albertin[1], Marilisa Romito[1], Eva Peccenini[2,3,4], Matteo Bettuzzi[2,3,4], Rosa Brancaccio[2,3,4], Maria Pia Morigi[2,3,4], Monica Del Rio[5], Dorit Raines[6], Giorgio Margaritondo[1], and Demetri Psaltis[1]; [1]École Polytechnique Fédérale de Lausanne (EPFL) (Switzerland), [2]Enrico Fermi Center, [3]University of Bologna, [4]Italian National Institute for Nuclear Physics (INFN), [5]Venetian State Archive, and [6]University of Ca' Foscari (Italy)* **14**

In recent years, research and technology made considerable progress in increasing the speed and the safety of the entire digitization process of ancient collections. Despite this, imaging ancient, fragile or unopened documents remains a formidable challenge. We employ an alternate digitization technique for hand-written documents, exploiting x-ray tomography: Virtual X-ray Reading. Thanks to the high penetration of x-rays, we can acquire 3-dimensional (3D) tomographic images of a multi-page document without opening it. The x-ray contrast necessary for the readability is produced by the chemical composition and the consequent strong x-ray absorption of the iron gall inks—largely used for European handwritten documents. We present the development of this technology, from the chemical investigations of the inks to the tomography of an unopened Venetian testament and of an 18th century, 200-page, handwritten book.

14:25 **Precise 3D Documentation—Between the Need of a High Resolution and the Limit of Visualization Possibilities,** *Eryk Bunsch, Museum of King Jan III's Palace at Wilanów, and Robert Sitnik, Warsaw University of Technology, (Poland)* **19**

Since 2007 Museum of King Jan III's Palace at Wilanów and Warsaw University of Technology, Faculty of Mechatronics are together developing precise, structured light-based scanning methods. This cooperation was established in order to elaborate solutions allowing documenting characteristics of the surface of different cultural heritage objects. The gathered data are intended to support the processes of conservation, education, historical analysis, as well as the sharing of the visualizations of the especially fragile objects. In order to fulfill those requirements for most of the historic artifacts, scanning with a spatial resolution of at least 2500 points per square millimeter is needed. As a result of those assumptions, files of a very large size are produced. Today's software environments for such a huge data processing and applications to visualize those data are very limited. Consequently, this raises the need for either a significant simplification of visualization process or a reduction of the shared results of measurements (by showing data concerning only a small parts of the heritage objects). Which path would be most suitable for the end-users? Should this insufficiency of the visualization software cause a reduction of the quality of the measurement processes?

14:50 **Image-based Relighting Using Environment Maps,** *Michael Tetzlaff and Gary Meyer, University of Minnesota (USA)* **23**

An image-based relighting algorithm has been extended so that it can accommodate environment based lighting. Camera mounted flash photographs, employed in the original relighting algorithm, are also used to achieve the environment map based relighting results. In addition to preserving the simple equipment and setup utilized in the original relighting approach, the new method allows professional studio lighting effects, simulation of museum gallery illumination, and outdoor lighting at particular times of the day and year.

15:15 – 15:55
COFFEE BREAK / EXHIBITS OPEN / PREVIEW POSTERS

Conference Center Lobby, level -1

15:55 – 17:10
ACCESS, DISSEMINATION, AND USE I
Erik Landsberg, Cultural Heritage Digitization Consulting (USA)

15:55 **Simple Image Presentation Framework (SIPI)—An IIIF-based Image-Server,** *Lukas Rosenthaler, Peter Fornaro, Andrea Bianco, and Benjamin Geer, University of Basel (Switzerland)* **28**

The International Image Interoperability Framework (IIIF) is a widely accepted and fast growing standard to present images as web-resources. The IIIF-standard defines an URL-syntax to access, transform and reformat the desired image. An IIIF-server converts the image on-the-fly based on the desired parameters and transfers the image using the HTTP protocol to the client. We designed and implemented an advanced, extremely flexible, fully IIIF compliant server in C++11 offering advanced features that go beyond the IIIF standard. Due to its flexibility, can easily be integrated into existing environments and thus facilitates the transformation of existing archiving platforms to support the IIIF protocol.

16:20 **Content-based Interoperability: Beyond Technical Specifications of Interfaces,** *Tobias Schweizer, Lukas Rosenthaler, and Peter Fornaro, University of Basel (Switzerland)* **34**

On the technical level, the International Image Interoperability Framework (IIIF) offers a common standard for referring to images or parts of images using a well-defined URL-based syntax. This allows transformations like scaling, rotation, mirroring, and format conversions to be applied. However, for the full integration of digital objects in other environments, interoperability has to go beyond mere technical specifications for access. Non-technical metadata such as image descriptions, annotations, and references to other objects (e.g., texts, transcriptions, and other visual representations) must be available as well. Using the example of digital facsimiles of the scientific notebook of Jacob I. Bernoulli and its transcriptions, we show how XML together with the IIIF Image API allows for a seamless integration and merging of image and text sources in the same environment.

16:45 **Advances in Integrated Research Infrastructures for Science and Humanities Linked Data,** *Fenella France, Library of Congress (USA)* **39**

The continued challenge for data in any discipline is sustainable access, open source file formats, and the capacity for linked data. Collaborations with European and American colleagues indicates a shared concern, but with a less focused effort for establishing and recognizing the need for a more integrated approach to truly linked data, and the need for high level metadata embedded within datasets. Many related fields and disciplines have begun to focus on the need to integrate and assess approaches from colleagues—from materials science to archeology, botany, biology, and chemistry. The Research Data Alliance (RDA) has brought together a more cohesive approach to data management on the global scale. Developments for linked scientific data generated on heritage materials has continued to develop within the Library of Congress Preservation Research and Testing Division has engaged with colleagues in RDA and internationally to build upon existing standards and authorities, allowing greater credence for humanities and cultural heritage linked data. Further developments in the CLASS-D database structure enable the unique capability to link a range of types of scientific instrumental analyses back to original source materials, track samples and derivatives over time, and further the capability for web-accessible access to heritage collections.

WEDNESDAY MAY 17, 2017

9:00 – 10:10
WEDNESDAY KEYNOTE AND AWARDS
Session Chair: Ulla Bøgvad Kejser, Det Kongelige Bibliotek/The Royal Library (Denmark)

Collecting and Preserving the Born-Digital Heritage—New Aspects of an Old Challenge, *Raivo Ruusalepp, National Library of Estonia (Estonia)*

Under the new Legal Deposit Act the National Library of Estonia is receiving digital print files of all legal deposit publications since the beginning of 2017. The steady stream of a variety of new file formats into the preservation repository is posing new challenges to the preservation policy, preservation planning and tools, as well as to the rights management and various services the digital repository has to support. The new workflows also permit to transfer several quality control and compliance checking features into the pre-ingest phase, making the task of assessing preservation-readiness easier for the digital archive. At the dissemination end of the repository, new text mining services are expected by academic users that demand a radical re-thinking of both the structure of the AIP the repository is storing and the automated interfaces a library preservation repository should support.

This paper will build on the Estonian experience with handling the digital legal deposit and the approach that has been taken to begin solving the preservation challenges it poses. The main take-aways from this paper would be:

- implementing appraisal through pre-ingest processing in a legal deposit library;
- re-designing the AIPs for born-digital content and for matching with preservation policies;
- preparing digital preservation repositories for text-mining services.

10:10 – 10:30
2-MINUTE INTERACTIVE PAPER PREVIEWS
Session Chair: Don Williams, Image Science Associates (USA)

A Bottom-up Approach to Carry out Pre-Studies for the Implementation of Electronic Archives—A Case Study from a Swedish Organization, *Hugo Quisbert, ArkivIT (Sweden)* **44**

In this paper a bottom-up approach to carry out preliminary studies for the implementation of electronic archives is presented. Pre-studies are a standard procedure ahead of putting forward a project for the implementation of electronic archives. The top-down method, fundament for presented approach relies on the activities of Problem Analysis, Goal Analysis, Business Analysis, Change Needs Analysis, and Determination of Change Actions (iteratively and in that order). The bottom-up approach proposes change, positioning Goal Analysis after Business Analysis due to the way of stipulating goals from this approach become stronger since it put forward employees goals as the driving forces for change. The bottom-up approach more clearly discloses complexity and to some extent competing and disperse goals within studied organisation.

Using 3D Digitization in the Preservation of Industrial and Agricultural Heritage, *Tine Verroken, Texture; Bert Lemmens, PACKED; and Renee Mestdagh, CO_7 (Belgium)* **49**
The project titled 'Diving into the Machine' explores the use of existing 3D technology for the preservation and dissemination of industrial and agricultural heritage objects. The project runs a series of case studies purposely identifying best practices for the use of 3D models in collection management and presentation. The digitization technologies used vary in quality, efficiency, and cost. Each case study researches the sustainability of the 3D files produced and the feasibility of applying them in collection management practice. The project advocates the sustainable use of 3D technology by heritage organizations and raises awareness among 3D service providers about the use of archival standards and guidelines for producing sustainable 3D documents.

Open Source Software to Manage Digitalization Projects—The Kitodo Example, *Frank Ulrich Weber, Zeutschel GmbH (Germany)* **54**
Zeutschel today has more than fifty years of experience in the field of developing analog and digital exposure systems. The time goes by and now the digital business displaced the analog. Besides just producing digital images it is more and more interesting how to manage the digitization and how to make a valuable source of information out of the digital nuggets. Those questions arise in all sizes of institutions like libraries, archives, museums, and service providers. To offer them, beside just deliver the scanners, a complete solution for the whole range of digitization we decided to support the open source software suite Kitodo as one important part in our solution portfolio. That solution generates a valuable surplus for the client.

Digital Color Restoration from Slide Images which use the Color Target Kodak Q-13, *Alexandre Cruz Leão, Arnaldo de Albuquerque Araújo, and Luiz Antônio Cruz Souza, Federal University of Minas Gerais (Brazil)* **57**
This paper intent to collaborate with some Archive's collection, where there are color slides with necessity to restore their colors. To make the restoration, this research considers the images which has the Kodak Q-13 Color target. The Kodak Q-13 reference color targets (Gray Scale and Color) were, at first, developed to evaluate and correct chemical color photographs and not for digital images. The purpose of this research is the development of a methodology which makes possible the restoration of colors and tones in the pictures through digital image processing. For this purpose, it was made the colorimetric study in these targets, and the development of the methodology for digital processing. The results indicate better color matching consistency for the Gray Scale target than the color one. Results obtained from the experiments using different methodologies to development show that it is possible to accomplish color restoration of the pictures, which has used the Kodak Q-13 Gray Scale reference target.

Set of Methodologies for Archive Film Digitization and Restoration with Examples of Their Application in ORWO Region, *Karel Fliegel, Stanislav Vítek, and Petr Páta, Czech Technical University in Prague and Film and TV School of Academy of Performing Arts in Prague; and Miloslav Novák, Jirí Myslik, Josef Pecak, and Marek Jicha, Film and TV School of Academy of Performing Arts in Prague (FAMU) (Czech Republic)* **62**
In this paper, we present a set of verified methodologies suitable for application to a particular problem of archive films' restoration and digitization, especially when a nonstandard laboratory or creative techniques were used, which is typical for so-called ORWO region. The umbrella of the presented techniques is formed by established Digitally Restored Authorizate (DRA) methodology, with its aim to achieve the appearance of the audio and visual components of the digitized film as close as possible to the original author's concept. Among the methodologies, we present tools for objective assessment of perceived differences in the outcomes of the color grading process. These techniques are suitable for evaluation of appearance match among various available versions of the digitized film in respect to the DRA outcome.

Implementing a Video Framework based on IIIF: A Customized Approach from Long-Term Preservation Video Formats to Conversion on Demand, *Julien A. Raemy, Peter Fornaro, and Lukas Rosenthaler, University of Basel (Switzerland)* **68**
This paper addresses the issue of elaborating a structure for digital video assets based on the International Image Interoperability Framework (IIIF) concepts for the use in archival environments. With a view to tailoring a solution to fit the end user's needs, the dissemination copies of video material could be automatically converted on demand from their master files. Such a reduced data structure simplifies access to digital video sources but leads as well to simplified preservation due to reduced data volume and data complexity. Dissemination copies do not require specific dispositions for digital archiving anymore.

Memory institutions would greatly benefit from a technology that can be integrated into a Web-based infrastructure. In such a way video content can for example be embedded into flexible Virtual Research Environments which allow scholars to work and cite more accurately video resources using IIIF.

Developing ARCLib—An Open-Source Solution for a Bit-level and Logical Long-term Preservation, *Andrea Miranda, The Czech Academy of Sciences, and Zdenek Hruska, Moravian Library (Czech Republic)* **74**
This poster informs about the Czech ARCLib project. One of the main goals of the project is the development of an open-source solution for a bit-level and logical preservation of digital documents, respecting the national and international standards as well as the needs of all types of libraries in the Czech Republic. The mission of the ARCLib project lies, among others, in creating a solution that will allow institutions to implement all of the OAIS functional modules and entities, considering institutions' information model. The architecture is planned as open and modular and the final product will be able to ingest, validate, and store data from a majority of software products used for creating, disseminating, and archiving libraries' digital and digitised data in the Czech Republic.

10:30 – 11:20
INTERACTIVE PAPER (POSTER) SESSION / COFFEE BREAK / EXHIBITS OPEN

11:20 – 12:35
IMAGING PERFORMANCE AND STANDARDS I
Session Chair:Thomas Rieger, Library of Congress (USA)

11:20 **Automatization in (Mass) Digitization QA-workflows,** *Martina Hoffmann, National Library of the Netherlands (the Netherlands)* **78**
In setting up a QA workflow—or any other type of workflow—one tries to make processing faster, better and more efficient. As we are often dealing with vulnerable originals, work on those documents can only be automated to a certain extent, but within the scope for automation, all opportunities should be used. Based on the example of the Netherlands

large digitization program Metamorfoze (specifically the Archives and Collections section) this paper will give an example on how to achieve such optimum automation for QA control on data-integrity and will try to answer key questions on automation as they are the starting point for a better QA-workflow.

11:45 **Extensions to OpenDICE: Batch Image Assessment and Additional Target Support,** *Lei He, Library of Congress (USA)* **83**
In this paper we present extensions to our digital imaging quality assessment software, OpenDICE. We have added three features; batch image assessment, the support to very large size targets, and two additional color patch targets. Batch image assessment provides the capability to monitor imaging device performance over long time periods. The addition of the very large size targets provides more comprehensive and accurate resolution assessment of the imaging system at different locations and orientations. The inclusion of two more color targets enhances color target profiling performance assessment.

12:10 **Evaluating Perceived Capture Quality for the Digitization of Cultural Heritage Objects,** *Susan Farnand, Rochester Institute of Technology, and Franziska Frey, Harvard Library (USA)* **88**
Perceptual experiments were used to evaluate the capture quality attainable of four digitization systems. The study results showed that two of these systems produced images of limited use as digital masters. The perceived image quality for the other two systems was comparable for digitization purposes. While a variety of system characteristics must be given careful consideration when identifying equipment to purchase, a system unable to attain the perceived quality needed for usable images is of little value no matter how inexpensive and ergonomic it may be. Also, image quality cannot be defined by number of pixels. In this study, the system producing the largest files was not well rated. Using perceptual experiments helped clarify the utility of digitization systems.

12:35 – 13:45
GROUP LUNCH
Restaurant Klīversala, level 1

13:45 – 15:00
ASSET AND COLLECTION MANAGEMENT II
Session Chair: Kathrine Hougaard Edsen Johansen, Copenhagen City Archives (Denmark)

13:45 **Work Ethics for the *Digitizer*—Opportunities and Best Practices for Production of Digital Archives: The Working Experience of the Photographic Archive of Pompeii,** *Patrizio Gianferro, University degli Studi di Roma La Sapienza, and Rosa Myriam De Lillo, Luigi Sturzo Institute (Italy)* **93**
In this presentation, we intend to examine the ethical and work-related issues relative to mass digitization projects of photografic archives. Our paper asks: who and what is there behind these productions? Our goal is to analyze and contextualize the activities of the digitizer, the frontline and central figure in the transformation of the archives from analogue to digital.

Historically, the transmission of knowledge is based upon a continuous copying process; whereas monks once manually transferred content from manuscript to manuscript, contemporary digitizers follow in their footsteps, giving old analogue photographs a new digital life by utilizing contemporary transcription pathways.

These aspects will be examined starting with our working experience within the photographic archive of the Pompeii excavations.

During this experience we elaborate upon the idea that digitizers must no longer be considered as mere mechanical performers in the process of digital acquisition, but as consciously trained professionals, able to engage critically in the various processes involved in the creation of a digital archive.

14:10 **A Context Metadata Collection and Management Tool for Computational Photography Projects,** *Carla Schroer and Mark Mudge, Cultural Heritage Imaging (USA), and Erich Leisch and Martin Doerr, Institute for Computer Science, FORTH (Greece)* **99**
This paper will present the first module of an advanced set of metadata and knowledge management tools to record a "Digital Lab Notebook" (DLN), the equivalent of the traditional scientist's lab notebook. The DLN:Capture Context (DLN:CC) tool describes the means and context of photographic data capture. The tool is designed for broad use across computational photography technologies. The DLN:CC has already been implemented for Reflectance Transformation Imaging (RTI) and implementation for photogrammetry is underway. The collection and organization of contextual metadata is highly automated, facilitating use during the time the image data is captured and processed, rather than afterward. This project adds ISO-standard compliant metadata, which establishes the provenance of the imaging subject's digital surrogate. The captured photographic sequences and the DLN metadata contain all the information needed to generate and/or regenerate advanced, image-based 2D and 3D digital surrogates, such as Reflectance Transformation Imaging or photogrammetry's 3D models with texture. The DLN also provides each digital surrogate a scientific account of their collection and generation.

14:35 **Identifying Top Performing TF*IDF Classifiers Using the CNN Corpus,** *A. Marie Vans and Steven J. Simske, HP Inc. (USA)* **105**
TF*IDF (term frequency times inverse document frequency) is a common metric used to automatically discover keywords in documents for use in classification and other text processing applications. We are interested in determining whether these measures can help in classifying documents. There are multiple ways to define TF*IDF, but there has been no real attempt to determine the value of these different forms. We explore a large family of 112 TF*IDF measures (corresponding to an a priori estimate of 20 degrees of freedom among these measures) applied to 588 CNN articles belonging in 12 classes such as Business, Sport, and World. We postulate that at least some sets of these measures must be effective for classification. The goal is to use a set of TF*IDF measures that best match the a priori classifications by CNN. We also show that by combining the results of a few well-performing TF*IDF measures can increase classification results.

15:30 – 17:45 (times vary)
BEHIND-THE-SCENES TOURS

19:00 – 21:30
CONFERENCE RECEPTION
Maza Gilde (Small Guild building)
Amatu iela 3, Riga

THURSDAY MAY 18, 2017

9:00 – 10:00
THURSDAY KEYNOTE

Session Chair: Don Williams, Imaging Science Associates (USA)

Resonating Spaces: 3D Imaging of the Berlin Philharmonie,
Chris Edwards, J. Paul Getty Trust (USA)
The Getty Research Institute's (GRI) exhibition Berlin/LA: Space for Music (April 25–July 30, 2017) celebrates the 50th anniversary of the sister city partnership between West Berlin and Los Angeles by focusing on two buildings that have captured the public imagination and become iconic features of the urban landscape of both cities: the Berlin Philharmonic (1963), designed by Hans Scharoun, and the Walt Disney Concert Hall (2003), designed by Frank Gehry. Original physical working models created by Gehry in designing the Walt Disney Concert Hall give visitors insight into how its expansive interior was created, however no working or presentation models of Scharoun's Philharmonie are extant. The interior of Hans Scharoun's Berlin Philharmonic, which ranks among the most influential concert hall designs of the 20th century, is a very complex and multifaceted space. The bold decision made by the GRI's curatorial team to commission a 3D digital and printed model of the Philharmonic interior allows the curators to communicate to gallery visitors the innovative nature of this complex space in a manner much more evocative than photographs alone could convey, making this a truly groundbreaking undertaking.

10:00 – 12:45
ADVANCED IMAGING II

Session Chair: Steven J. Simske, HP Inc. (USA)

10:00 **Reflectance Transformation Imaging in Daguerreotype Investigation,** *Hembo Pagi[1], James Miles [2], Andres Uueni[1], Stephen Hogarth[3], and Kadi Sikka[4]; [1]Archaeovision (Estonia) [2]Arcaeovision (UK), [3]shogarth.com (Canada) and [4]Estonian Photographic Heritage Society (Estonia)* **116**
Reflectance Transformation Imaging (RTI) has been used for cultural heritage documentation since its introduction by Tom Malzbender in 2001. The technique allows for the recording of 3D surface reflectance properties and visualise them as 2D interactive images. The method can be used to investigate objects in various lighting conditions to enhance small surface changes, to bring out cracks, tool marks, scratches, pencil impressions, and many more features that are not visible to the naked eye. The method is a valuable tool when examining coins, writing tablets, and daguerreotypes, as features such as fine polishing lines, retouches, and deteriorations can be identified.

In this paper, an outline of the RTI technique will be given and a case study will be provided that encompasses a new processing algorithm for RTI that can extract greater levels of information than was previously available.

10:25 **Advances in Spectral Imaging Curve Analysis for Humanities Studies and Heritage Science,** *Fenella G. France, Meghan A. Wilson, and Chris Bolser, Library of Congress (USA)* **122**
Establishing standardized digital protocols for Spectral Imaging creates opportunities for non-invasive analysis of a wide range of heritage and archival materials. In addition to the capacity to reveal hidden and non-visible information, the creation of spectral curves from the response of materials throughout the visible and non-visible wavelengths allows us to identify and characterize inks and colorants as well as track changes due to environment or conservation treatments. The use of spectral curves for this purpose requires a spectral library of reference materials that can be used for comparison and identification of these heritage materials. Advancing our capacity to non-invasively analyze documents, manuscripts, textiles, and objects requires a rigorous standardized protocol that is reproducible and repeatable. Spectral curve analysis necessitates that all imaging metadata and parameters are consistent and materials are monitored to assure accurate replication.

10:50 – 11:30
COFFEE BREAK

Conference Center Lobby, level -1

11:30 **Next Generation Camera Calibration Target for Archiving,** *David Wyble, Avian Rochester, LLC (USA)* **127**
The desired features of a new camera characterization target are described. Any new target must minimally be justified by meeting or exceeding the characterization performance targets used in current imaging processes. Additionally, a target should be physically robust against normal laboratory handling, including the ability to be cleaned after being soiled. To be less sensitive to illumination and camera geometry, all patches in the target should have identical gloss and more generally the same BRDF characteristics. Finally, some colors should be included that are similar to those in anticipated materials to be imaged. These features will allow a more productive and continuous workflow without interruptions imposed by inadvertent target mishandling, and bring cost and time savings by eliminating unnecessary recalibrations. After describing the process for selecting the colors of such a target, the target camera characterization performance is compared against targets in common use.

11:55 **The Combination of 3D and Multispectral Imaging for Scientific Visualization—Tool for Conservation and Heritage Specialists,** *Andres Uueni and Hilkka Hiiop, Estonian Academy of Arts (Estonia), and Fabrizio Ivan Apollonio, University of Bologna (Italy)* **133**
Information and communication technology applications and the preservation, research, and popularization of heritage have attained an important position in the management and studies of cultural heritage. Over the past couple of decades, instrumental studies, imaging, and survey technologies have developed very rapidly due to information technology solutions and have become ever more important tools in documentation. Cultural heritage has lot of different aspects and there is often a need of continuous scientific research and preservation to gather valuable information from our history and provide better understanding for future generations. 3D survey and multispectral imaging provides enriched information about heritage objects that may considerably help conservation and heritage specialists especially when it is accessible over the Web.

12:20 **The Documentation and Investigation of Surface Deposits on a Tutankamun's Pottery Jar Using Advanced Imaging Techniques,** *Mahmoud Hassan, Grand Egyptian Museum (Egypt)**
The digitization and investigation processes play an important role in conservation science, as they construct the fundamental information that conservation treatments relies on and for the preservation of the object

*Paper not available at time of publication.

status for future treatments and studies. In this research a pottery jar from the collection of Tutankhamen is under investigation. The use of different techniques for digitization and investigation such as detailed photography, microscopy, and imaging spectroscopy, in addition to the historical research, gives the opportunity to document the object and identify different materials.

12:45 – 14:00
GROUP LUNCH
Restaurant Klīversala, level 1

14:00 – 15:15
ACCESS, DISSEMINATION, AND USE II
Session Chair: Jonas Palm, Swedish National Archives (Sweden)

14:00 **Unlocking the Archive: The US Defense Department's Analysis & Implementation of its Authority to Publicly Release Audiovisual Records,** *Julia Hickey, Defense Media Activity (USA)* **138**
Defense Media Activity (DMA) is the Department of Defense's (DoD) direct line of communication for news and information to U.S. military forces worldwide. The agency informs DoD audiences, entertains DoD audience overseas, trains Public Affairs and Visual Information professionals, and manages the DoD's visual information or audiovisual records.

The central audiovisual archive of the Department of Defense (DoD) holds imagery—still and motion media—dating to the 1890s, with millions of images ranging from World War I to current operations in the Middle East. The Defense Imagery Management Operations Center (DIMOC) of the Defense Media Activity centrally collects, processes, disseminates, and archives these audiovisual records documenting the many activities of the Department of Defense (DoD) and U.S. military around the globe. In an effort to make the entire DoD collection publicly available online, and as a cascading effect of the DMA digitization and storage contract initiated in 2013, DIMOC has developed and implemented a process to address a large portion of the collection that has not been reviewed for public release.

This paper is a follow-on to the 2016 IS&T: Archiving Conference paper titled "Unlocking the Archive: The Defense Department's Plan to Make Unreleased Audiovisual Records Public," which explored the historical context for the mixed DoD collection and the innovative solutions implemented by DIMOC to overcome barriers to clearing this content for public access. Specifically, this paper will focus on the implementation and standard operating procedures (SOPs) and processes implemented into DIMOC's workflow. The paper will also discuss the DoD's risk assessment for each review area: Freedom of Information Act (FOIA), Operational Security (OPSEC), Sensitive Subject (SS) and Public Affairs Guidance (PAG), and a Presidential Executive Order (EO) that established a mandatory 25-year review for declassification. Tied closely to DIMOC's contracted asset management system, the use of (metadata) automation and workflow process reengineering are results of the imagery collection's analysis conducted for this new public release responsibility.

14:25 **Using a Large Set of Weak Classifiers for Text Analytics,** *Steven J. Simske and A. Marie Vans, HP Inc. (USA)* **146**
TF*IDF is a common approach used for text mining and information retrieval. We have described a method for using 112 variations on the TF*IDF equation for the classification of 588 CNN news articles belonging to 12 different classes. We found that no single TF*IDF could accurately classify all the documents. In fact, the highest accuracy attainable by any single TF*IDF was 45%. In this article, we take the work further to show how different measurements utilizing the TF*IDF classification results can be used to show that some classes may be logically inconsistent as classes. These methods also may be used to create more cohesive classes.

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

Latvian National Archives

Latvian National Museum of Art

Museum of Energy (Latvenergo)

Museum of the Occupation of Latvia

National Library of Latvia

Special thanks to the Latvian Museum of Photography, the National Library of Latvia, and their staff for support of Archiving 2017.

14:50 **The Evolution of the US National Archives Catalog: From Access to Engagement,** *Michael Horsley, National Archives and Records Administration (USA)* **152**
Improvements to the United States National Archives and Records Administration's (NARA) on-line catalog enable delivery of high-resolution images to the public. New transcription and tagging tools in the catalog allow citizen archivists to engage with digitized records and increase access to archival material. NARA has adopted a multitude of social media platforms that have connected millions of patrons to digital content. In an ever-expanding online world, the public expects to find everything on the Internet. With holdings in the billions of archival records of every conceivable format, NARA has had to embrace an adaptive and scalable digitization approach. NARA has a long history of developing digitization guidelines that have proved invaluable to ensure image quality, meet mandates for preservation and access, and create useful master and derivative digital objects. This paper analyzes the parallel evolution of NARA's digitization standards and how they are applied in online catalogs and social media platforms.

15:15 – 15:45
COFFEE BREAK
Conference Center Lobby, level -1

15:45 – 17:15
IMAGING PERFORMANCE AND STANDARDS II
Session Chair: Christoph Voges, Hochschule für angewandte Wissenschaft und Kunst (HAWK), and consultant (Germany)

15:45 **JPEG2000 as a Preservation Format for Digitization: Lessons Learned from a Library,** *Laurent Duplouy, Bibliothèque Nationale de France (France)* **157**
This article attempts to present the methodology used to respond to ques-

tions and issues raised by the adoption of JPEG2000 format at the National Library of France for mass digitization. It attempts to describe particularly the methodology used to define a compression ratio for heritage digitization.

Finally, it presents lessons learned after two years of mass production.

16:10 **Automation of Data Integrity Checks in QA for Mass Digitization—A Case Study,** *Martijn van der Kaaij, Heron Information Management LLP (UK)* **160**

In early 2014, Heron Information Management LLP got the opportunity to put a generic approach to workflow automation into practice for the data integrity part of the Metamorfoze Archives & Collections Quality Assurance workflow as hosted by the Royal Library of the Netherlands. This approach, characterized by standardization, modularity and ease of operation, has resulted in a QA-workflow for mass digitization that covers the whole process of data integrity checks: from the production of inventories to the logging of results in such a way that they can be used in long term preservation.

16:35 **A Decade of Experience with Digital Imaging Performance Guidelines: The Good, the Bad, and the Missing,** *Don Williams, Image Science Associates, and Peter D.Burns, Burns Digital Imaging (USA)* **165**

As with any initiative, despite design intentions, the first efforts have unexpected positives and how-did-we-miss-that negatives. The FADGI (Federal Agency Digital Guideline Initiative) and Metamorfoze guidelines are no exceptions. Whenever such efforts are brought to practice in the field we learn. And that is good, because it forces behavior, software, and hardware to evolve to be more resilient. We address these developments in digitization for cultural heritage collections. We look-back on the last decade of our experience with these guidelines and discuss progress, limitations, and future directions.

17:00 **Closing Remarks,** *Ulla Bøgvad Kejser, Det Kongelige Bibliotek/The Royal Library (Denmark) and Don Williams, Image Science Associates (USA)*