iPRES2021

17th Internatinal Conference on Digital Preservation

iPRES 2021
17th International Conference
on Digital Preservation

PROCEEDINGS

Beijing · China // 19-22 October, 2021

ipres2021.ac.cn

Table of Contents

Long Papers

Short Papers

Posters

Panels

Workshops & Tutorials

Demonstration

Hackathon

Lighting Talks

Preserving Landmark Legacy Software with the Software Heritage Acquisition Process

Roberto Di Cosmo

Inria and University of Paris

Carlo Montangero

University of Pisa

Guido Scatena

University of Pisa

Laura Bussi

University of Pisa

Abstract

The source code of landmark software developed since the beginning of the computer era is a precious part of our cultural heritage, and needs to be properly rescued, curated, archived and made available to present and future generations. In this article, we present the Software Heritage Acquisition Process, that has been designed to provide detailed guidelines on how to perform this important task, preserving important historical information. This process has been validated extensively on several important pieces of software source code of historical relevance in the University of Pisa, in collaboration with UNESCO, and is open to all for adoption and improvement.

Keywords

Software Preservation
Legacy Software
Source Code
Acquisition Process

LONG
PAPER

COMPARING HOW TO TAKE CARE OF HUMAN' AND BIT-STREAMS' LIVES

Eld Zierau

Royal Danish Library

Abstract

This paper investigates a number of practices to ensure safety for human lives and compare them with practices to ensure the 'lives' of bit-streams. The selected practices for human lives are: A. common emergency preparedness practices for securement of places like shopping centers, B. safety critical systems commonly used for airplanes, space crafts and nuclear power plants, and C. pandemic preparation planning.

The results of the comparison are used to illustrate how human security precautions can be used in similar ways in a bit preservation case involving methods and systems on both the technical and the organizational level.

Keywords

Bit Preservation
Safety Critical Systems
Safety Procedures
Pandemics
Risk Management

L ONG
PAPER

IDENTIFICATION OF MULTI-PART DIGITAL OBJECTS

Jack O'Sullivan

Preservica Ltd

Richard Smith

Preservica Ltd

Jonathan Tilbury

Preservica Ltd

Abstract

Digital information often relies on the interactions of multiple individual digital files, rather than being completely encapsulated in a single file. Where this happens, all necessary files should be considered to be part of the same Multi-Part digital object. This paper describes a data model for describing the format of such digital objects, or more specifically, the Multi-Part Representations of such objects. As with the identification of file formats, having a Representation Format like this enables us to determine which tools to use to undertake particular preservation actions for a Representation.

Keywords

Format Identification
Data Model
Complex Objects

**L ONG
PAPER**

RESEARCH AND PRACTICE ON DOMAIN ONTOLOGY OF ANCIENT CHINESE ARTIFACTS

Yipei Ye

The Palace Museum

Abstract

Collection information is one of the important resources that has long-term preservation value to museums. The paper proposes a domain ontology centered on ancient Chinese artifacts, aiming to help any organizations or individuals who own ancient Chinese artifacts collection better organize and manage their collection information, furthermore, improve collection search experience so that users can retrieve objects in the collection more easily and efficiently to ensure the availability of information as the digital object. The paper elaborates the creation process of "Ancient Chinese Artifacts Conceptual Reference Model" (ACACRM), summarizing and sharing practical experience for providing reference for museums with similar needs and application scenarios.

Keywords

Museum
Collection Searching
Ontology
Knowledge Graph
The Palace Museum

L ONG
   PAPER

Scalable and Sustainable Long Term Digital Preservation of Scientific Datasets

Matthew Addis

Arkivum

Abstract

The European Commission supported ARCHIVER project (Archiving and Preservation for Research Environments) aims to "introduce significant improvements in the area of archiving and digital preservation services, supporting the IT requirements of European scientists and providing end-to-end archival and preservation services, cost-effective for data generated in the petabyte range with high, sustained ingest rates, in the context of scientific research projects". This paper presents a software solution developed by Arkivum to meet the needs of long-term digital preservation of scientific datasets in ARCHIVER. We present and discuss how this solution is scalable (able to process and store very large volumes of research data) and sustainable (both economically and environmentally). This is achieved through a combination of serverless computing, deployment on hyperscale infrastructure, and implementation of configurable 'Minimum Effort Ingest' workflows. In particular, we show how high-performance and scalable Long Term Digital Preservation (LTDP) of very-large datasets can be done in a way that is entirely compatible with high levels of cost-efficiency and minimized environmental impact.

Keywords

Scalability
Sustainability
Environment
Cost
Research Data

L **ONG**
PAPER

DIGITAL SERVICE DESIGN FOR MUSEUMS BASED ON DATA CURATION: TAKE THE CONSTRUCTION OF ONE-STOP ONLINE DIGITAL PLATFORM OF THE PALACE MUSEUM AS AN EXAMPLE

Xiaolu Kang

The Palace Museum

Abstract

With a collection of more than a million works and an early start in informatization, the Palace Museum manages a large number of cultural artifacts and digital assets. At present, the Palace Museum is committed to improving the active management of its digital content. In particular, the Museum aims to strengthen the long-term preservation of digital content to meet the preservation and utilization challenges brought about by the phenomenal growth of its digital content. This paper analyzes the top-level design, implementation, and evaluation of the Palace Museum's one-stop online digital platform through the concept of data curation and the DCC curation lifecycle model. The analysis discusses the positive role that digital curation plays in practice for digital content production, preservation, and service and provides a reference point for other museums.

Keywords

Data Curation
Digital Content
One-Stop Digital Service of the Palace Museum
Curation Lifecycle
Network Information Resources

L **ONG**
PAPER

Research on 3D Digitization Scheme of Cultural Relics Preserved for a Long Time

Peipei Zhang

The Palace Museum

Huan Liu

China Ordnance
Industry Survey and
Geotechnical Institute

Chenchen Hou

The Palace Museum

Abstract

With the development of digital technology, museums are facing a great challenge in the preservation and utilization of 3D data of cultural relics. Based on the demands on long-term preservation of 3D data of cultural relics by museums, a solution of 3D digitization of cultural relics is proposed in this paper for the unbalanced cost-efficiency-result, rare data utilization and inconsistent data storage scheme in the process of digitization. To solve the problem of unbalanced cost-efficiency-result, an evaluation model of 3D digitization has been designed by the Palace Museum, namely comprehensive "cost-efficiency-result" evaluation model. By adding a QR code control device for close range photogrammetry of cultural relics the high digitization cost, low digitization efficiency and inconsistent data quality are solved. Based on the specific needs of preservation and utilization, different grades are specified for 3D data of cultural relics, and a three-layer hierarchical structure and storage scheme is designed in this paper, which solves data inconsistence and difficult data sharing. The above methods have been widely used in the digitization project of the Palace Museum, and the related studies and practices also provide reference for the further development of national 3D digitization standards.

Keywords

Museum
Cultural Relics
3D Data
Photogrammetry
Accuracy Control

LONG
PAPER

CONSTRUCTION OF MANAGEMENT PLATFORM FOR 3D MODEL OF CULTURAL RELICS

Fang Liu
....

National Museum of China

Huabiao Li
....

National Museum of China

Chao Yang
....

National Museum of China

Abstract

With the continuous enrichment of Museum digital resources, more and more attention has been paid to the management of them. This article introduces the practice of the National Museum of China in the long-term preservation of high-definition three-dimensional images of cultural relics in terms of data acquisition, quality control, storage and security protection mechanisms.

Keywords

3D Model
Cultural Relics
Preservation
Management Platform

LONG
PAPER

The Significant Properties Of Spreadsheets

Remco Van Veenendaal

National Archives of the Netherlands

Lotte Wijsman

National Archives of the Netherlands

Asbjørn Skødt

Danish National Archives

Kati Sein

National Archives of Estonia

Jacob Takema

National Archives of the Netherlands

Jack O'sullivan

Preservica

LONG PAPER

Abstract

When preserving spreadsheets, in certain cases it is decided to convert the spreadsheet file format to an image-based file format such as TIF. However, during this conversion, a loss of information can occur. Certain functionalities cannot be translated to the new file format and therefore lose their meaning. An example of this are formulas. When converting towards an image-based file format, the outcome of the formula will still be displayed, but the calculation that is behind it is lost. It is therefore imperative to find out which properties in spreadsheets are significant, such that these can be preserved. The Open Preservation Foundation's Archives Interest Group (OPF AIG) decided to research this using the InSPECT methodology, which consists of both an Object Analysis and a Stakeholder Analysis. These two types of analysis eventually led to a longer report which was published, in which not only our results were shared, but also a reusable and practice-tested method to apply the InSPECT methodology. In this paper, we present the main findings of our research.

Keywords

Significant Properties
Spreadsheets
Preservation
Object Analysis
Stakeholder Analysis

Exploration of Preservation Metadata Towards Medical Resources Long-Term Archiving

Chenliu Yang

Institute of Medical Information, Chinese Academy of Medical Sciences

Jiahui Hu

Institute of Medical Information, Chinese Academy of Medical Sciences

Qian Wang

Institute of Medical Information, Chinese Academy of Medical Sciences

Lei Wang

Institute of Medical Information, Chinese Academy of Medical Sciences

Kuanda Yao

Institute of Medical Information, Chinese Academy of Medical Sciences

An Fang

Institute of Medical Information, Chinese Academy of Medical Sciences

Abstract

Preservation metadata is the significant component of digital long-term preservation, includes the information about data processing, and undertakes the functions maintaining the reliability, integrity, authenticity, renderability, understandability of archived resources. This paper describes MedPRES preservation metadata-related work, construction requirements with the foundation of Open Archival Information System (OAIS) architecture, corresponding to PREMIS data model which contains object, event, agent, rights entities, as well as data dictionary, illustrates metadata practice in MedPRES key processes as well.

Keywords

Long Term Preservation
PREMIS
Preservation Metadata
OAIS
Semantic Units

LONG
PAPER

WIKIDATA: A MAGIC PORTAL FOR SIEGFRIED AND ROY

Ross Spencer

Ravensburger AG

Katherine Thornton

Yale University Library

Richard Lehane

International Atomic
Energy Agency

Euan Cochrane

Yale University Library

Abstract

A project team was established in late 2019 to create a Siegfried/Wikidata integration. The project was established to make Yale University Library's work curating digital preservation data in Wikidata actionable; that is, enable it to be consumed by the Siegfried utility for the purpose of file format identification. Once accessible by such a tool, file-system data can be linked directly to Wikidata records and thus the breadth and depth of the Wikidata dataset. The combination opens up possibilities around the EaaSI work also being developed with Wikidata in mind. The Siegfried/Wikidata collaboration was an intentional one that would have been difficult to make happen through normal circumstances; be that geographic or organisational dispersal, opportunity, or cost. The project was seen as a worthwhile endeavour to pursue independently of these restrictions and other projects may in the future need to resort to non-traditional approaches to achieve their aim. The first iteration of this work successfully went live in October 2020 and for it to continue to succeed there is a call to arms from those in the digital preservation and Wikidata communities to help push it further. Work in the pipeline is believed will continue to demonstrate the worthwhile nature of today's commitment.

Keywords

Wikidata
WikiDP
Format-Identification
Siegfried
Collaboration

LONG
PAPER

Preserving for People: Observing Humanities Scholars' Research Practices in a Hybrid Archive Environment

Alexandra Leigh

City, University of London; The National Archives, UK

Stephann Makri

City, University of London

Alex Taylor

City, University of London

Alec Mulinder

The National Archives, UK

Sarra Hamdi

The National Archives, UK

Sonia Ranade

The National Archives, UK

LONG PAPER

Abstract

In order to assess the potential suitability of digital preservation efforts for future research, it is necessary to understand how users interact with information in the present. Yet there is very little information on how humanities researchers – a key user group for archives – interact with archives beyond discovery. In the following, we show the importance of recognising end-users as part of wider information workflows that comprise not only discovery but the reuse of information and an unfolding interpretation of materials to construct new knowledge. We make our case through the presentation of findings from a naturalistic empirical observation of 11 humanities researchers engaging in research at a national archive. Our work identifies two research practices important to knowledge construction – reading and collecting – through which scholars create an interpretation of the archival record situated in its wider context.

Keywords

Archives
Human Information Interaction
Knowledge Construction
Information Use

PRESERVING ACCESS TO WEB SERVERS

Rafael Gieschke

University of Freiburg

Klaus Rechert

University of Freiburg

Susanne Mocken

University of Freiburg

Abstract

With the advent of the so-called Web 2.0 technologies, web resources became the primary platform for internal and external collaboration, knowledge exchange, and self-publication. Many of these services cannot be fully preserved by solely relying on client-side technologies like crawlers. In particular, their dynamic and reactive character will be lost by generating static snapshots. In this article, we develop concepts and describe infrastructure built on emulation to address the problem of retiring obsolete web service infrastructure while keeping the service long-term accessible and usable. We show the practicability of this approach by the example of three use-cases from academic projects.

Keywords

Web Server
Preservation
Emulation
Case Study

L ONG
   PAPER

3D Preservation Models and Modalities: Advancing Research Reproducibility and Capacity at Virginia Tech

Alex Kinnaman

Virginia Tech

Maureen Saverot

Virginia Tech

Yinlin Chen

Virginia Tech

Kara Long

Virginia Tech

Nicholas Polys

Virginia Tech

Sterling Nesbit

Virginia Tech

Michelle Stocker

Virginia Tech

Shuhai Xiao

Virginia Tech

Nathan Hall

Virginia Tech

LONG PAPER

Abstract

This paper presents findings from a multi-method study in archiving, accessing, and preserving 3D data. Specifically, we seek to develop a reusable and robust model for preserving and accessing 3D data. We further seek to identify links between digital triage, preservation actions, and archiving in multiple disciplines in order to make recommendations to embed data stewardship processes in research ecosystems. This work happens through interdisciplinary collaboration to validate curation practices across fields and disciplines.

Numerous disciplines and sectors are producing 3D data for research and instruction, but without the guidance of published standards or practices generalizable across fields. This paper will present our findings from identifying researcher needs, preservation and access requirements, and metadata models for our 3D models of specimens from the Virginia Tech Entomology Department. Our outcomes are an evaluation of the output using quantitative and qualitative methods, a working 3D metadata schema for access and preservation, and an access platform for our 3D models.

Keywords

Digital Preservation
3D Curation
3D Access
3D Metadata
3D Modeling

EXECUTABLE ARCHIVES: SOFTWARE INTEGRITY FOR DATA READABILITY AND VALIDATION OF ARCHIVED STUDIES

Natasa Milic-Frayling

Intact Digital Ltd

Marija Cubric

University of
Hertfordshire

Abstract

Long-term readability of electronic data is a key regulatory requirement for archived data integrity in life sciences and pharmaceutical research. However, this has been difficult to achieve within current data and software preservation practice due to data dependence on specialty software which becomes unusable as a result of rapid obsolescence of hardware and operating systems. This paper introduces a novel Executable Archive framework that extends traditional data archives with a platform for hosting legacy software and with processes for installing, use, and long-term maintenance of the software. Through a case study of a scientific software de-commissioning, we demonstrate the use of the framework for designing a solution for GLP-compliant software transition from operational to archival use and a secure processing of raw archived data to reconstruct past research studies. The framework is flexible and opens up opportunities for preservation planning and action that consider both data access and software management together, ensuring that the archived data integrity is fully supported by the long-term software integrity.

Keywords

Data Integrity
Software Integrity
Study Reconstruction
Significant Properties
Executable Archive

LONG
PAPER

When Digital Remembers Analogue - Conservation Metadata for Analogue Film as Preservation Description Information in a Digital Archive

Michelle Lindlar

TIB - German National Library of Science and Technology

Merle Friedrich

TIB - German National Library of Science and Technology

Miriam Reiche

TIB - German National Library of Science and Technology

Abstract

While "Digitization is not Digital Preservation", the two processes are often closely connected. This is especially true for audiovisual (AV), where the analogue materials, much like their digital counterparts, come in a variety of formats and are only accessible via rendering equipment such as projectors. This indirect accessibility of analogue audiovisual makes autopsy a resource-intensive process. Nevertheless, it's a necessary process prior to digitization, as the digital objects' quality depends on the quality of its analogue source material. But is analogue conservation data helpful in the preservation process of the digital object as well? Can it be used as Preservation Description Information in the OAIS sense?

TIB, the German National Library of Science and Technology, has authored a Conservation Metadata Schema for Analogue Film as part of the large digitization and digital preservation project DELFT. The schema includes metadata which TIB considers important knowledge for the contextualization and interpretation of the digital AV object. This paper outlines motivation and background for this development, discusses the connection of analogue and digital in an overarching analogue conservation and digital preservation workflow and gives a detailed insight into the TIB Conservation Metadata Schema.

Keywords

Audiovisual Preservation
Conservation Metadata
Preservation Description Information

L ONG PAPER

Progress with Improving Preservation and Reuse of Scientific Research Data

Peter Cornwell

Data Futures

Donat Agosti

Plazi

Laurence Benichou

Museum of Natural History, Paris

Jose Benito Gonzalez Lopez

CERN

Madeleine Herren

Institute for European Global History, Basel

Patrick Ruch

Swiss Instiutute of Bioinformatics

LONG **PAPER**

Abstract

We report progress towards automatically transforming existing analyses of scientific literature into annotations based on W3C's Web Annotation Data Model (WADM). Case studies are presented from the life sciences, and social sciences and humanities, in which these developments have led to the creation of new unrestricted data services for the research community. We discuss the cross- domain potential of annotation infrastructure for releasing scientific facts reported in research literature from copyright restrictions, and demonstrate the utility of common standards-based preservation and discovery methods in disparate activities. We suggest that scientific treatments of literature using WADM annotation can lead to new mechanisms for access to and reuse of research data, and accelerate convergence with the FAIR Principles.

Keywords

Research Data Preservation
Scientific Literature
Copyright, Biodiversity
Taxonomy
Infectious Disease
Global History

Collaborative Mechanism for Public Digital Preservation Service in China

Zhenxin Wu

National Science Library, CAS; University of CAS

Xiaolin Zhang

National Science Library, CAS; University of CAS

Jiancheng Zheng

National Science and Technology Library

Dongrong Zhang

National Science Library, CAS; University of CAS

Honghu Fu

National Science Library, CAS

Abstract

National Digital Preservation Program is built as a public service of, by, and for the national library community, and a trustworthy partner in the chain of digital scholarly communications. Its inherent collaboration is built on three layers: a national collaborative commitment by the library community and the government enforced by community auditing and certification, a multi-stakeholder contractual participation in the preservation of targeted resources, and a cooperative network of qualified archiving nodes from and answering to the communities, together with reliable partnerships with international publishers. By designing collaboration into the operation of the service, it has achieved a powerful, robust, trusted, and sustainable digital preservation service for the Chinese library community.

Keywords

Digital Preservation
Collaboration
Library
NDPP
China

LONG
PAPER

Dependence Rescission and Semantic Interpretation: Key Strategies to Long-Term Preservation of Digital Archives

Qian Yi

Renmin University

Ma Linqing

Renmin University

Abstract

There are three states of archival objects and their corresponding management measures. Clarifying the continuity or discreteness of signals and semantics of three states is very important in preservation. For objects at digital state, low-order logic dependencies with software and hardware dependencies as the core need to be removed. And for objects at data state, high-order logic dependencies with semantic associations as the core need to be removed.

Keywords

Analog State
Digital State
Data State
Dependency Removal
Semantic Interpretation

SHORT
PAPER

Cultural Heritage Preservation Using Multimedia and AI

Juan Du

Library, Southern University of Science and Technology

Xiaoqing Zhang

Department of Computer Science and Engineering, Southern University of Science and Technology

Kailun Jiang

Cultural Heritage Laboratory, Center for Social Sciences, Southern University of Science and Technology

Yufan Hu

Shuli College, Southern University of Science and Technology

Jigen Tang

Cultural Heritage Laboratory, Center for Social Sciences, Southern University of Science and Technology),

Jiang Liu

Department of Computer Science and Engineering, Southern University of Science and Technology

Henian E

Library, Southern University of Science and Technology

SHORT PAPER

Abstract

Cultural Heritage plays an important role in a modern city like Shenzhen. An archaeological discovery of Shenzhen 5000 years ago is innovatively preserved in library digital repository using digitalization and multimedia technologies at Southern University of Science and Technology(SUSTech). Multimedia modalities including interactive text, audio, image, graph, and video contents are developed, curated, and preserved in the project. Multimedia animation of a unique ritual culture from the archaeological discovery is further developed through deploying new computer graphics and Artificial Intelligence (AI) techniques. The project is materialized through inter-disciplinary collaboration among members from various teams with expertise in archaeology, multimedia, AI, and digital preservation. This collaboration enables the re-imaging of Shenzhen 5000 years ago innovatively with modern and technical elements.

Keywords

Interdisciplinary Collaboration
Digital Preservation
Digital Archaeology
Multimedia

Text File Format Identification: An Application of AI for the Curation of Digital Records

Santhilata Kuppili Venkata

The National Archives

Paul Young

The National Archives

Alex Green

The National Archives

Abstract

File format identification is a necessary step for the effective digital preservation of records. It allows appropriate actions to be taken for the curation and access of file types. The National Archives has existing processes for dealing with binary file format types, using tools such as PRONOM and DROID. These methods rely on using header information (metadata) and consistent binary sequences. However, these are not appropriate for the identification of text le formats as these do not contain recognisable header information or consistent patterns. Most text formats can be opened as plain text files, however file type information is often needed to understand the files use and context. Automated methods are necessary for text file format identification due to the scale of digital records processed by The National Archives, UK. An Artificial Intelligence methodology was tested and implemented using representative data collected from the GitHub repositories of UK Government departments. The first prototype developed has achieved reasonably good performance in successfully detecting five file formats with similar characteristics. The results encourage us to carry out additional experiments to include further text file format types.

Keywords

Text File Formats
Supervised Learning
Digital Preservation

SHORT
PAPER

BUILDING POLICY & CAPACITY: LAUNCHING THE U.S. LIBRARY OF CONGRESS DIGITAL COLLECTIONS MANAGEMENT COMPENDIUM

Trevor Owens

U.S. Library of Congress

Camille Salas

U.S. Library of Congress

Jesse Johnston

University of Michigan

Andrew Cassidy-Amstutz

U.S. Library of Congress

Pedro Gonzalez-Fernandez

U.S. Library of Congress

Kristy Darby

U.S. Library of Congress

SHORT PAPER

Abstract

In 2019 the Library of Congress launched the Digital Collections Management Compendium (DCMC). This publicly available resource establishes and communicates institutional digital collections management policy. A core function of the DCMC is to articulate the connections between broader digital collections management policy and that of planning resources produced and/or consulted by Library staff. This paper provides information regarding the development of the DCMC and its key features, and supports use of the DCMC as a public resource. We intend this paper to be of use to others working on organization-wide digital collections management initiatives. More broadly, the paper contributes to the discussion on how work around digital collections policy and planning connects with practice.

Keywords

Policy
Planning
Storage Systems

A Usability Study of Emulation

Jakob Moesgaard

Royal Danish Library

Claus Jensen

Royal Danish Library

Abstract

In this paper, we study the usability of a range of emulators using video games as an example digital material. We investigate four different emulators using four different games from a period spanning 1963 to 2011. Four students who each played all of the four games assisted us in the study. From their responses we identified a number of usability challenges. Finally, we discuss possible solutions to the usability challenges founded.

Keywords

Usability
Emulation
Access
Rendering
Interactivity

SHORT
PAPER

Harmonizing Human Infrastructure: A Case Study of Bringing Preservation Workflows of a Library, Archive, and Museum into Alignment

Hanna Bertoldi

University of Notre Dame

Peggy Griesinger

University of Notre Dame

Mikala Narlock

University of Notre Dame

SHORT PAPER

Abstract

The Hesburgh Libraries and Snite Museum of Art at the University of Notre Dame were awarded a 3-year grant that provided funding for a unified, cross-institutional discovery and exhibition platform that promotes serendipitous discovery of digitized cultural heritage materials. Although many digital preservation concerns were outside the scope of the grant, one beneficial output of the project has been renewed discussion and interest around robust digital preservation implementations appropriate for each institutions' specific needs. From derivative access copies to preservation TIFF images, the lightweight technical solution was designed to meet the discovery needs of the grant partners and also leave room to connect the project with future digital preservation systems and infrastructure. In this paper, we will discuss how our efforts to bring together different types of cultural heritage materials was also informed by an opportunity to consider digital preservation needs. We will describe the flexible, human-centered workflows that the team developed to prioritize education and collaboration, while leaving space for future preservation initiatives. This case study will provide concrete examples of how to bring workflows from disparate library, archive, and museum (LAM) units into harmony while being sensitive to both current, local practices as well as perceived future needs.

Keywords

Libraries Archives Museums
Digital Collections
Workflow

Understanding Storage Intermediaries

Helen Hockx-Yu

University of Notre Dame

Don Brower

University of Notre Dame

Abstract

Storage intermediaries are software, and sometimes hardware appliances that act as a link between applications and storage media, performing a range of tasks, such as protocol translation, caching, compression or even encryption. This paper describes storage intermediaries and their key functions that librarians and archivists should be aware of, as these introduce technical dependencies that can impact digital preservation.

Keywords

Archival Storage
Storage Gateway
Cloud Storage
Tape Storage
Digital Preservation
Redundancy
Diversification

SHORT
PAPER

A Collaborative Approach to Preserving At-Risk Open Access Journals: "Journals Preserved Forever"

Alicia Wise

....

CLOCKSS

Abstract

Scholars rely upon ongoing access to the published research in their fields. For journals to disappear is to harm the ability of scholars to conduct their work. Recent research has indicated that journals published by the long tail of smaller publishers may be especially at risk for disappearing from the Web. Five leading organizations are collaborating to significantly improve the number of Open Access journals that are being preserved for the long-term, in a project named "Journals Preserved Forever" (nicknamed "JASPER").

Keywords

Digital Preservation
Open Access
Journals

SHORT
PAPER

THE TIP OF THE VALIDATION ICEBERG

Jay Gattuso

National Library of New Zealand

Andrea Goethals

National Library of New Zealand

Abstract

This paper describes the volume of JHOVE-based file validation warnings in the National Library of New Zealand's digital preservation repository, our motivation for addressing them, and the start of our automated workflow that seeks to significantly reduce the number of indicated files.

Keywords

File Format
JHOVE
Validation
Preservation Action

SHORT
PAPER

FILEDRILLER: MARRYING SIEGFRIED AND THE NATIONAL SOFTWARE REFERENCE LIBRARY

Steffen Fritz

German Literature Archive

Abstract

Working with huge collections of unstructured data is a common yet still challenging task in digital preservation. This paper presents a tool for finding irrelevant files in large data sets to spot the relevant. The tool builds on two well-known and frequently used applications, respectively data sets and combines and extents them in a meaningful way.

Keywords

Format Identification
Tool
Automation
Siegfried
NSRL

SHORT
PAPER

Back to Basics: The Minimum Preservation Tool

Maureen Pennock

British Library

John Beaman

British Library

Peter May

British Library

Kevin Davies

British Library

Abstract

This paper presents the Minimum Preservation Tool (MPT), designed and developed by the British Library to provide a basic technical digital preservation service for collections awaiting ingest to a more formal digital preservation repository. The MPT can satisfy fundamental preservation storage requirements that are not typically otherwise supported in a standard corporate technical environment more focused on cyber-security. Replication, checksum generation and validation, and regular reporting are all key features of the MPT, written as a set of Python Utilities and freely available on Github. MPT is an entry-level tool that lowers the bar for early participation in preservation endeavors, in contrast with larger scale and more expensive, complex end-to-end technical solutions.

Keywords

Minimum Preservation Tool
Checksums
Integrity
Replication
Accessibility
Open Source
File Preservation
Risk Reduction
Assurance

SHORT PAPER

FDA-DBRepo: A Data Preservation Repository Supporting FAIR Principles, Data Versioning and Reproducible Queries

Martin Weise

TU Wien

Cornelia Michlits

U.S. Library of Congress

Moritz Staudinger

TU Wien

Eva Gergely

Universität Wien

Kirill Stytsenko

Universität Wien

Raman Ganguly

Universität Wien

Andreas Rauber

TU Wien

SHORT PAPER

Abstract

Database preservation frequently happens post-factum: databases are transferred and migrated into preservation formats and environments after a project has ended.

This increases the risks concerning incompatibility and pushes the preservation burden after the initial lifetime and use of the data.

We propose a database repository infrastructure, where databases are created, used and preserved directly in the data curation environment. This increases the FAIRness of the data curated as professional data stewardship activities accompany the databases right from the onset. We present the FAIR Data Austria Database Repository (FDA-DBRepo) infrastructure and provide a first version of an open-source reference implementation.

Keywords

Database Preservation
Research Data
FAIR Data
Data Citation
Reproducibility

THE DESIGN OF COLLABORATIVE MARKING MODE FOR MUSEUM COLLECTIONS-TAKING "GUGONG MINGHUAJI" AS AN EXAMPLE

Sun Jing

The Palace Museum

Abstract

Collaborative tagging is a way of information collection after the popularity of the Internet. Digital resource preservation and management organizations are very interested in this way, and have done a lot of research. However, this method is affected by the uneven level of information providers, the difficulty in confirming the accuracy of information, and the risk of unverified information exposure on the platform. In this paper, we will design a collaborative tagging model, which uses the content of user's spontaneous tagging as the basis, through the three stages of self-use tag, collective tag audit and public tag, to supplement the collection information and improve the efficiency of digital resource retrieval.

Keywords

Collaborative Tagging
Tag
Information Organization
User Generated Content

SHORT PAPER

A Smart Guide to Preferred Formats

Sam Alloing

KB, National Library of the Netherlands

Remco de Boer

ArchiXL

Marjolein Steeman

Netherlands Institute for Sound and Vision

Abstract

The Dutch Digital Heritage Network has developed an online tool that helps archives and heritage organisations to formulate their policy on file formats. It builds a knowledge base, via a smart combination of codification of formats on one hand and personalisation of tailor made policies on the other. This paper will explain how the online tool is set up, what technologies it uses and what possible next steps might be in the near future.

Keywords

Preferred Formats
Community
Knowledge Base
Linked Data
Registers

SHORT
PAPER

Developing a Holistic Research Data Management Strategy for a University - Making Preservation Planning and Long Term Access First Grade Citizens in RDM

Dirk Von Suchodoletz

University of Freiburg

Jan Leendertse

University of Freiburg

Klaus Rechert

University of Freiburg

Rafael Gieschke

University of Freiburg

Saher Semaan

University of Freiburg

Björn Goldammer

University of Freiburg

Dimitri Tolkatsch

University of Freiburg

S HORT PAPER

Abstract

An integrated research data management (RDM) enables reproducible and verifiable research, linking of interdisciplinary expertise, sharing of research for comparison, and integration of different analysis results and metadata studies. An increasing adoption of FAIR principles and requirements by funding agencies has significantly benefited overall quality, reuse, and sharing of research results. For scientists, a proper data management is a crucial element to prove their findings and make them reproducible. As a consequence, RDM had to become an integral part of the science support infrastructure in todays research institutions.

Scientist of various disciplines should be supported over the data lifecycle starting from holistic planning of future projects to RDM related services provided: As RDM is a multidimensional endeavor requiring various skills; tasks ranging from community specific to community needs are optimally handed to the best qualified provider. The presented concepts and considerations are work in progress while establishing an organizational frameworks for a research university. Completely reproducible preferably open data publications including the relevant data-sets' context are the ultimate goal. These require appropriate service components like EaaSI. The university strives to profit from the overlaps in RDM and digital preservation and to define the handover of tasks from the first to the latter.

Keywords

Research Data Management Planning
Continuous Access
Re-Use
Data Management Plan
Data Publication
Federated Services
RDM Ecosphere

DNA4DNA: Preserving Culturally Significant Digital Data with Synthetic DNA

Eugenio Marinelli

EURECOM

Eddy Ghabach

EURECOM

Thomas Bolbroe

Rigsarkivet

Omer Sella

Imperial College London

Thomas Heinis

Imperial College London

Raja Appuswamy

EURECOM

Abstract

The growing adoption of AI and data analytics in various sectors has resulted in digital preservation emerging as a cross-sectoral problem that affects everyone from data-driven enterprises to memory institutions alike. As all contemporary storage media suffer from fundamental density and durability limitations, researchers have started investigating new media that can offer high-density, long-term preservation of digital data. In the European Union-funded Future and Emerging Technologies project OligoArchive, we are exploring one such media, namely, synthetic Deoxyribo Nucleic Acid (DNA). In this paper, we provide an overview of the ongoing collaboration between project OligoArchive and the Danish National Archive in preserving culturally important digital data with synthetic DNA.

Keywords

DNA Storage
Long-Term Storage
Preservation
SIARD-DK

SHORT
PAPER

GROWTH RINGS AND DIGITAL PRESERVATION THREE MEDITATIONS ON STRATEGIES TO PRESERVE PERFORMANCE-DEPENDENT ITEMS

Louise Curham

School of Information Studies and Communication, Charles Sturt University

Louise Curham

University of Canberra

Abstract

This short paper consists of three meditations on potentially fruitful concepts for the preservation of performance-dependent heritage derived from a study of the use of 1970s media art. The meditations cover the following topics: 1) there are aspects to ritual that can guide principles for digital preservation; 2) as a result of use, growth rings that reflect that use can be added to items as a form of enriched metadata; and 3) in literary translations, 'clouds of meaning' can be shared with readers by showing both the original and the translation [1]. Some future users will want to understand not just the content of item that has undergone migration or emulation, they will want to investigate that process itself, able to interrogate both the original item and its iterations.

Keywords

Media Art
Audiovisual Preservation
Participatory Archiving
Digital Preservation
Metadata

SHORT PAPER

SOFTWARE PRESERVATION IN THE NETHERLANDS: LOWERING THE THRESHOLD FOR CULTURAL HERITAGE INSTITUTIONS

Jesse de Vos

The Netherlands Institute for Sound and Vision

Claudia Röck

The Netherlands Institute for Sound and Vision

Eoin O'Donohoe

The Netherlands Institute for Sound and Vision

Abstract

Thus far, software preservation in Dutch Heritage institutions is far from common practice. Over the course of 18 months the collaborative Software Archiving-project aimed to lower the threshold for institutions to start or advance software preservation. The project investigated the status quo, made recommendations and offered practical resources on emulation as a service, browser emulation and metadata for software preservation.

Keywords

Emulation
Software
Web Browsers
Metadata

SHORT
PAPER

Towards Levels of Digital Preservation as a Service

Jefferson Bailey

Internet Archive

Peggy Lee

Internet Archive

Abstract

Since 1996, the Internet Archive (IA) has provided storage, preservation, and access infrastructure and services to over 1,000 cultural heritage organizations around the world. It has also provided customized digital preservation services on a contractual basis to a handful of large institutions. In 2020, IA began building a more generalized digital preservation service in response to the needs of a broader range of institutions and to leverage IA's self-owned data centers, non-profit cloud services, and demonstrated expertise in both small and petabyte-scale digital stewardship. This system is being developed in direct dialogue with 30+ organizations, including universities, public libraries, arts organizations, and cultural heritage organizations, over the course of the 2021 - 2022 year. This paper shares key takeaways from the information collected from this pre-pilot phase and serves as a lean landscape review of the current gaps within the digital preservation landscape, particularly as they relate to the distinct needs and goals of nonprofits, libraries, and cultural heritage organizations, that this service aims to address.

Keywords

Digital Preservation
Product Development
Archives
Open Infrastructure
Sustainability

SHORT
PAPER

The Italian Guidelines on Creation, Management and Preservation of Digital Records—A Proposed Methodology for File formats Assessment

Stefano Allegrezza

University of Bologna

Abstract

This paper aims to address the issue of file formats assessment for the preservation of digital records, which is fundamental because the chances of preserving records produced in these formats over time depend greatly on it. In particular, the paper presents and discusses the methodology proposed by the Italian Agency for Digital Government who recently published the "Guidelines on the Creation, Management and Preservation of Digital Records". This methodology is based on a quantitative method that evaluates some properties of file formats and assigns them a score; the sum of these scores is the so-called "interoperability index" that provides useful information in order to establish whether the format is prone to obsolescence or not. The methodology is explained with examples that show its potential. Some suggestions for improvement and further developments are then discussed.

Keywords

File Formats
Assessment
Evaluation
Selection
Digital Preservation

SHORT
PAPER

Project Pipeline: Preservation, Persistence, and Performance

Jane Greenberg

Drexel University

Christopher B. Rauch

Drexel University

Mat Kelly

Drexel University

Abstract

Preservation pipelines demonstrate extended value when digitized content is also computation-ready. Expanding this to historical controlled vocabularies requires additional steps if they are to be fully leveraged for research. This paper reports on work addressing this challenge. We report on a pipeline and project progress addressing three key goals: 1) transforming the 1910 Library of Congress Subject Headings (LCSH) to the Simple Knowledge Organization System (SKOS) linked data standard, 2) implementing persistent identifiers (PIDs) and launching our prototype ARK resolver, and 3) importing the 1910 LCSH into the Helping Interdisciplinary Vocabulary Engineering (HIVE) System to support automatic metadata generation and scholarly analysis of the historical record. The discussion considers the implications of our work in the broader context of preservation, and the conclusion summarizes our work and identifies next steps.

Keywords

Computational Archival Science
Historical Vocabularies
Digital Preservation
Persistence
Pipelines

SHORT PAPER

Trans-Pe/ar(man)ent Engage Virtualization as a Preservation Strategy to Create an Open, Sustainable, Collaborative Infrastructure in an Extensive Art Museum

Almut Schilling

aBITpreservation

Franziska Butze-Rios Butze-Rios

Donau-Universität Krems Austria

Abstract

This paper briefly describes the establishment of a digital archival infrastructure to preserve one artwork from the late 90's and a complete artist archive. Within those digital preservation processes an accessible, transparent and sustainable infrastructure will be designed, implemented and evaluated, fitting the existing collection management structure. The most critical point is building a system relying on open source software in an institutional environment without dedicated knowledge and experience. While building, testing and integrating the Linux working station to virtualize and transform those artworks it is confirmed that this is an efficient method to built transparent institutional capacities and capabilities.

Keywords

Software-Based Artwork
Digital Sustainability
Virtualization
Open Source
Cooperation

SHORT
PAPER

ARE WE WINNING? OTHER MEASURABLES FOR DIGITAL PRESERVATION

Tim Evans

....

Archaeology Data
Service

Abstract

This short paper reflects on the past 25 years of the Archaeology Data Service (ADS), a digital repository for UK heritage data. The paper focusses on the benefits of the Digital Preservation Coalition (DPC) Rapid Assessment Model (RAM). Experience of using the RAM has complimented the strengths of accreditation and has done much to focus on capacity and capability. It has also helped highlight successes that may otherwise be overlooked. For a smaller organization, these successes may not be intrinsically measurable in terms of bytes and processes, or the capacity to build a complex infrastructure, but simply demonstrating the reuse of the data so forensically preserved and the individuals that ensure that it matters.

Keywords

Maturity Modelling
Designated Community
Data Reuse
Capability

SHORT PAPER

DIGITAL PRESERVATION STORAGE CRITERIA AND RELEVANT STANDARDS - LATEST DEVELOPMENT ON THE DIGITAL PRESERVATION STORAGE CRITERIA

Eld Zierau

Royal Danish Library

Cynthia Wu

National Library of New Zealand

Andrea Goethals

National Library of New Zealand

Nancy McGovern

Massachusetts Institute of Technology

Sibyl Schaefer

University of California

Abstract

This poster provides the latest updates on the Digital Preservation Storage Criteria. These Criteria are intended to help with developing requirements and evaluations of preservation storage solutions, to seed discussions about preservation storage, and/or to use as digital preservation instructional material. The Criteria are grouped into categories including content integrity, cost considerations, flexibility, information security, resilience, scalability & performance, support, and transparency. The working group created an accompanying usage guide covering risk management, independence, elements in establishing bit safety, and cost considerations as related to the Criteria, with new topics to be added in the next version. This year, the working group mapped each criterion to international standards that are relevant to Digital Preservation Storage such as ISO 16363 Audit and Certification of Trustworthy Digital Repositories and the ISO 27000 Information Security series. The aim of this poster is to discuss the mapping that was done between the criteria and these standards, as well as the proposed changes to the criteria and the associated usage guide as a result.

Keywords

Digital Preservation Storage
Archival Storage
Criteria
Risk Management
Standards

POSTER

Introducing the Format Library Helper

Jay Gattuso

National Library of New Zealand

Abstract

This poster describes the Format Library helper tool built by the National Library of New Zealand to aid file format research and analysis.

Keywords

Format Identification
PRONOM
Format Library
Knowledge Base
Risk

POSTER

48

IS THIS IMAGE THAT IMAGE?

Jay Gattuso

National Library of New Zealand

Abstract

This poster describes the image comparison technique developed at the National Library of New Zealand. This process has been developed to help demonstrate / document any changes to image content as a result of a preservation or conservation treatment.

Keywords

Integrity
Image Comparison
Rmse

POSTER

LaVaH - Distributed Long-Term Preservation with Diverse Roles and Responsibilities

Natascha Schumann

hebis Verbundzentrale, Goethe University Frankfurt

Martina Sinkovic

hebis Verbundzentrale, Goethe University Frankfurt

Abstract

Distributed long-term preservation with diverse roles and responsibilities is the goal of the LaVaH project. The poster describes the division of work between stakeholders from the university libraries, the data centers, and the service provider for library infrastructure. Data management is done locally at the universities, central digital archiving tasks by the service provider (hebis) and storage infrastructure is hosted by two data centers.

Keywords

Collaboration
Data Management
Distributed Preservation

POSTER

Digital Lifecycle Management in Maryland: Establishing the Foundations for Consortial Approaches to Digital Preservation at Maryland Universities and Affiliated Institutions

Joseph Koivisto

University of Maryland

Laura Bell

University of Baltimore

Kristin Conlin

University of Baltimore

Allison Fischbach

Towson University

Fatemeh Rezaei

University of Baltimore

Ashley Todd-Diaz

Towson University

Abstract

The Digital Lifecycle Management Interest Group (DLMIG) of the University System of Maryland and Associated Institutions (USMAI) formed in 2020 and provides a space for multi-institutional collaboration. The group is working toward a USMAI-wide digital lifecycle management strategy that will provide sustainable and scalable digital preservation and practices at consortium institutions of all sizes. This poster will share the group's experiences and challenges thus far as a model for other consortia and collaborations.

Keywords

Digital Preservation
Consortium
Collaboration

POSTER

A Data-Visiting Infrastructure for Providing Access to Preserved Databases that Cannot be Shared or Made Publicly Accessible

Martin Weise

TU Wien

Andreas Rauber

TU Wien

Abstract

Databases preserved in archives contain highly valuable information that frequently cannot be made freely accessible for analyses via standard data portals, be it due to legal, commercial or ethical issues. We present OSSDIP, the reference implementation of a high-security data visiting infrastructure initially conceived as a safe-compute environment for medical data. It provides highly controlled and monitored data visiting services while ensuring to the largest degree possible that data cannot be extracted from the infrastructure. This may offer archives a viable alternative for providing restricted access to sensitive data in a more flexible manner.

Keywords

Secure Data Infrastructure
Data Visiting
Data Dissemination

POSTER

PHS-PRES: MULTI-SIDED PROTECTION FOR POPULATION HEALTH DATA

Chenliu Yang

Institute of Medical Information, Chinese Academy of Medical Sciences

Jiahui Hu

Institute of Medical Information, Chinese Academy of Medical Sciences

Yunman Fan

Institute of Medical Information, Chinese Academy of Medical Sciences

An Fang

Institute of Medical Information, Chinese Academy of Medical Sciences

Abstract

This poster presents the challenges of population health data long-term protection, put forward the security system and multi-sided approaches to defend these threats.

Keywords

PHS-PRES
Long-Term Preservation
Multi-Sided Protection Security Strategy
Safeguard Procedures

POSTER

ENHANCING FRENCH DIGITAL PRESERVATION COOPERATION

Yannick Grandcolas

BIBLIOTHEQUE NATIONALE DE FRANCE

Lorène Bechard

CINES

Marion Humbert

Archives de la Moselle

Martine Sin Blima-Barru

Archives Nationales

Emeline Levasseur

Archives Nationales

POSTER

Abstract

Since 2019, four groups have been created within the French association "Aristote", including members of National Library & Archives and several Ministries.

These four groups are working on these subjects:
1. Digital Formats knowledge: in all our institutions we must deal with many different formats, it was necessary to share our experience, help each other, and propose tools to help the French digital preservation community to think about the development of format preservation strategies.
2. Translation: because most of the documentation about digital preservation is not available in French.
3. Expertise: it's necessary to identify and share our experts' knowledge on file formats between institutions because of the lack of human resources in the digital preservation domain.
4. Tools: many tools are available, and we can offer a national guidance through our experimentations.

In conclusion, the creation of these groups has set up new ways to improve our National digital preservation strategy.
The poster is made with BDNF (https://bdnf.bnf.fr/), a free application from the French national Library, it was created to use digital objects and images available in Gallica, the French digital national library online (https://gallica.bnf.fr).

Keywords

Digital
Preservation
Cooperation
Formats
Expert
Translation
Tools
Francophonie
France

CHANGING PRACTICES IN FIXITY: RESULTS FROM THE 2021 NDSA FIXITY SURVEY

Nick Krabbenhoeft

New York Public Library

Carol Kussmann

University of Minnesota

Sibyl Schaefer

University of California San Diego

Abstract

Beyond the encouragement to check fixity, the field of digital preservation does not have a best practice on how to verify the stability of digital collections. Fixity verification is a contextual practice that is implemented differently at every organization. This poster presents the results of the 2021 NDSA Fixity Survey to better understand fixity practices at cultural heritage organizations.

Keywords

Fixity
Survey
Workflow

POSTER

LONG-TERM PRESERVATION AND UTILIZATION OF DIGITAL DUNHUANG RESOURCES

Peter Zhou

University of California

Linqiang Luo

Seagate Technology

Jian Wu

Dunhuang Academy

Kendell Chilton

Futurewei

Tianxiu Yu

Dunhuang Academy

Abstract

Dunhuang Mogao Grottoes, a world cultural heritage originally built in the 4th century, has now 735 remaining grottoes. Since the 1990s, Dunhuang Academy has taken the lead in digitizing the architectural forms, statues and murals of the grottoes through non-contact data collection and optical measurement technology, so that Dunhuang resources can be fully protected digitally. Based on that, through "digital Dunhuang", numerous images, videos, three-dimensional digital objects, research documents and archaeological information are collected to build a digital, integrated, large-scale mural resource library and a sharing guarantee system for Dunhuang digital resources, to further link with relevant international resources and support knowledge mining and intelligent services.

Keywords

Digital Dunhuang Resources
Long-Term Preservation
Utilization

P ANEL

BUILDING A LOCAL DIGITAL PRESERVATION INFRASTRUCTURE:
EXPERIENCES IN SELECTING AND IMPLEMENTING DIGITAL PRESERVATION SYSTEMS

Oya Y. Rieger

Ithaka S+R

Michael Boock

Oregon State University

Lindsay McCormack

University of Oxford

Nick Ruest

York University

Abstract

As cultural heritage organizations seek solutions, open source, community-based, and commercial digital preservation systems and tools have proliferated. Nevertheless, given the magnitude of digital content and the requirements of specific file formats, institutions struggle to integrate disparate tools and systems to support their institutional preservation programs. The purpose of the panel discussion is to provide insights from on-the-ground implementations of current preservation tools and to consider the opportunities and challenges in deploying and integrating these tools with local digital content management processes.

Keywords

Preservation Systems
Digital Asset Management
Sustainability
Preservation Processes

PANEL

DIGITAL DOCUMENTARY HERITAGE: SECURING RELIABLE AND SUSTAINABLE ACCESS

Lai Tee Phang

UNESCO Preservation Sub-Committee

Natasa Milic-Frayling

UNESCO MoW Preservation Sub-Committee

Peter Doorn

Data Archiving and Networked Services

Tobias Steinke

German National Library

Ryder Kouba

American Center of Research

Eric Chin

National Library Board, Singapore

Abstract

UNESCO Memory of the World Programme (MoW) and its Preservation Sub-Committee are concerned with the long term and sustained access to the World's documentary heritage, including digital. They are keen to engage with the digital preservation professionals around the world, to gather the requirements for effective international policies to support digital preservation actions and, more importantly, work towards reliable access to the world's documentary heritage, now and in the far future. The past two decades have seen unprecedented increase in publishing, documenting, and communicating in the digital form, giving rise to a wealth of knowledge, insights and experiences. Digital information is becoming critical for education, skill building and participation in the global digital economy. This panel is organized by the UNESCO Preservation Sub-Committee to discuss the critical technological, policy and sustainability factors in providing effective access to the World's digital documentary heritage.

Keywords

Digital Preservation
Digital Documentary Heritage
Access
Technology
Digital Preservation Policy

P ANEL

Digital Library Maturation: The Evolution of Digital Collections Management Programs at National Libraries

Trevor Owens

Library of Congress

Andrea Goethals

National Library of New Zealand

Jeffrey van der Hoeven

National Library of the Netherlands

Lisandro Pablo Olivares

Biblioteca Nacional de México

Paul Wheatley

Digital Preservation Coalition

Michael Day

The British Library

Abstract

This panel will present and discuss the development of digital collections management programs at a series of national libraries. Panelists will each give brief overviews of the development of their institutions programs, policies, and technical infrastructure and share out about upcoming and ongoing major initiatives in these areas.

Keywords

Digital Preservation
National Libraries
Policy
Infrastructure
Formats
Scale
Access

P ANEL

EMAIL ARCHIVING: A COLLABORATIVE JOURNEY

Ruby Martinez

University of Illinois

Chris Prom

University of Illinois

Stephen Abrams

Harvard University

Gregory Wiedeman

University of Albany, SUNY

Matt Teichman

University of Chicago

Abstract

This panel examines the current and future developments in email archiving. Such developments include 1) the Email Archiving: Building Capacity and Community (a regrant program), 2) Integrating Preservation Functionality into ePADD, 3) Attachment Converter: Preserving the Context of Electronic Correspondence, 4) Mailbag: A Stable Package for Email with Multiple Formats, and 5) Developments in EA-PDF. In addition, the five panelists will consider how their projects' collaborative nature contributes to the broader community's efforts to build flexible, efficient services to meet email archiving challenges.

Keywords

Email Archiving
Best Practice
Community
Collaboration

PANEL

Trustworthiness Auditing for Sustainable Digital Preservation Services

Xiaolin Zhang

National Science Library, Chinese Academy of Sciences

Zhenxin Wu

National Science Library, Chinese Academy of Sciences

Yunhai Tong

Peking University

Abstract

Digital preservation (DP) service can be provided at various levels and by different agents. But the very nature of DP requires that the service be provided in a reliable, resilient, trustworthy, and sustainable way. This further demands a systematic thinking covering the contextual, legal, organizational, technical, administrative, financial, and personnel aspects of the services, and an evidence-based operations approach ensuring the requirements for trustworthiness is understandable, implementable, measurable, verifiable, retraceable, accountable, transparent, and efficiently executable. The auditing is a tool and an opportunity, but the resulting trustworthy service is the starting point and the desired result. The workshop will summarize international understanding, standards, and best practices, but will mainly use the experience of NDPP to illustrate how a real life digital preservation service develops itself in response to the requirements of trustworthiness auditing in regard of those principles. The audience will be asked to discuss their requirements and efforts or plans from their individual needs for digital preservation services.

Keywords

Trustworthiness Auditing
Digital Preservation Services
Auditing & Certification
Contextual & Organizational & Technical & Administrative Requirements

WORKSHOP

GUIDELINES FOR THE SELECTION OF DIGITAL HERITAGE FOR LONG-TERM PRESERVATION: WHAT, WHY, WHEN

Ingrid Parent

University of British Columbia; IFLA

Anthea Seles

ICA

Jenna Murdock Smith

Library and Archives Canada

Claire McGuire

IFLA

Abstract

The PERSIST Content Task Force, under the Preservation Subcommittee of UNESCO Memory of the World, offers this workshop based on the 2nd edition of the UNESCO Guidelines for the Selection of Digital Heritage for Long-term Preservation. This ninety-minute workshop is geared to an audience of practitioners of digital preservation. It will be informal, allowing for active exchange among the workshop participants and leaders. The content of this workshop will include an overview of the UNESCO PERSIST programme, a look at how the Guidelines have been used in the past, an introduction to the updated edition of the Guidelines and key concepts, and an interactive discussion on how they can be applied in professional practice. This workshop will bring the publication of the 2nd edition of the Guidelines to the attention of practitioners working to preserve digital information and provide an opportunity for a lively discussion on practical applications in a range of memory institutions.

Keywords

Content Selection
Digital Preservation
Libraries
Archives
Museums

WORKSHOP

From Capture to Replay: Web Archiving with Webrecorder Tools

Lorena Ramirez-Lopez

webrecorder

Ilya Kreymer

webrecorder

Emma Dickson

webrecorder

Abstract

Web archiving can take many forms, but usually involves fully capturing websites, storing the captured content, and faithly reproducing, or replaying the archived sites.

Through this workshop, participants will gain a working knowledge on how web archiving works by using a simple web archiving process from capture to replay using two open-source tools from Webrecorder: archiveweb.page and replayweb.page. Archiveweb.page is an open-source capture tool that turns your browser into a full-featured interactive web archiving system. Replayweb.page is an open-source browser-based viewer that loads and renders web archive files for replay in the browser. Both these tools combined create high-fidelity web archives using a basic workflow from capture to replay that participants can control, save, and embed onto other platforms. All the code can be viewed and accessed via github: http://github.com/webrecorder/ The webrecorder project builds quality open source tools to enable 'web archiving for all' to allow anyone with a browser to create their own web archives, and to accurately replay them at a later time.

Keywords

Web Archiving, Capture, Replay, Open-Source Tools, Web Collecting, Personal Digital Archiving, High Fidelity Web Archives, Managing Web Archives Skills Building, Hands-on Workshop

W ORKSHOP

Understanding and Implementing METS

Karin Bredenberg

Kommunalförbundet Sydarkivera

Juha Lehtonen

CSC – IT Center for Science

Tobias Steinke

Deutsche Nationalbibliothek

Abstract

This half day tutorial will provide participants with an introduction to the Metadata Encoding and Transmission Standard (METS) and the METS Primer. It will give a basic overview of the standard and explore different models of implementation. The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library as well as digital archives, expressed using the XML schema language of the World Wide Web Consortium. It is maintained by the METS Board and the METS Maintenance Activity is managed by the Library of Congress.

Keywords

Metadata and Information Strategies and Workflows
Infrastructure and Systems
Tools
Case Studies
Best Practices and Novel Challenges
Training and Education

TUTORIAL

RIGHTS MANAGEMENT FOR RESEARCH DATA MANAGEMENT, SHARING, AND PRESERVATION

Zeyu Zhang

Computing and Networking Information Center, Chinese Academy of Sciences

Lulu Jiang

Computing and Networking Information Center, Chinese Academy of Sciences

Zhihong Shen

Computing and Networking Information Center, Chinese Academy of Sciences

Abstract

Research data (RD) is a fast developing area of digital resources that requires long-term preservation. But the acquisition, processing, share, and preservation of RD involve complicated rights issues among multi-stakeholders with different or conflicting interests. Only with well-designed, easy-to-implement, and balanced rights management, we can achieve an effective, reliable, robust, sustainable research data management and preservation. This tutorial aims to summarize issues and best practices in this regard to stimulate further discussion on it.

Keywords

Research Data
Digital Research
Management & Sharing & Preservation & Publishing
Rights Management
Data Protection and Data Security

T UTORIAL

Understanding and Implementing PREMIS

Karin Bredenberg

Kommunalförbundet Sydarkivera

Eld Zierau

The Royal Danish Library

Michelle Lindlar

TIB Leibniz Information Centre for Science and Technology

Abstract

This half day tutorial will provide participants with an introduction and walkthrough of the Data Dictionary and the Semantic Units found in PREservation Metadata Implementation Strategies (PREMIS) after the sessions start with an introduction to the background and resources available. Thereafter the focus is implementation which will explore the different ways that PREMIS can be implemented, there are more than one way to implement PREMIS. The tutorial will focus on a basic overview of the standard and explore different models of implementation giving you the skills to continue your PREMIS journey.

Keywords

Preservation Strategies and Workflows
Systems and Tools
Case Studies
Best Practices and Novel Challenges
Training and Education

T UTORIAL

AI VIDEO GAME LEGACY ASSETS CLASSIFIER

Stefan Serbicki

Electronic Arts.

Abstract

Our department collects vast chunks of assorted data from legacy servers left over after teams disband and/or studios close. In most cases the domain experts that created that data and knew how to use it moved on resulting in a loss of the knowledge needed to make sense of it. Our solution to this issue was to train an AI model that could help identify and classify this data and subsequently feed the output to our catalog application.

Our demonstration would show how our model was trained, how it carries out predictions, and how well it performs on those predictions. To illustrate performance we will present confusion matrices, F1 scores, and precision-recall graphs. We will also outline future plans regarding model improvements and uses.

Keywords

Machine Learning
Data Dumps
Legacy
Classification
Artificial Intelligence
Video Games

DEMONST RATION

COPTR-EDIT-A-THON TOOLING UP FOR DIGITAL PRESERVATION

Stacey Jones Erdman

Arizona State University

Ania Molenda

Dutch Digital Preservation Network

Paul Wheatley

Digital Preservation Coalition

Abstract

Community Owned digital Preservation Tool Registry (COPTR) was developed in 2013 to provide efficient "one stop shopping" for practitioners performing research on digital preservation tools and services. It contains information on over 500 tools and since 2021 has been equipped with semantic MediaWiki capabilities. This makes a large part of the information within the registry machine-readable and possible to query by API. It opens new possibilities for creating tool overviews and allowing users to find comparisons that will be useful to them. It also makes new integrations possible for projects, which are based on data harvested from COPTR. COPTR Edit-A-Thon during iPRES2021 invites the international digital preservation community to come together in a common effort to enrich the data in COPTR, explore the possibilities of the recently added semantic annotation, and discuss the future of the registry. It will combine hands-on editing, group discussions and brainstorms conducted in a hybrid manner. With this edit-a-thon we aim to develop new/improved entries, new ideas for making the most of the new functionality, and new items to add to the "to do" list for future improvements. This event is for everyone interested in tools and workflows for digital preservation and does not require any specialized knowledge.

Keywords

preservation tools
open source tools
preservation solutions
collaborations

Virtual Research Environment with Preservation Tools

Ania Molenda

DDHN, Netherlands

Martin Speller

OPF, UK

Carl Wilson

OPF, UK

Abstract

In this lightning talk Ania Molenda from the Dutch Digital Heritage Network (DDHN), Martin Speller and Carl Wilson from Open Preservation Foundation (OPF) will present the Virtual Research Environment (VRE) with preservation tools, which they recently developed.

L IGHTING TALK

TO PRESERVE, YOU MUST ADAPT.

Alex Kinnaman

Virginia Tech University
Libraries

Rachel Howard

University of Louisville
Libraries

Deanna Ulvestad

Greene County Public
Library

Abstract

The MetaArchive Cooperative has made proactive adjustments to its community model, technical interoperability, and the ways it engages with the broader distributed digital preservation field in an intentional effort to remain relevant, useful, and impactful. As a multi-cache preservation system, we rely on our membership to sustain our preservation infrastructure. As the digital preservation landscape evolves, the Cooperative also evolves to meet the needs of our diverse membership. The Cooperative will share the impact of its Changing for Continued Impact series, which has resulted in newly established Membership Levels, allowing a broader range of organizations and individuals to participate in the Cooperative. Increasing membership eligibility increases the engagement and involvement from a wider community of content creators, users, and decision-makers in the digital preservation community. This presentation will give the digital preservation community a look at the Cooperative's new and future directions.

LIGHTING **TALK**

ARCHIVING AND PRESERVATION OF LARGE SCIENTIFIC DATASETS

Matthew Addis

Arkivum Ltd

Abstract

In this lightning talk, we would like to present recent developments by Arkivum in the ARCHIVER project where we have been working on new approaches to very large-scale digital preservation of scientific datasets. The European Commission supported ARCHIVER project (Archiving and Preservation for Research Environments) aims to "introduce significant improvements in the area of archiving and digital preservation services, supporting the IT requirements of European scientists and providing end-to-end archival and preservation services, cost-effective for data generated in the petabyte range with high, sustained ingest rates, in the context of scientific research projects". Arkivum, along with other contractors in the ARCHIVER project, have been building prototype LTDP solutions in Phase 2 of the project. Initial results of a working prototype for ARCHIVER are already available. Further results of Phase 2 will be ready in time for presentation at iPRES. The focus of the talk will be to highlight how new cloud computing models such as serverless computing allows for LTDP systems to be constructed and operated in a way that is economically and environmentally sustainable as well as to preserve and provide access to very large datasets – not just for scientific data, but for all forms of digital content.

LIGHTING TALK

The "PUPS" Model: Private funding, University staff, Publicrecords, and State archives

Gregory S. Hunter

Palmer School of Library and Information Science, LIU Post

Jaime Karbowiak

Palmer School of Library and Information Science, LIU Post

Emily Antoville

Palmer School of Library and Information Science, LIU Post

Judi Yuen

Palmer School of Library and Information Science, LIU Post

Abstract

The Robert Moses Collection Project (RMCP)is an ambitious venture to create widespread public access to historic records produced during the era when "master builder" Robert Moses developed most of Long Island's public parks and parkways. While head of multiple public commissions and authorities, Moses opened a largely rural Long Island up to the urban masses of New York City for public recreation and subsequently enabled the advance of suburbanization. He essentially shaped Manhattan and Long Island as we know them today.

While renowned in some circles, charges of racism, classism, and segregation have also plagued the Moses legacy. Questions regarding the implications his actions and how they have reverberated to shape modern society continually arise. Since the recent unveiling of President Biden's new plan to address historical racial inequities in American infrastructure, the collection has moved beyond regional to take on national significance.

Through funding provided by the Robert David Lion Gardiner Foundation, Long Island University archivists are using New York State Archives guidelines and procedures to create digital access to over 500 cubic feet of New York State Department of Parks records dating to the Moses era. Through an extensive digitization effort, this long-hidden collection of government records will finally be made publicly accessible on the New York State Archives website. Providing widespread access to these important records will empower and encourage responsible factual research and assist in expanding governmental transparency and accountability.

The RMCP archivists propose the presentation of a lightning talk at iPRES2021 regarding our work and the collaborative efforts of the public and private institutions invested in the project. While based in the United States, our work (particularly our project framework) has international relevance. The "PUPS Model" framework developed for the project consists of four main elements (Private funding, University staff, Public records, and State archives) and can be used to guide other organizations interested in pursuing similar important ventures.

LIGHTING TALK

ADVANCING DIGITAL CURATION SCHOLARSHIP THROUGH ICA'S STUDY SCHOOLS

Forget Chaterera-Zambuko

Sorbonne University Abu Dhabi

Abstract

ICA's training programmes are indisepensable in advancing scholarship and capacitating archival educators towards addressing topical issues in the broader discipline of records nad archives management. In August 2019, I was privileged to participate in a Digital Records Curation Programme Study School facilitated by James Lowry and Margaret Crockett as part of ICA's Africa Programme. The study school occurred at an opportune time when the Records Management Department at the National University of Science and Technology in Zimbabwe was in the process of rebranding its Bachelor's Degree programme in Records and Archives Management and reviewing the curriculum to match the many digital advancements that have engulfed the world across all disciplines, including records and archives management. Using the content and knowledge I acquired during the training programme, I was able to contribute immensely to the curriculum review and rebranding exercises. The proposed changes have since been approved by Zimbabwe's Council on Higher Education and the department's name changed to Information Management and Technology. The newly approved programme is expected to commence in the next academic year beginning August 2021. Using the knowledge I gained from the training programme, I also developed a Digital Curation syllabus for the Bachelors' Degree in Records management and Archival Science at Sorbonne University Abu Dhabi.

LIGHTING **TALK**

DEVELOPING A FLEXIBLE SKILLS FRAMEWORK FOR DIGITAL PRESERVATION

Sharon McMeekin

Head of Workforce Development, Digital Preservation Coalition

Amy Currie

Training Officer, Digital Preservation Coalition

Abstract

It is widely accepted that skilled staff are the most importance resource an organization can have when aiming for digital preservation success. But what skills do those staff members need to have? Past projects have addressed this issue, including DigCurV and DigCCurr, but the outputs are around a decade old and no longer represent the current landscape of quickly advancing digital preservation practice. The roles defined by existing frameworks are also relatively fixed and did not necessarily align well with the realities of the diverse staffing structures and individual roles at organizations around the world.

With many of the Digital Preservation Coalition (DPC)'s members identifying difficulties with pursuing a structured approach to skills development, the DPC is working to develop a new skills framework aligned to their popular DPC Rapid Assessment Model. The new framework will represent current good practice and offer more flexibility for users by offering different "views" of skills. The framework will also be accompanied by additional supporting resources including a skills audit toolkit. This lightning talk will report on current progress with the development of the framework and an overview of next steps.

LIGHTING
TALK

PRESERVATION WATCH: BUILDING TOWARDS A NEW GENERIC FACILITY WITHIN THE DUTCH DIGITAL HERITAGE NETWORK

Tamara van Zwol

Coordinator Preservation Watch for The Dutch Digital Heritage Network (NDE), The Netherlands Institute for Sound And Vision

Abstract

Preservation of digital materials can and will be influenced by all kinds of entities, like for example advancing technologies, organisational policies, the changing needs from your designated communities or even climate change. Some of these developments can propose a risk to the life cycle and sustainability of digital materials. Therefore it is important to monitor, to watch those internal and external developments in order to take appropriate measures in time. This monitoring function is called Preservation Watch and needs to be implemented within your organisation and preservation policy. But how do you keep track of this array of different developments and possible risks? Especially if you are an organisation with limited resources? Being part of a network helps: a community of heritage organisations where you can find and share (practical) expertise, signal and observe developments, do research and ask experts to address specific topics within the context of preservation watch, e.g. the influence of artificial intelligence on preservation. That is why we started in March 2021 building towards a generic facility for Preservation Watch. The Netherlands Institute for Sound and Vision is the coordination party, working together with a group of experts from The Dutch Digital Heritage Network. In this talk, I will address how we are building up and organising this facility, how to involve the (international) community and which topics are to be monitored on this communal level.

LIGHTING TALK

COMMUNITY HISTORY WEB ARCHIVING PROGRAM

Lori Donova

Community Programs Manager, Web Archiving & Data Services, Internet Archive

Abstract

The Internet Archive's Community History Web Archiving Program (Community Webs) provides public libraries and cultural heritage organizations with technical services, training and cohort development to build expertise in web archiving and documenting local civic life, especially of marginalized groups and those overlooked or misrepresented in the historical record. This lightning talk would highlight the goals and outcomes of this program over the last 4 years, as well as opportunities and initiatives for international expansion.

L IGHTING TALK

COVID-19 WEB ARCHIVE

Lori Donova

Community Programs Manager, Web Archiving & Data Services, Internet Archive

Abstract

Our current global moment reminds us of the central role of the web as a historical record and communication platform, if an ephemeral and technically challenging one, and has presented the institutions responsible for archiving web-published materials an opportunity to expand their collecting activities to document the pandemic's wrenching changes upon our institutions, communities, neighborhoods, and lives. Additional opportunities exist in facilitating the use of these materials by the The Internet Archive has worked with a diverse set of over 200 libraries from around the world to enable them to create hundreds of COVID-19 digital collections. Taken together these collections include over 300 million unique web documents, social media posts, audio and video files, datasets, and other born-digital resources archived from the web, totaling over 90 terabytes of archival materials documenting the impact of the pandemic on communities and citizens. We will give a brief overview of the state of COVID-19 Web Collecting as well as work to enhance access and encourage research use of the collections.

L IGHTING TALK

ARK ALLIANCE: EMPOWERING 800 INSTITUTIONS AND 8 BILLION IDENTIFIERS SINCE 2001

John Kunze
....
California Digital Library

Abstract

The ARK Alliance comprises organizations and individuals who are growing the community of researchers and scholars that rely on ARK identifiers for long term reference. Archival Resource Keys (ARKs) serve as persistent identifiers, or stable, trusted references for information objects. ARKs aim to be web addresses (URLs) that don't return 404 Page Not Found errors. Since 2001 some 8.2 billion ARKs have been created by 800 institutions — libraries, data centers, archives, museums, publishers, government agencies, and vendors. ARKs are open, mainstream, non-paywalled, decentralized persistent identifiers that you can start creating in under 48 hours. They identify anything digital, physical, or abstract. In contrast to other persistent identifiers schemes, ARKs are cheaper, more flexible, and less centralized, empowering you to • create unlimited identifiers without paying for the right to do so, • add any kind of metadata, including no metadata, • append extensions and query strings during resolution, • link directly to an article, image, or spreadsheet that is immediately usable by people and software without making them first stop at a landing page, and • make millions of ARKs resolvable by managing just one ARK, via a mechanism called suffix passthrough. This lightning talk will present an update on the ARK Alliance and the diverse and rapidly growing body of ARK identifiers.

L IGHTING TALK

TOWARDS MAKING YOUR REPOSITORY MORE TRUSTWORTHY AND FAIR-ENABLING

DOI: 10.17605/OSF.IO/7XG5N

Linas Cepinskas

Data Archiving and Networked Services (DANS), The Hague, Netherlands

Abstract

FAIRsFAIR builds the capacity and capability of data repositories through a tailored support programme. Since February 2020, the programme has been supporting 10 selected European data repositories, representing a diversity of countries and disciplines, in their journey to becoming more trustworthy and FAIR-enabling. Towards the end of the programme, it is expected that these repositories will achieve a CoreTrustSeal certification, marking their key milestone in this programme.

The lightning talk will briefly cover the main aspects of the support programme, emphasising the services and hands-on tools developed and used within the programme as well as the role of the participating repositories in informing the project with best practices for enabling FAIR data. It highlights the importance of peer network support as well and how repositories can help each other on their journey towards repository trust and data object FAIRness.

The lightning talk may interest data managers, data professionals, and policy makers engaged in long-term digital preservation and FAIR-enabling practices. FAIRsFAIR is an EU-funded project offering practical solutions for the use of FAIR principles throughout the research data life cycle. The project spans from 2019 to 2022 and involves 22 partners from 8 Member States. Its key objective is to accelerate the European Open Science Cloud by opening up and sharing all knowledge, expertise, guidelines, training and tools on FAIR matters.

LIGHTING TALK

PREMIS Rights: Keeping Pace with New Requirements

Marjolein Steeman

Preservation Officer
Netherlands Institute for
Sound and Vision,
member PREMIS
Editorial Committee

Abstract

One of the main entities of the PREMIS datamodel is the Rights entity. It has been designed to support a minimal registration of the rightsbasis for preservation actions. However, over time the usage of it has broadened. For instance: repositories wish to include other rights based actions than those strictly aimed at preservation. Also users ask us how to document obligations that the repository is committed to, in addition to granted rights. In general concerns have been raised whether the data model sufficiently supports the complexity of multiple restrictions and rules, rights holders, (internal) users that have different roles, uncertain expiration dates, review outcomes and so on. The PREMIS Editorial Committee has created a working group to do research on this topic. In this talk, the working group will present the first findings. In short three main issues will be explained and illustrated. These issues are open for discussion and the working group now invites you to add your comments or use cases to bring the work forward!

LIGHTING TALK

WHAT ARE YOUR DIGITAL PRESERVATION HEADACHES, AND HOW CAN WE TREAT THEM TOGETHER?

Lotte Wijsman

National Archives of the Netherlands

Abstract

We want to learn about your digital preservation headaches. And start to work on cures, together with you.

As the Open Preservation Foundation's Archives Interest Group (AIG) we have spent the last couple of years on an issue that was a headache for all AIG members: preserving spreadsheets. We soothed this headache by investigating the significant properties of spreadsheets. We have finished this work and are now looking for new headaches to treat. We will present an online platform for headache sharing. Headaches that were already shared with us will be included in this platform. Examples are explaining file format validation errors and warnings, and measuring preservation action successfulness.

Together with you, we want to start a new working group for finding practical, reusable cures for common, high-priority headaches – not deadly digital preservation diseases. The group will be supported by the OPF and isn't limited to OPF members.

L IGHTING TALK

The Radboud Data Repository: Digital Preservation Throughout the Research Lifecycle

Inge Slouwerhof

Radboud University Library, Nijmegen, The Netherlands

Didi Lamers

Radboud University Library, Nijmegen, The Netherlands

Abstract

At Radboud University a new tool has recently been developed to archive, publish and share digital research data acquired, processed, and analysed by researchers of the University. This novel repository, named the Radboud Data Repository, serves researchers throughout different stages in the research lifecycle.

To accomplish this, the repository allows researchers to archive their data into three types of data collections, collectively corresponding to a single research project. Raw, unprocessed data of the project are preserved in Data Acquisition Collections (DACs). The research process is documented in Research Documentation Collections (RDCs). Data on which a scientific publication is based are stored and publicly shared in Data Sharing Collections (DSCs). The collection types serve the goals of longterm internal data preservation for re-use, reproducibility, and scientific integrity (DACs and RDCs) and of open access data sharing with the external scientific community (DSCs).

The Radboud Data Repository is suitable for daily data handling and researchers can collaborate on their data with colleagues from inside and outside of the university by extensive role-based access management. When a data collection is complete, it can be archived (DACs and RDCs) or published (DSCs). Access to published DSCs can be managed by the researcher based on a wide variety of Data Use Agreements. All data collections are made findable by metadata indexing in a searchable resource, the assignment of a persistent identifier (DOI), and the availability of rich metadata fields. With this new repository, launched at the start of 2021, Radboud University strives to enable FAIRness of its research data, in line with the university's Research Data Management policy, and to enhance impact of its research.

LIGHTING TALK

THE HULL 2017 CITY OF CULTURE DIGITAL ARCHIVE

Laura Giles

....

University of Hull

Abstract

Hull in the UK was the holder of the UK City of Culture title in 2017. The City of Culture scheme defers the title to a different city every 4 years and guarantees funding for a year of cultural celebration and regeneration. Hull is one of the 5 local authorities in the UK with the most deprived neighbourhoods in England so a spotlight was thrown on the city as it looked to answer the question as to whether culture (art, dance, music, performance and drama) has the power to help struggling local economies and improve health, wellbeing and educational outcomes.

Recognising both the academic and local interest in preserving data and stories from the 2017 culture year the University of Hull seized the opportunity to develop digital archive storage, preservation and access for the first time with a newly acquired pilot digital collection drawn from major partners of, and artistic contributors to, the 2017 Culture year. Building on an existing university commitment to and relationship with open source technology the solution (that was developed in partnership with CoSector at the University of London) pulls together the open source components: archival software Archivematica, Hyrax repository and Blacklight online catalogue system with proprietary software Box cloud storage and Calm cataloguing software. This mix of components was designed to try and ensure as much consistency with our existing workflow for traditional archives as possible - both for our team of archive professionals and for our researchers.

In this lightening session I propose to showcase this technical architecture and share some brief insights from the development process. It should be of interest to any archive service looking to integrate digital archiving components into their existing archives workflows.

LIGHTING TALK

PEOPLE MAKE DIGITAL PRESERVATION

Sarah Middleton
....

Head of Advocacy &
Community
Engagement, Digital
Preservation Coalition

William Kilbride
....

Executive Director,
Digital Preservation
Coalition

Abstract

Exploring New Horizons

The iPRES conference is coming to Glasgow in 2022 and the DPC is looking forward to welcoming old friends and new to our home city. 2022 is a special year for the DPC: it is our 20th Anniversary, and we are excited to include iPRES attendees in some of the celebrations we have planned.

However, we also recognize that the past 18 months have affected our community in different ways, and we must adapt to provide accessible and inclusive means to come together again. As in Beijing, iPRES2022 will be a hybrid conference and we are exploring ways of creating a conference experience which is accessible to as many people as possible.

Scanning New Development and Building Capacity

The pandemic taught the whole community that whatever happens we need a space to exchange ideas, solutions, works in progress and the supportive conversation required to build skills and networks. #WeMissiPRES helped fulfill part of this need in 2020 and we are delighted to support colleagues in Beijing to deliver iPRES2021. We aim to learn from these experiences to ensure that iPRES2022 meets this critical need too.Enhancing Collaboration Continuing and complementing the work of the iPRES Working Group, since it was convened after iPRES2018 to undertake a review of the current structure and governance of iPRES, the DPC aims to make sure that iPRES2022 is accessible to any individual across the community who wishes to participate in the conference and its activities.

Therefore, this Lightning Talk will be an invitation to attend the launch of iPRES2022 at the end of this conference and discover some of what we have planned, but also to attend an iPRES2022 Community Consultation in November 2021 to share ideas about how we can best create a meaningful conference experience for all.

LIGHTING TALK

Long Term Preservation Scheme for Digital Assets by Macrowing Archivals

Xufeng Liu

Macrowing Software Technology Co., Ltd., Shanghai, China

Abstract

Innovative management of Digital Archives in the whole industry chain, Achieve the goal of permanent preservation of human memory.

LIGHTING
TALK

LEARNINGS FROM ARCHIVER PROJECT HOW TO PRESERVE RESEARCH DATASETS

Antonio G Martinez

LIBNOVA SL

Teofilo Redondo

LIBNOVA SL

Maria Fuertes

LIBNOVA SL

Abstract

LIBNOVA leads one of the consortia selected for the ARCHIVER Project. This Lighting Talk presentation would be an overview of the proposed solution, together with the University of Barcelona, the CSIC, David Giaretta, Voxility and Amazon Web Services. The solution developed will provide a Research, Management and Preservation Platform to solve obstacles for research dataset management (including preservation) identified at the beginning and throughout the ARCHIVER project.

LIGHTING TALK

LIBNOVA RESEARCH LABS LIBNOVA's Lines of Research in Technological Innovation

Antonio G Martinez
....

LIBNOVA SL

Teofilo Redondo
....

LIBNOVA SL

Maria Fuertes
....

LIBNOVA SL

Abstract

LIBNOVA will explain the lines of research to be followed in technological innovation by LIBNOVA Research Labs, its research dedicated department.

LIGHTING TALK

PRESERVING 300TB OF DATA IN 8 MINUTES

Antonio G Martinez

LIBNOVA SL

Teofilo Redondo

LIBNOVA SL

Maria Fuertes

LIBNOVA SL

Abstract

8 minutes is the time used by LIBSAFE to preserve 300TB of data. In this Lightning Talk, LIBNOVA explains how it has achieved to improve the ingest rate performance of its digital preservation platform.

LIGHTING TALK

PRESERVATION AS THE PRIORITY

Paul Stokes
....
JISC

Tamsin Burland
....
JISC

Abstract

The use of research management systems (such as CRISs, repositories and systems to manage research grants and contracts, ethics or impact) is well established across UK universities. With the drive to open research, these systems are crucial to an institution's ability to effectively manage their response to funder, legal, ethical and good practice mandates around research.

However, universities face many challenges in maintaining and supporting their portfolio of research management systems and research outputs, and ensuring they remain fit for purpose against a backdrop of a constantly evolving and sometimes unpredictable environment. (Digital) Preservation of research outputs persists as a particular problem for many institutions. There's a lack of understanding, lack ofknowledge, lack of systems and, perhaps most damming, a lack of funding.There is a school of thought that suggests that digital preservation should be the first thing a researcher should consider when preparing to embark on their research. Where are the outputs to be stored and kept usable for the long term. Data management plans are (relatively) well established and these theoretically are supposed to cover this, but the digital preservation aspect is still often glossed over or forgotten.

In this lightning talk we intend to explore how a research data management eco system might be envisioned with preservation at its heart; the Archival Information package as the primary home for research outputs.

LIGHTING TALK

The DPC in Australasia: Implementing A Global Strategy

Jaye Weatherburn
....

Digital Preservation
Coalition

Abstract

Digital preservation is a global concern which needs to be addressed as such. To foster the growth of the global digital preservation community in new geographies, the Digital Preservation Coalition (DPC) adopted a global objective as a supplement to its Strategic Plan (2018-2022) in June 2019. This global objective states the need for a "A Coalition Scaled to the Digital Preservation Challenge".

To achieve this objective, the DPC identified Australasia as one vibrant region of the global digital preservation community, particularly recognised by the active community of practice, Australasia Preserves. In January 2020, a memorandum of understanding for a two-year project was signed by the DPC and the University of Melbourne: a partnership established to explore and develop sustainable operations for the DPC in the Australasian region.

Since the establishment of the Australasian DPC office in early 2020, Australasian-based DPC members have grown from 3 to 14, highlighting the keen response from the digital preservation community in the region that is contributing overwhelming support for this initiative.

LIGHTING TALK

This lightning talk will share the collaborative work in Australasia that has brought greater attention and global access to the good practice and excellence within this dynamic regional community.

PRESERVING IMMERSIVE MEDIA

Osman Serhat Karaman

Sabancı University Sakıp Sabancı Museum, Istanbul, Turkey

Abstract

On April 8, 2021, we held an conference entitled Preserving Immersive Media with new media artist Jeffrey Shaw, Tate and ZKM's Time-based Media Conservation Department teams and the computer scientist as part of the Technological Arts Preservation Project, which is co-organized by the Sakıp Sabancı Museum and Sabancı University.

During this conference, we focused on Shaw's pioneering augmented reality installation entitled Virtual Sculptures (1981) together with Agnes Hegedüs's VR artworks entitled Memory Theater VR (1997) and delineated the technical and nontechnical central points which would ensure that the artworks remain exhibitable in the future.

In my talk, I would like to depart from this conference and delve into the non-technical difficulties of preserving this type of new media artworks. Later on, I would like to explore the following questions: Can the "impact" of an artwork be conserved, considering it has been technically conserved "just as it was"? How can we conserve the historicity of VR artworks? What should we take into account during documentation?

LIGHTING TALK

LIFECYCLE-BASED PRESERVATION MANAGEMENT OF RESEARCH DATA

Qing Lu

Shanghai Tech University, Shanghai, China

Yingdong Zhang

Shanghai Tech University, Shanghai, China

Yinbing Sun

Shanghai Tech University, Shanghai, China

Xiaoying Sun

Shanghai Tech University, Shanghai, China

Abstract

Research data management (RDM) is now an integral part of university management of research, research integrity, and research infrastructure. But RDM is hampered by the segmented treatment and the un-disciplined management of the research lifecycle. RDM fails when looking at it from a repository perspective and when considering it only when the data "is ready". ShanghaiTech University takes a lifecycle based approach to RDM to ensure the context of, interaction with, and evolution of research data along the research process are managed and preserved. Based on a brief discussion of the challenges to real life RDM, the paper describes, as a work in progress, the underlining framework, and its four components of the lifecycle-based RDM: project-based RDM, active research data management, deposit and sharing of supporting data for research, and distributed preservation of layered research data.

LIGHTING TALK

Consistent Evaluation of Trustworthiness Across Guiding Principles and Standards Is Crucial to Ensure A Robust Digital Repository Ecosystem

David Giaretta

PTAB LTD

John Garrett

Garrett Software

Mark Conrad

Adv. Info. Collaboratory

J Steven Hughes

Jet Propulsion Laboratory

Terry Longstreth

Washington Acad. Sciences

Roberta Svanetti

Deda.Cloud Srl Supporting ESA

Dawei Lin

NIAID

Robert R. Downs

CIESIN, Columbia University

Felix Engel

FernUniversität in Hagen

Matthias Hemmje

FernUniversität in Hagen

Iolanda Maggio

Rhea GROUP Supporting ESA, Italy

LIGHTING **TALK**

Abstract

This presentation calls for attention to established and emerging principles and standards developed from various perspectives to maintain and improve the robustness of the digital repository ecosystem. While the guiding principles and standards share a common goal to ensure the trustworthiness of digital repositories, a comparison can help identify areas where they may be clarified and strengthened in the future if they are to be used to perform consistent evaluations of repositories. Therefore, a group, whose members individually have participated in developing or revising the OAIS framework, the ISO16363 requirements, the CTS requirements, the TRUST Principles, or NESTOR, attempted to compare those standards and principles.

The effort aims to improve the consistency of audits using any one set of requirements rather than making the audits consistent between sets of requirements since they each serve a different purpose. The approach is to clarify the relationships between the concepts and terminology of ISO 14721(OAIS), ISO 16363, and ISO 16919 compared to the CoreTrustSeal Trustworthy Data Repositories Requirements (CTS), NESTOR, the FAIR Guiding Principles for Scientific Data, and the TRUST Principles for Digital Repositories. The comparison will help improve efforts to revise each of the instruments. Furthermore, such comparisons can improve the interoperability among these instruments and, consequently, enable future auditors or audit teams to have consistent interpretations among all the requirements. In addition, the group examined the auditing methods and will talk about lessons learned that help improve the consistency of audits and have identified specific improvements for the various sets of requirements.