

**Title:** Defining a Roadmap for Economically Efficient Digital Curation – a 4C Project Workshop

**Leaders:** Neil Grindley, Katarina Haage, Paul Stokes

**Type:** Workshop

**Duration:** Half-day

**Abstract:**

The 4C Project (a Collaboration to Clarify the Costs of Curation) is a European Commission funded two year coordination action which has been funded to provide useful, useable resources that provide better support to identify and quantify the cost of digital curation. From the outset, however, the project has taken the view that costs cannot be dealt with in isolation from a number of other related concepts (e.g. benefits, risk, quality, sustainability) and this holistic view might more accurately be described as an *economic* perspective on digital curation.

Borrowing the language of economics and mapping it onto digital curation needs to be done selectively and carefully. Digital assets do not have the same attributes as other kinds of (financial) assets and equally, it may not be possible to define when digital assets become (economic) liabilities in any objectively quantifiable way. However, there is still terminology from the field of economics that may help to define what the digital curation community might aspire to over the next few years and the starting point for this workshop is the concept of ‘economic efficiency’- which might be defined as the optimised situation where it is no longer possible to add quantity or value given a finite availability of resources.

The 4C Project is tasked with delivering a Roadmap report and it is this drive towards ‘economic efficiency’ in relation to digital curation that will be central to the agenda that it sets out. The consultation, stakeholder engagement, analysis and modelling work that have been done allow some principles to be proposed and some assertions to be made that will form the backbone of the report. The purpose of a Roadmap – particularly where it seeks to set out an action agenda for a range of stakeholders across various communities – is to make politically astute observations and to arrive at plausible conclusions. This is only possible via early interaction with stakeholders and by achieving some level of community validation before publication and this is the purpose of the workshop. One of the guiding principles of the 4C Project is to create a better understanding of the economics of digital curation through collaboration; and also to be an ‘open and social’ project and to listen to the needs of the community. iPRES 2014 occurs at roughly the three quarter point of the two year project and provides a timely opportunity to check and refine the draft Roadmap.

Early ideas and discussions about the structure and content of the Roadmap have indicated that it will need to address various questions.

- What vision should we advocate and what principles should we espouse to bring about economically efficient digital curation?
- What current economic inefficiencies do we need to eliminate?
- What or who is the most influential mechanism to bring that about and where will that influence most be felt?
- What is the policy, business and regulatory framework for digital curation and how is it likely to change?
- Over what timescales should we advocate action?
- How can we most economically sustain and exploit existing work? (including the 4C Project outputs)
- How are the economic requirements of stakeholders changing?
- Is it possible and economically desirable to try and align digital curation practice (including standards and terminology)?
- How can we most effectively invest in digital curation at the institutional, national and international level?

This workshop is an important opportunity to connect with stakeholders and get input for a critical deliverable of the project. But it is also an opportunity for participants to learn more about the economics of digital curation and to critically assess the efficiency and sustainability of their own services and solutions.

-----

**Title:** Born Digital Appraisal, Ingest, and Processing

**Leaders:** Jessica Moran, Leigh Rosin, Douglas Elford, Emma Jolley, Somaya Langley, Donald Mennerich, Ben Fino-Radin, Cal Lee, Erin O'Meara

**Type:** Workshop

**Duration:** Half-day

**Abstract:**

The workshop will bring together experienced digital preservation practitioners, specifically digital archivists, digital curators, and other practitioners with a responsibility for, or interest in, born digital preservation workflows and systems. A number of digital archivists will give short presentations sharing their experiences and insights around particular aspects of working with born digital collections, including initial technical appraisal, media transfer, ingest workflows, and specific tools. Discussion will focus around these workflows and tools, as well as issues or problems encountered, successes and failures, lessons learned, and future planning.

The workshop is designed to further the articulation of international current and best practices for digital archivists. The workshop will deal explicitly with the technical appraisal and ingest of born digital materials. As the field of born digital collecting and preservation has grown, the body of knowledge and experience in this area has also grown and there is a greater need than ever to share that knowledge among a wider group of practitioners.

The workshop will combine formal presentations, contributions from participants, and group discussion. As such, the workshop would best benefit those who already have experience as practitioners in the area of digital preservation workflows and systems. It is hoped the workshop will contribute to further networking and information sharing among digital archivists, curators, and others with a responsibility or interest in born digital preservation.

-----

**Title:** PREMIS Implementation Fair Workshop

**Leaders:** Eld Zierau, Peter McKinney, Rebecca Guenther

**Type:** Workshop

**Duration:** Half-day

**Abstract:**

This workshop provides an overview of the PREMIS Data Dictionary for Preservation Metadata, a standard addressing the information you need to know to preserve digital content in a repository. It includes an introduction to PREMIS and reports from the preservation community on implementation of the standard in various systems or contexts.

The PREMIS Implementation Fair Workshop is one of a series of events organized by the PREMIS Editorial Committee [<http://www.loc.gov/standards/premis/>] and that has been held in conjunction with previous iPRES conferences.

At iPRES 2014, the workshop will give the audience a chance to understand the PREMIS data dictionary and give implementers, and potential implementers, of the PREMIS Data Dictionary for Preservation Metadata an opportunity to discuss topics of common interest and find out about latest developments.

-----  
**Title:** ICA-AtoM, Archivemata and Digital Preservation

**Leaders:** Lise Summers, Meg Travers

**Type:** Workshop

**Duration:** Half-day

**Abstract:**

ICA-AtoM is an open source, web-based archival description software application that is based on International Council on Archives standards. The first version of it was released in 2008 with funding from a variety of organisations including UNESCO, World Bank Group Archives, the Dutch Archiefschool, the Direction des Archives de France and the United Arab Emirates Centre for Documentation and Research. In 2013, the State Records Office of Western Australia (SROWA) has invested in further development of the software, chiefly to include support for the Australian Series Registration System as well as simple preservation workflow. These additions to ICA-AtoM will be completed in the middle of 2014 and made freely available as part of the open source package to any archives wishing to download the software.

This work has aligned the Australian Series Registration System within an international standard ISAD compliant system, and will go some way to eventually bringing the two together.

This workshop will provide an overview of using ICA AtoM with special attention to archival description using the Australian Series Registration System; ingest into the complementary Archivemata digital preservation system and attaching digital objects to AtoM.

-----  
**Title:** From Preserving Data to Preserving Research: Curation of Process and Context

**Leaders:** Angela Dappert, Rudolph Mayer, Stefan Pröll, Andreas Rauber, Raul Palma, Kevin Page, Daniel Garijo

**Type:** Tutorial

**Duration:** Half-day

**Abstract**

Awareness of the need to provide digital preservation solutions is spreading from the core memory institutions to other domains, including government, industry, SME and consumers. In many of these settings we are, however, faced with preserving more than just data. In the domain of eScience, for example, investigations are increasingly collaborative. Most scientific and engineering domains benefit from building on the outputs of other research by sharing information to reason over and data to incorporate in the modelling task at hand.

This raises the need for preserving and sharing entire eScience workflows and processes for later reuse. We need to define which information is to be collected, create means to preserve it and approaches to enable and validate the re-execution of a preserved process. This includes and goes beyond preserving the data used in the experiments, as the process underlying its creation and use is essential.

The TIMBUS project and Wf4Ever project team up for this half-day tutorial to provide an introduction to the problem domain and discuss solutions for the curation of eScience processes.

-----

**Title:** Memento. Uniform and Robust Access to Resource Versions

**Leader:** Herbert Van de Sompel

**Type:** Tutorial

**Duration:** Half-day

**Abstract:**

The Memento protocol tightly integrates the Web of the Present and that of the Past, making it possible to seamlessly navigate between both. The protocol defines an interoperable approach to access versions of a resource in web archives or content management systems such as wikis that leverage the URI of that resource and the datetime of the required resource version. Technically, the Memento protocol is an extension of HTTP that is fully based on the primitives of Web interoperability: URIs, resource representations, links, content negotiation. The tutorial will give an in-depth insight in various aspects of the Memento protocol that meanwhile has been published as RFC 7089.

-----

**Title:** Digital Preservation Systems Showcase

**Leaders:** Ross King, Archivematica, RODA, Duraspace, Preservica, Ex Libris

**Type:** Showcase

**Duration:** Full-day

**Abstract:**

This session will explore the functionality of digital preservation systems available to the user community. System developers will showcase their systems in line with a pre-determined set of functions. These functions are derived from current state-of-the-art as described in standards, key literature and interest groups (e.g. PREMIS, OAIS, International Internet Preservation Consortium Preservation Working Group).

-----

**Title:** Modelling file formats and technical environments using the NSLA Digital Preservation Technical registry (DPTR)

**Leaders:** Jan Hutař, Ross Spencer, Libor Coufal, Kevin DeVorse, Jay Gattuso, Steve Knight, Peter McKinney

**Type:** Workshop

**Duration:** Full-day

**Abstract:**

This workshop introduces the work of the National and State Libraries funded work on a Digital Preservation Technical Registry. In particular it will allow participants to gain an understanding of the new model for modelling formats. They will be tasked with working through exercises designed not only to give participants an understanding of the model, but to test and critique it.

The heart of contemporary digital preservation is multi-faceted. Where once the technical registry sat at its core, we now seek 'war stories' from the community that detail experiences with legacy digital information. We seek high quality information about file formats, carrier mediums, software, and the complete picture of the technical environments that we're dealing with. With archival principle at the centre of our work we also look for detailed provenance, validation and verifiability of that information. Along with that, the technical registry still holds a key role in the community's multi-faceted approach. The technical registry, through enabling file format identification and validation, can aid the filtering and routing of content at the pre-deposit, and pre-ingest stages of the digital preservation lifecycle. The technical registry is still the core information source for the migration and maintenance of content as part of any technology-watch capability or preservation action.

We believe that the technical registry should provide information that is accessible to all involved in digital preservation at all levels of skills and knowledge. The information also needs to be actionable, that is, machine readable information that can be accessed by the tools in the digital preservation toolkit. Registries should use relationships to describe more complete technical environments – the links between specific instances of software, hardware, carrier mediums and file formats. Registry information also needs to be augmented with user-level, 'community text' describing war-stories, domain expert knowledge, and the institutional relevance of specific registry entries. Our aim is to create a registry flexible enough to contain and identify real life file format instances with all their specialities and varieties.

Striking that balance, we'll be introducing our work on the NSLA Digital Preservation Technical Registry, providing an overview of its core features before introducing our more radical changes in thinking. The biggest advancement we'll introduce is the overhaul of the format model traditionally used in digital preservation - presenting three interpretations of file format that we believe encompass the many different ways we talk about the subject in the community. Workshop participants will be given a more thorough introduction to this side of the work package, learning about these three components and the building blocks used to create them – format Aspects.

Using Aspects and the knowledge of how to create our three format objects, participants will engage in a modelling activity to help us challenge the work we've completed this far and help us to reinforce a step-change in thinking about the requirements of a modern, comprehensive technical registry.

-----

**Title:** Acquiring and processing Born-digital data using the BitCurator environment

**Leader:** Christopher A. Lee

**Type:** Tutorial

**Duration:** Full-day

**Abstract:**

This tutorial will prepare participants to use the open-source BitCurator environment to acquire and process born-digital data. There will be a brief lecture and discussion that focuses on the motivation for using the tools and several foundational technical concepts. The remainder of the tutorial will be devoted to demonstration and hands-on exercises that demonstrate specific tools and methods. Participants will learn how to mount media as read-only, create disk images, mount forensically packaged disk images, export individual files or entire directories from disk images, use Nautilus scripts to perform batch activities, generate and interpret Digital Forensics XML (DFXML), generate a variety of standard and customized reports (including PREMIS records), and identify various forms of sensitive data within collections.

-----

**Title:** Applying the TIMBUS Approach to Preserving Context in Digital Libraries

**Leaders:** Carlos Coutinho, Paul Gooding

**Type:** Tutorial

**Duration:** Half-day

**Abstract:**

To date, digital preservation has generally focused on the preservation of specific data in the form of artefacts. However, in order to ensure that this data is accessible in the long term, it is vital that we consider how to extract and preserve information on the software and hardware contexts which this

data depends upon to operate. We therefore need tools to assist in identifying, preserving and validating the processes which underpin the creation of data in digital libraries.

In particular, we need to consider the importance of preserving not just individual digital artefacts, but the platforms which allow digital libraries to render or execute their items. Digital libraries rely on this software to render their items, and it is therefore important to know configuration details and software dependencies to ensure these items remain fully operational in the future. In the case of digital libraries, the TIMBUS framework provides the tools necessary to assist practitioners in identifying relevant processes, undertake risk analysis, and then to assist the user in extracting, preserving and revalidating the necessary processes.

This half-day tutorial introduces the TIMBUS approach to process preservation, and demonstrates how it can be applied to issues relating to digital libraries. TIMBUS focuses primarily on business processes, but this tutorial will show its approach to process-oriented preservation is also relevant to digital libraries. It provides a methodology for process preservation and a set of tools which help to semi-automatically validate and preserve processes so that they can be recreated at a later date. Participants will be given the knowledge to understand the importance of technical environments for collection items, and learn more about the TIMBUS solutions through examples relevant to the digital library domain. They will also gain an understanding of digital preservation as a risk mitigation strategy.

-----

**Title:** Surveying ISO Standards for PDF: archive, accessibility, engineering, metadata, 3D data and PDF itself

**Leader:** Duff Johnson

**Type:** Tutorial

**Duration:** Half-day

**Abstract:**

This tutorial provides high-level information on the features, development and current status of ISO standards for PDF relevant to archiving.

Presented by an industry association representative, the workshop will begin with an exploration of the world of PDF and PDF-related standards. A series of breakout sessions will challenge the participants to carefully define and characterize the problems they currently experience and suggest a variety of solutions.

The workshop participants will all review each of the breakout session discussions and offer a collective opinion on the various problems and solutions identified.

-----

**Title:** Leveraging Web Archiving tools for Research and Long-term Access

**Leader:** Lori Donovan

**Type:** Workshop

**Duration:** Half-day

**Abstract:**

Web archiving is an important part of the digital preservation field. While most are familiar with the Wayback Machine available at archive.org, less are aware that there are a number of tools and services developed for organizations and individuals to create their own web archives, including the

capability to search and analyze large data sets built around the WARC file format, an ISO standard for web archiving. In addition, web archives provide permanent URLs for citation and can show how a website has changed over time at a single URL, even if no longer available on the live web. In short, web archives can provide very necessary preservation tools for researchers and archivists to manage content that is only posted on the web.

This workshop will introduce participants (15-20) to basic web archiving concepts and challenges. Using the Archive-It ([www.archive-it.org](http://www.archive-it.org)) web application, participants will have a hands-on opportunity to build a collection of content archived from the web, which can include their own organization's web presence, social media, digital exhibitions, data sets, or topical content publicly available on the web. Following the workshop participants will have a searchable archive available to them, including the option of downloading WARC files for long-term preservation or research.

The target audience for this workshop includes interested scholars researching the web and professionals responsible for digital library services or digital archives. No prerequisite knowledge of or experience with web archives is necessary, and the session does not require any programming or advanced technical knowledge of the web. The workshop will not be oriented towards those with deep knowledge of web archives or the WARC format, although there could be time allotted to a demonstration of another web archiving tool or project related to web archiving.