

DELIVERABLE

Project Acronym: DCH-RP

Grant Agreement number: 312274

Project Title: Digital Cultural Heritage Roadmap for Preservation - Open Science Infrastructure for DCH in 2020

D3.2 Standards and interoperability best practice report

Revision: 1.0

Authors:

Börje Justrell (Riksarkivet – National Archives of Sweden)
Eva Toller (Riksarkivet – National Archives of Sweden)

Reviewers:

Tim Devenport (EDItEUR)
Roberto Barbera (INFN)
Giovanni Ciccaglioni (ICCU)

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	P
C	Confidential, only for members of the consortium and the Commission Services	

Revision History

Revision	Date	Author	Organisation	Description
0.1	2013-09-30	Börje Justrell	Riksarkivet	First draft
0.2	2013-10-17	Eva Toller	Riksarkivet	Inserted contributions from Tim and corrected minor errors
0.3	2013-10-18	Eva Toller	Riksarkivet	Added potential FP7 projects in section 3.1. Inserted corrected/missing links (supplied by Gordon). Formatted section 4.4 (eCSG). Structured the items in section 3.2 (Key topics). Started on the Executive Summary.
0.4	2013-10-22	Eva Toller	Riksarkivet	Added suggestions from Gordon for Chapter 4. Added PREMIS and MoReq2 in section 4.1.1. Cleaned up Executive Summary and section 3.1. Wrote last chapter (Issues and Challenges; Conclusions). Renamed section 4.4 "Standards used in different countries and institutions" and made "Standards used by the e-csg" a sub-section; added another sub-section "Standards used by the ICCU". Included cloud standards (tip from Roberto) in section 4.1.5. Inserted EPIC and APARSEN (tip from Sanja) in section 3.1.
0.5	2013-10-25	Eva Toller	Riksarkivet	Made changes suggested by Gordon (including correction of English language in the whole document). Added DC-NET in section 3.1 and SPARQL in section 4.1.4 (tip from Roberto). Inserted hyperlinks for projects in section 3.1 and sorted the "dotted" projects in alphabetic order. Added EUDAT and Europeana Inside in section 3.1 (tip from Maciej and Sanja, respectively). Re-ordered items in tables in sections 4.1.1 and 4.1.2. Added MusicXML in 4.1.1. Added ODF in 4.1.2, and also all standards suggested by Tim/Graham.
0.6	2013-10-28	Eva Toller	Riksarkivet	Made some language corrections and added section 4.4.3.
1.0	2013-10-28	Claudio Prandoni	Promoter	Formal check

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

TABLE OF CONTENTS

1	EXECUTIVE SUMMARY	4
2	INTRODUCTION	5
2.1	STRUCTURE OF THE DOCUMENT	5
2.2	OBJECTIVES OF THE DELIVERABLE	5
3	PROJECTS AND INITIATIVES IN ARCHIVING AND PRESERVATION	6
3.1	RELEVANT INITIATIVES.....	6
3.2	KEY TOPICS FOR THE CULTURAL HERITAGE AREA.....	8
4	STANDARDS, BEST PRACTICE, AND IDENTIFIERS	10
4.1	IMPORTANT STANDARDS	10
4.1.1	<i>Descriptive Metadata Standards for Particular Domains</i>	10
4.1.2	<i>Standards for Types of Digital Content</i>	14
4.1.3	<i>Standard Licenses and Methods of License Expression</i>	21
4.1.4	<i>Standards for Search, Retrieval, and Harvesting</i>	22
4.1.5	<i>Basic Standards</i>	23
4.2	BEST PRACTICE AND “HOW-TO” GUIDES	24
4.3	PERSISTENT IDENTIFIERS FOR DIGITAL ARTEFACTS AND ASSOCIATED ENTITIES.....	25
4.4	STANDARDS USED IN DIFFERENT COUNTRIES AND INSTITUTIONS	26
4.4.1	<i>Standards used by the E-CSG</i>	26
4.4.2	<i>Standards used by the ICCU</i>	26
4.4.3	<i>Standards used by the KIK-IRPA</i>	26
5	ISSUES AND CHALLENGES; CONCLUSIONS	27

1 EXECUTIVE SUMMARY

In this document we present the standards, best practices, and identifiers that are of interest for the Digital Cultural Heritage (DCH) sector. More often than not, other projects have already identified a large part of these issues. We refer directly to these projects and their web sites in order avoid duplication of effort. However, we also directly give short descriptions and references to various types of important standards, and discuss issues and challenges regarding these standards.

Note that specific software tools for DCH preservation tasks are not described in this document; instead, we refer to [Task D3.3, Registry of Services](#) .

The “digital world” is international and interdisciplinary in a way that makes it possible to recommend standards (which once were developed in a specific country and/or for a specific (scientific) discipline) to a very wide range of users and for a very wide range of uses. Therefore, we do not separate general standards from standards developed specifically for the DCH sector, but rather divide them according to basic usage (for example, standards for search and retrieval). The exception is standards for metadata, which in general are rather domain-specific.

We also list some best practices “How-To” guides, which give practical advice for digital preservation.

2 INTRODUCTION

DCH-RP is a 24-month EU co-funded project (coordinated action) with the main goal to create a Roadmap for the implementation of a Preservation Infrastructure for Digital Cultural Heritage (DCH) which will be the first instance of an Open Science Infrastructure for DCH in 2020.

The Roadmap will be supplemented by practical tools which can help the monitoring of activities and will therefore hopefully be of great benefit for the stakeholders, such as the EC and national ministries of culture.

DCH-RP's activities are divided into five work packages (WPs) which carry out activities involving partners, infrastructure communities and other contributors. All results will be shared with the expert community and disseminated through the website as well as through workshops and dedicated events.

2.1 STRUCTURE OF THE DOCUMENT

This deliverable is a part of the work in WP3 and its Task 3.2, *Standards and interoperability*. The deliverable reports the results of Task 3.2 with particular regard to existing projects and initiatives as well as standards, tools, workflows, approaches, solutions, demonstrators, and applications. All these are useful for the two communities of DCH and e-Infrastructures, and also approach the digital preservation issues.

In Chapter 3, these projects, initiatives etc. are presented under a subtitle for each category. In Chapter 4, the standards, best practices, and identifiers are briefly described. In Chapter 5, some issues and challenges are presented, together with some conclusions.

2.2 OBJECTIVES OF THE DELIVERABLE

Task 3.2 has very broad objectives, aiming in the first place to do research on current projects, standards etc. During the work it became clear that much of the core knowledge that was planned to be the content of this report is already published or available on the Web. Instead of using resources on just repeating what can be studied elsewhere, this deliverable gives a short overview of existing projects and initiatives, standards, tools, etc., that can be seen as best practices.

Results of the work of Task 3.2 will also be integrated in

- Deliverable D3.3 Registry of services, where the compliancy to standards will be assessed;
- Deliverable D.3.4 Intermediate version of the Roadmap, which will give examples of best practices on (and if possible, also role models for) interoperability between the DCH community and e-Infrastructure.

3 PROJECTS AND INITIATIVES IN ARCHIVING AND PRESERVATION

3.1 RELEVANT INITIATIVES

During the last few decades, archival science research has been carried out in the area of digital preservation. This has focused on the practical testing of strategies, the analysis of risks attached to different strategies, and different systems architectures for the different strategies.

Research projects as the EU FP6 projects [NEDLIB](#), [ERPANET](#), [PRESTO](#), [MINERVA](#), [DELOS](#), CEDARS and CAMiLEON, and the EU FP7 projects [DC-NET](#), [INDICATE](#), and [APARSEN](#), have contributed to the development of practical action models and an increased awareness of the difficulties that need to be addressed. [Europeana Inside](#) is a CIP ICT-PSP network, working to transform the ability of European cultural heritage institutions (museums, archives and libraries) to participate in the Europeana platform.

Other international projects, for example [InterPARES](#), [NARA/SDSC](#), tested strategies for digital preservation and organisational and technical issues in conjunction with implementation of the strategies. Some national archives around the world are also very active, e.g. the National Archives of Australia, the National Archives of United States, the UK National Archives, the Swiss National Archives, and the Danish National Archives. Universities such as Monash University in Australia and San Diego Supercomputer Center at University of California have also carried out interesting work.

Among the most prominent and relevant digital preservation and digitisation projects and networks (listed without priority) are:

- BRICKS which established the organisational and technological foundations of an open networked system in order to integrate distributed collections of multimedia resources in museums, libraries, and archives (<http://www.brickcommunity.org>);
- CASPAR which focused on building archive information systems based on the OAIS standard (the conceptual model Open Archival Information System, ISO 14721:2003) (<http://www.casparpreserves.eu>);
- DELOS which integrates European research in the field of digital libraries, including library architectures, information access and personalisation, audio-visual and non-traditional objects, user interfaces, knowledge extraction, semantic interoperability, preservation, and evaluation (<http://www.delos.info>);
- DIAS/kopal (koLibRi) which provided a flexible and scalable open deposit library solution for storing and retrieving massive amounts of electronic documents and multimedia files (http://kopal.langzeitarchivierung.de/index_koLibRI.php.en);
- Digital Preservation Coalition which is an advocate and catalyst for digital preservation by raising "*awareness of the importance of the preservation of digital material and the attendant strategic, cultural and technological issues*" (<http://www.dpconline.org/>);

- DPE which is a network for the coordination of and co-operation between actors to create a common basis of knowledge (<http://www.digitalpreservationeurope.eu>);
- ENSURE which is extending the state of the art in digital preservation by looking at: costs and value; the preservation lifecycle management; content-aware long term Data Protection; and the utilisation of emerging ICT (<http://ensure-fp7-plone.fe.up.pt/site/>);
- EPIC which provides persistent identifier services to the European Research Community (<http://www.pidconsortium.eu/>);
- EPOCH which is a network of approximately one hundred European cultural institutions, joining their efforts to improve the quality and effectiveness of their use of information and communication technology (<http://www.epoch-net.org>);
- EUDAT which is developing and running the Collaborative Data Infrastructure which is a layered set of data storage and management services for data preservation targeted to various scientific disciplines including humanities, Earth and climate research, biodiversity and ecosystems, neuro-informatics, virtual human and cultural heritage organisations (<http://www.eudat.eu>);
- IMPACT which deals with “*large scale digitisation (projects) transforming Europe’s printed heritage into digitally available resources*”, utilizing innovative OCR technology (<http://www.impact-project.eu>);
- InterPARES which is developing a knowledge-base needed for the long-term preservation of authentic digital records (<http://www.interpares.org/>);
- LDP Centre which is a competence centre for research and technical development and testing of methods and technologies for long-term digital preservation and access (<http://www.ltu.se/centres/Centrum-for-langsiktigt-digitalt-bevarande-LDB?l=en>);
- LiWA which dealt with conversion of web content to achieve long term interpretability and improved archive fidelity for a wide variety of content (<http://liwa-project.eu>);
- Papyrus which developed a cross-disciplinary library engine, allowing the exchange of information between distinct groups of users (<http://www.ict-papyrus.eu>);
- PLANETS which deals with the automation of preservation planning and the collection of information concerning file formats (<http://planets-project.eu>);
- PRELIDA which is targeting the stakeholders of the linked data community (who have not in general been targeted by the digital preservation community) in order to make them aware of the digital preservation solutions that are already available (<http://prelida.eu/>);
- SCAPE which is directed towards long term digital preservation of large-scale and heterogeneous collections of digital objects (<http://www.scape-project.eu/>);

- SHAMAN which centred on storage solutions and grid technology (<http://www.shaman-ip.eu>).

3.2 KEY TOPICS FOR THE CULTURAL HERITAGE AREA

The projects in the framework of Cultural Heritage concentrate on some special topics in multidisciplinary research areas but with different goals. Nevertheless, the results of these projects can be related to digital preservation. Some of the listed project goals can also be extracted and seen as more general needs and requirements in a digital preservation context. Examples (listed regardless of priority) are the following:

- **Miscellaneous issues**
 - Reliability and robustness
 - Assurance of valid licensing procedures, commercial conditions, and transactions
 - Open, scalable, and flexible solutions (built on open industry standards like J2EE and XML)
 - Ease of use (for example, user-friendly interfaces)
 - OAIS compliance
 - Multilingualism
- **Content/information issues and metadata issues**
 - Mechanisms for integration and automation of appraisal and ingestion of digital material
 - Automatic metadata capture and extraction
 - Separation of content (information) and metadata
 - Various content formats (from print-based document to digitized images)
 - Ontologies for both visual and textual concepts
 - Annotation services
- **Performance issues**
 - Scalability (up to hundred terabytes or more)
 - Performance for hundreds of thousands of electronic documents
- **Trust issues and security issues**
 - Authenticity and integrity of data
 - Continuity (which means the handling of information, both data and metadata, for at least the next 100 years)
 - Identification of digital objects which are in danger of becoming inaccessible due to changes in technology
 - Security during transmissions of files between countries

- Validation (certification) of software and hardware environments required to render the digital objects
- **Infrastructure-related issues**
 - Distributed systems
 - Virtualisation
- **Hardware-related issues**
 - Support of many storage media and devices
 - Backup and restore

4 STANDARDS, BEST PRACTICE, AND IDENTIFIERS

4.1 IMPORTANT STANDARDS

The extensive use of relevant and open standards is a vital pre-requisite if the Cultural Heritage community is to promote interoperability, encourage widespread access and control costs in its digital preservation programmes.

Extensive reviews under the auspices of the Minerva (2008), Athena (2009) and Linked Heritage (2011) projects have already categorized and described many of the standards that are most applicable or recommended in this area. Rather than repeat those exercises, the approach taken in the current work is therefore to provide pointers to the main areas covered earlier, whilst highlighting topics or frameworks that have emerged or gained traction more recently.

The more relevant deliverables from the earlier projects are available as follows:

- Athena: <http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents>
 - *D3.1, Report on Existing Standards Applied by European Museums.*
 - *D3.2, Recommendations and Best Practice Report.*
- Linked Heritage: <http://www.linkedheritage.eu/index.php?en/142/documents-and-deliverables>
 - *D2.1, Best practice report on cultural heritage linked data and metadata standards.*
 - *D2.2, State of the art report on persistent identifier standards and management tools.*

4.1.1 Descriptive Metadata Standards for Particular Domains

Descriptive metadata standards for a number of domains – for example, museums, archives, libraries, etc. – were surveyed in the Athena and Linked Heritage projects and elsewhere. The following table summarizes a number of important standards in the Cultural Heritage area, together with several standards such as LIDO that are of a more general or cross-cutting nature.

Standard	Domain	Description	More information
<i>EAD</i> Encoded Archival Description	Archive	An XML standard for encoding archival finding aids	http://www.loc.gov/ead/ead.xsd (W3C schema)
<i>ISAD (G)</i> General International Standard Archival Description	Archive	General rules for archival description that may be applied irrespective of the form or medium of the archival material	http://www.icacds.org.uk/eng/ISAD(G).pdf
<i>OAIS</i> Open Archival Information	Archive, cross-domain	A conceptual reference model for an open archival information	http://public.ccsds.org/publications/archive/650x0m2.pdf

Standard	Domain	Description	More information
System, ISO 14721:2012		system (OAIS). An OAIS is an archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a “designated community”	
<i>MoReq2</i> Model Requirements Specification for the Management of Electronic Records	Archive, electronic records	An XML standard for electronic records, developed by the DLM Forum	http://www.moreq2.eu/home
<i>ISAAR (CPF)</i> International Standard Archival Authority Record for Corporate Bodies, Persons and Families	Archive, party information	A standard for authority records for corporate organizations, persons and families	http://www.icacds.org.uk/eng/ISAAR(CPF)2ed.pdf
<i>ONIX for Books</i>	Book publishing and supply	An XML-based standard for communicating book product information	http://www.editeur.org/83/Overview/
<i>VCARD</i>	Business, party information	A standard for describing electronic business cards	http://tools.ietf.org/html/rfc6350
<i>Indecs</i> Interoperability of data in e-commerce systems	Conceptual, e-commerce	Provides a framework for metadata requirements for e-commerce in content (intellectual property), focusing on semantic interoperability	http://www.doi.org/topics/indecs/indecs_framework_2000.pdf
<i>DBpedia Ontology</i>	Cross-domain	A cross-domain ontology, based on the “infoboxes” of Wikipedia	http://dbpedia.org/Ontology
<i>Europeana Data Model</i>	Cross-domain	Created for structuring data for Europeana ingestion, management and publication, and improves on Europeana’s basic data model, the Europeana Semantic Elements	http://pro.europeana.eu/edm-documentation

Standard	Domain	Description	More information
		(ESE)	
<i>VRA Core</i> Visual Resources Association	Cross- domain	A data standard for the description of works of visual culture and the images that document them	http://www.vraweb.org/projects/vracore4/index.html
<i>Premis</i> Preservation Metadata: Implementation Strategies	General heritage	An XML-based metadata framework	http://www.loc.gov/standards/premis/
<i>CIDOC-CRM</i> CIDOC Conceptual Reference Model	General heritage, conceptual	Provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation	http://cidoc.ics.forth.gr (CRM website)
<i>Dublin Core</i>	General, cross-domain	A simple metadata element set intended to facilitate discovery of electronic resources	http://dublincore.org/documents/dc-rdf
<i>SKOS</i> Simple Knowledge Organization System	General, cross-domain	Designed for the publication of controlled structured vocabularies for the Semantic Web, including thesauri, classification schemes, taxonomies, and subject headings	http://www.w3.org/TR/2009/REC-skos-reference-20090818
<i>Basic Geo</i>	Geography	An RDF vocabulary for basic geographical information: latitude, longitude, and altitude	http://www.w3.org/2003/01/geo
<i>MIDAS Heritage</i>	Historic environment	A standard for the management of the historic environment	http://www.english-heritage.org.uk/publications/midas-heritage/midas-heritage-2012-v1_1.pdf
<i>FRBR</i> Functional Requirements for Bibliographic Records	Library	A conceptual entity-relationship model for use with online library catalogues and bibliographic databases	http://www.ifla.org/VII/s13/frbr/frbr.pdf
<i>MAB2</i> Maschinelles Austauschformat	Library	An automated exchange format for libraries, since superseded by MARC-	http://www.dnb.de/EN/Standardisierung/Formate/MAB/mab_node.html

Standard	Domain	Description	More information
für Bibliotheken		21	
MARC MAchine Readable Cataloging	Library	A set of digital formats for the description of items catalogued by libraries, such as books	http://www.marc21.ca/index-e.html (MARC21) Five types of formats are supported, respectively for Authority (or authorized form), Bibliographic, Classification, Community Information and Holdings http://www.openarchives.org/OAI/2.0/guidelines-marxml.htm (MARCXML)
METS Metadata Encoding and Transmission Standard	Library	An XML-based standard for encoding descriptive, administrative and structural metadata regarding objects within a digital library	http://www.loc.gov/standards/mets/
MODS Metadata Object Description Schema	Library	An XML-based bibliographic description schema	http://www.loc.gov/standards/mods/
CDWA Categories for the Description of Works of Art	Museum	Describes the content of art databases by articulating a conceptual framework for describing and accessing information about objects and images	http://www.getty.edu/research/institute/standards/cdwa/index.html
<i>museumdat</i>	Museum	A harvesting format for providing core data from museum holdings	http://museum.zib.de/museumdat/museumdat-v1.0.xsd (XML schema)
<i>Object ID</i>	Museum	Description of cultural objects, especially of use if objects are stolen	http://archives.icom.museum/objectid/
SPECTRUM	Museum	A collections management standard including data elements for management and description	http://www.collectionslink.org.uk/spectrum-standard
LIDO Lightweight Information Describing Objects	Museum, general	A harvesting format for collections, particularly from museums, to portals	http://www.lido-schema.org/schema/v1.0/lido-v1.0-specification.pdf
MO	Music	Contains concepts and properties for describing	http://musicontology.com

Standard	Domain	Description	More information
Musical Ontology		music, for example: Artists, Tracks, Performances, Arrangements	
<i>MusicXML</i>	Music	An XML-based file format for representing Western musical notation	http://www.musicxml.com/
<i>FOAF</i> Friend Of A Friend	Party information, roles	A format, using RDF and OWL, for describing persons, their relations to other persons and things, and their activities	http://xmlns.com/foaf/spec
<i>BIBO</i> Bibliographic Ontology	Publications	A format, using RDF, for describing bibliographic items like books, magazines, and newspaper pages	http://bibliontology.com/specification

4.1.2 Standards for Types of Digital Content

Deliverable D3.2 of the Athena project looked at the mainstay standards for various categories of digital content commonly encountered in Cultural Heritage preservation. Athena described and recommended standards for the following content types: Text, Images, Audio, Video, Vector Graphics and Virtual Reality/3D Representation.

This work was then augmented by additions gathered during the Linked Heritage project. Athena and Linked Heritage deliverables presented a great deal of background detail and a headline summary is provided in the following table. Short references to EPUB-3 and HTML-5 have been added, given the increasing significance of those standards for text-based (and other) e-resources.

Standard	Content type	Description	More information
<i>AAC</i> Advanced Audio Coding	Audio	A standardized, lossy compression and encoding scheme for digital audio	ISO/IEC 13818-7:2006
<i>AIFF</i> Audio Interchange File Format	Audio	A non-compressed audio format most widely found on Apple Macintosh computers	http://www- mmsp.ece.mcgill.ca/Documents/Audio Formats/AIFF/Docs/AIFF-1.3.pdf (Version 1.3)
<i>AU</i>	Audio	A sound format for UNIX systems, the	http://www- mmsp.ece.mcgill.ca/Documents/Audio

Standard	Content type	Description	More information
		"standard" audio file format for Java	Formats/AU/AU.html
<i>FLAC/ALAC</i> Free Lossless Audio Codec	Audio	FLAC is similar to MP3 but offers better quality audio. ALAC is the Apple Mac version	https://xiph.org/flac/format.html (FLAC) http://alac.macosforge.org (ALAC)
<i>MP3</i> MPEG Layer 3	Audio	An audio compression format commonly encountered on the internet	ISO/IEC 11172, 13818, 14496
<i>RM</i> Real Media	Audio	A format optimized for delivery of audio over the Web	Proprietary format
<i>Vorbis</i> (Ogg Vorbis)	Audio	An audio format specification and software implementation (codec) for lossy audio compression	http://www.vorbis.com/
<i>WAV / RIFF</i> Resource Interchange Format	Audio	A format for sampled audio	http://tools.ietf.org/html/rfc2361
<i>WMA</i> Windows Media Audio	Audio	A proprietary competitor to MP3, optimized to deliver audio (particularly streamed content) over the Web	Based on ASF, see http://www.microsoft.com/windows/windowsmedia/forpros/format/asfspec.aspx
<i>ASCII</i> American Standard Code for Information Exchange	Character encoding	A 7-bit code to represent characters, such as letters and digits in computer systems	http://tools.ietf.org/html/rfc20
<i>ISO 8859-1 – ISO 8859-11; ISO 8859-13 – ISO 8859-16</i>	Character encoding	A number of 8-bit character encoding standards, extending the range of printable ASCII characters to support additional characters required in European languages	http://www.iso.org/iso/home.html

Standard	Content type	Description	More information
<i>Unicode</i>	Character encoding	A 16-bit code to represents characters, such as letters and digits in computer systems, created to overcome the limitations of 8-bit character sets and to form one universally usable encoding scheme for characters from both Western and non-Western languages and scripts	ISO/IEC 10646
<i>HTML-5</i>	Content structuring and encoding	A markup language used for structuring and presenting content (text and multimedia) for the Web: the fifth revision of the HTML standard	http://www.w3.org/TR/html5/ (HTML 5)
<i>TEI</i> Text Encoding Initiative	Document encoding	A consortium which collectively develops and maintains a standard and guidelines for the encoding and representation of texts in digital form	http://www.tei-c.org/index.xml
<i>BMP</i> BitMaP	Image	A proprietary format for monochrome and colour images	http://www.digicamsoft.com/bmp/bmp.html (Bitmap File Structure)
<i>DjVu</i>	Image	Designed mainly to store scanned images, especially those containing text and line drawings	http://djvu.org/docs/DjVu3Spec.djvu
<i>GIF</i> Graphical Interchange Format	Image	A proprietary format for monochrome or colour images, compressed using raster graphics to minimize file sizes	http://www.w3.org/Graphics/GIF/spec-gif89a.txt
<i>JPG or JPEG</i>	Image	A still-image compression	ISO/IEC 10918-1:1994

Standard	Content type	Description	More information
Joint Photographic Expert Group		algorithm	http://www.jpeg.org/jpeg2000/ (JPEG 2000)
<i>PNG</i> Portable Network Graphics	Image	An extensible format for the lossless, portable, well-compressed storage of raster images	ISO/IEC 15948:2003 (E)
<i>PSD</i> (Photoshop)	Image	Format for Photoshop documents	http://www.adobe.com/devnet-apps/photoshop/fileformats.html/#50577409_72092
<i>TGA, TARGA</i> Truevision Graphics Adapter; Truevision Advanced Raster Graphics Adapter	Image	A raster graphics file format	http://www.digitalpreservation.gov/formats/fdd/fdd000180.shtml
<i>TIFF</i> Tagged Image File Format	Image	A general purpose and device independent image data format, compatible with a wide range of scanners and image-processing applications	http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf (Version 6.0)
<i>SWF</i> Small Web Format	Interactive audio/video	For delivery of animated vector graphics, interactive audio and video	http://www.adobe.com/content/dam/Adobe/en/devnet/swf/pdf/swf_file_format_spec_v10.pdf (version 10)
<i>EPS</i> Encapsulated PostScript	Page layout and vector graphics	A page layout language based on PostScript and allowing PostScript encoded layouts to be “encapsulated” within other documents	http://partners.adobe.com/public/developer/en/ps/5002.EPSF_Spec.pdf (Version 3.0)
<i>ASF</i> Advanced Streaming Format	Streaming media	A proprietary format, designed to deliver, compressed, streaming video/audio content over the Internet	http://www.microsoft.com/windows/windowsmedia/forpros/format/asfspec.aspx (Revision 01.20.03)
<i>SGML</i> Standard Generalized	Text and document encoding	A language for structured data and document representation, on	ISO 8879:1986

Standard	Content type	Description	More information
Markup Language, ISO 8879		which XML and HTML were based	
XML eXtensible Markup Language	Text and document encoding	The most widely used structuring language for documents and data	http://www.w3.org/XML/
<i>DocBook</i>	Text encoding	A semantic markup language for technical documentation	http://www.docbook.org/
<i>EPUB-3</i>	Text and multimedia encoding	A distribution and interchange format standard for digital publications and documents. EPUB defines a means of representing, packaging and encoding structured and semantically enhanced Web content — including HTML5, CSS, SVG, images, and other resources — for distribution in a single-file format	http://idpf.org/epub/30
<i>HTML</i> HyperText Markup Language	Text encoding	A markup language based upon SGML with its original focus on the rendering (display) of documents rather than on document structure	http://www.w3.org/TR/html4/ (HTML 4)
<i>LaTeX</i>	Text encoding	A document preparation system and document markup language	http://www.latex-project.org/
<i>ODF</i> Open Document Format	Text encoding	An XML-based file format for spreadsheets, charts, presentations and word processing documents	http://www.opendocumentformat.org/
<i>PDF</i> Portable Document	Text encoding	A device-independent method of representing	ISO 32000-1:2008 – Document management -- Portable document

Standard	Content type	Description	More information
Format		document structure and layout, based on the PostScript page rendering standard	format -- Part 1: PDF 1.7
<i>RTF</i> Rich Text Format	Text encoding	A format for text and graphics interchange that can be used with different output devices, operating environments, and operating systems	http://msdn.microsoft.com/en-us/library/aa140277.aspx (Version 6.1)
<i>TeX</i>	Text encoding	A typesetting system, often used to typeset complex mathematical formulae	http://tug.org/
<i>Word</i>	Text encoding	A de facto, proprietary standard for document structuring and rendering	http://download.microsoft.com/download/2/4/8/24862317-78F0-4C4B-B355-C7B2C1D997DB/[MS-DOC].pdf
<i>SVG</i> Scalable Vector Graphics	Vector graphics	A language for describing two-dimensional graphics in XML	http://www.w3.org/TR/SVG11/ (Version 1.1)
<i>AVC/H.264</i> Advanced Video Coding	Video	A video compression format	http://www.mpegla.com/main/programs/AVC/Pages/Intro.aspx
<i>AVI</i> Audio Video Interleave	Video	An early and proprietary video format for PCs	http://msdn.microsoft.com/en-us/library/ms779636.aspx
<i>FLV</i> Flash Video Format	Video	A format used for the delivery of video over the Internet, used in conjunction with either separate "player" software or a web browser "plug-in"	http://download.macromedia.com/f4v/video_file_format_spec_v10_1.pdf (Version 10.1)
<i>Theora</i>	Video	A video compression format	http://www.theora.org/
<i>VP8</i>	Video	A Google-owned video compression	https://datatracker.ietf.org/doc/rfc6386

Standard	Content type	Description	More information
		format, but published with an open-source license	
<i>WMV</i> Windows Media Video	Video	Microsoft's proprietary competition to MPEG-4. Optimised to deliver video over the Web, particularly streaming	http://msdn.microsoft.com/en-us/library/ff819505(v=vs.85).aspx (WMV 9)
<i>MKV</i> Matroska	Video, audio	A multimedia container format for video, with audio and subtitles	http://www.matroska.org/technical/index.html
<i>MPEG or MPEG-1</i> Coding of Moving Pictures and Associated Audio for Digital Storage Media	Video, audio, TV	Designed as the equivalent of a video recorder format in the digital world	ISO/IEC 11172:1993 (Parts 1 to 5)
<i>MPEG-2</i>	Video, audio, TV	Improved version of MPEG-1 with improved encoding techniques	ISO/IEC 13818:2000 (Parts 1 to 11)
<i>MPEG-4</i>	Video, audio, TV	A high compression version of MPEG-2	ISO/IEC 14496 (Parts 1 to 10)
<i>MOV</i> Quicktime	Video, virtual reality/3D	Apple's proprietary video (and virtual reality) format and system	https://developer.apple.com/library/mac/documentation/QuickTime/QTFF/QTFFPreface/qtffPreface.html
<i>OBJ</i> Wavefront Object File	Virtual reality/3D	For the representation of 3D geometry	http://www.martinreddy.net/gfx/3d/OBJ.spec
<i>Quicktime VR</i>	Virtual reality/3D	Apple's proprietary virtual reality format and system for creation of content	Proprietary format
<i>VRML97</i> Virtual Reality Modelling Language	Virtual reality/3D	A format for describing static or dynamic 3D objects or "worlds", designed to be interactive and accessible over the Internet	ISO/IEC 14772-1:1997

Standard	Content type	Description	More information
X3D eXtensible 3D	Virtual reality/3D	An XML-based format expressing the functionality of VRML97	ISO/IEC FDIS 19775-1.2:2008

4.1.3 Standard Licenses and Methods of License Expression

It is becoming increasingly important to understand and communicate the license agreements and terms of usage associated with digital resources, whether these are “born digital” or are digitized representations of other cultural heritage artefacts. The Linked Heritage project investigated this topic and reported seven overall license types relevant here and broke these out further, for example describing at least four variants of the Creative Commons (CC) licenses in routine use.

The following table briefly summarizes the licenses mentioned and more detail can be found in Linked Heritage deliverables. The table also mentions a highly structured method for license expression, namely ONIX-PL; this latter is not a license in itself but rather a machine-readable framework for conveying licensing and usage terms, conditions and prohibitions.

License or standard	Description/purpose	More information
<i>BSD</i> Berkeley Software Distribution	One of a group of permissive software licenses, imposing minimal restrictions on the redistribution of the software covered by the license	http://en.wikipedia.org/wiki/BSD_licenses
<i>CC</i> Creative Commons	A series of public copyright licenses. Currently seven such license types exist	http://creativecommons.org/licenses/ See the website for more information on each license type: CC BY, CC BY-SA, CC BY-NC, CC BY-ND, CC BY-NC-SA , CC BY-NC-ND and CC0
<i>GNU FDL</i> GNU Free Documentation License	A “copyleft” licence designed for the free documentation of software, but which can be used for other text works	http://www.gnu.org/copyleft/fdl.html
<i>GNU GPL</i> GNU General Public License	A free software licence granting the licensee the right to change and redistribute the software free of the prohibitions of copyright law	http://www.gnu.org/copyleft/gpl.html
<i>ODbL</i> Open Database License	A license covering data in databases and allowing licensees, under certain conditions, to share create or adapt the database or its	http://opendatacommons.org/licenses/odbl/

License or standard	Description/purpose	More information
	content	
<i>ODC PDDL</i> Open Data Commons Public Domain Dedication and Licence	A license covering data in databases and allowing licensees, without attribution, to share create or adapt the database or its content	http://opendatacommons.org/licenses/pddl/1-0/
<i>ONIX-PL</i> ONIX for Publication Licenses	An XML format for the communication of license terms for digital publications in a structured and substantially encoded form	http://www.editeur.org/21/ONIX-PL/

4.1.4 Standards for Search, Retrieval, and Harvesting

The Athena project identified three protocols of particular importance in this area – OAI-PMH, SQL and Z39.50 – to which a fourth (SUSHI) is added here. Key points are listed in the following table and more detail is available in Athena deliverable D3.1. SPARQL, a query language for RDF, is a W3C standard for the Semantic Web.

Standard	Description/purpose	More information
<i>OAI-PMH</i> Open Archives Initiative Protocol for Metadata Harvesting	Provides access for harvesting programs to data stored in databases or repositories that cannot be harvested using “standard” http/html parsing	http://www.openarchives.org/OAI/openarchivesprotocol.htm
<i>SPARQL</i>	A query language for RDF.	http://www.w3.org/TR/rdf-sparql-query/
<i>SQL</i> Structured Query Language, ISO 9075	A query language for relational databases	http://en.wikipedia.org/wiki/SQL
<i>SUSHI</i> Standardized Usage Statistics Harvesting Initiative	A protocol defining an automated request and response model for the harvesting of electronic resource usage data	http://www.niso.org/workrooms/sushi/
<i>Z39.50</i> ISO 23950:1998	A protocol to implement search and retrieval in client-server applications	http://www.niso.org/standards/resources/Z39.50_Resources

4.1.5 Basic Standards

Linked Heritage also listed and explained a set of basic standards, whose use is widespread across digital preservation and other internet-enabled areas. They are listed in the table below.

Standard	Description/purpose	More information
<i>ORE</i> Object Reuse and Exchange	Used for the description and exchange of “aggregations” of web resources	http://www.openarchives.org/ore/1.0/toc.html
<i>OWL 2</i> Web Ontology Language	A Semantic Web computational logic-based language designed to represent rich and complex knowledge about: things; groups of things; relations between things	http://www.w3.org/TR/owl2-overview
<i>RDF</i> Resource Description Framework	A key standard for the expression, interchange and use of linked data	http://www.w3.org/RDF/
<i>URI</i> Uniform Resource Identifier	String of characters used to identify a name or a resource on the internet	http://tools.ietf.org/html/rfc1630
<i>URL</i> Uniform Resource Locator	A URI (string) that specifies where a resource is available and the mechanism for retrieving it	http://tools.ietf.org/html/rfc1738
<i>URN</i> Uniform Resource Name	A URI (string) acting as a persistent, location-independent, resource identifier, designed to make it easy to map to other namespaces	http://tools.ietf.org/html/rfc2141
<i>XML</i> eXtensible Markup Language	A simple text-based format for representing structured information	http://www.w3.org/XML/
<i>XSD</i> XML Schema Definition Language	Defines a class of XML documents in terms of a set of rules (structure and data types) to which a document must conform in order to be considered “valid”	http://www.w3.org/TR/xmlschema-0 (Primer) http://www.w3.org/TR/xmlschema-1 (Structures) http://www.w3.org/TR/xmlschema-2 (Datatypes)

See also [OCCI](#) (standards for cloud computing) and [CDMI](#) (standards for cloud storage interfaces).

4.2 BEST PRACTICE AND “HOW-TO” GUIDES

The Minerva, Athena, and Linked Heritage projects all offered considerable discussion and advice on best practice in a variety of areas. Additionally, “how-to” guides – such as those offered by SPECTRUM – provide practical guidance to both beginning and experienced practitioners.

For a selection of best practice advice on digital preservation in English:

Guide	Link
Kate Fernie, Giuliana De Francesco and David Dawson. <i>MINERVA Technical Guidelines for Digital Cultural Content Creation Programmes: Version 2.0, 2008, Sections 3.4 & 5.2.3</i>	http://www.minervaeurope.org/publications/MINERVA%20TG%202.0.pdf
Neil Beagrie, Maggie Jones and Digital Preservation Coalition. <i>Digital Preservation Handbook. 2001-2009.</i>	http://www.dpconline.org/advice/preservationhandbook
Canadian Heritage Information Network (CHIN). <i>Digital Preservation - Best Practices for Museums. 2013.</i>	http://www.pro.rcip-chin.gc.ca/contenu_numerique-digital_content/preservation_numerique-digital_preservation/index-eng.jsp
Priscilla Caplan. ‘Preservation metadata’ in <i>Curation Reference Manual.</i> Digital Curation Centre (DCC). 2006.	http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/preservation-metadata/preservation-metadata.pdf
David Holdsworth. ‘Preservation strategies’ in <i>Curation Reference Manual.</i> Digital Curation Centre (DCC). 2007	http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/preservation-strategies/preservation-strategies.pdf
Kevin Bradley (Editor). <i>Guidelines on the Production and Preservation of Digital Audio Objects. Second edition.</i> 2009. IASA Technical Committee.	http://www.iasa-web.org/tc04/audio-preservation
JISC Digital Media. <i>A Guide to Sustainability.</i> 2013.	http://www.jiscdigitalmedia.ac.uk/guide/a-guide-to-sustainability
The National Archives [UK]. <i>Digital Preservation – Guidance.</i> -2013.	http://www.nationalarchives.gov.uk/information-management/projects-and-work/guidance.htm
Library of Congress [USA]. <i>Digital Preservation.</i> -2013.	http://www.digitalpreservation.gov/
Digital Preservation Europe. <i>Briefing Papers.</i> 2007-2008.	http://www.digitalpreservationeurope.eu/publications/briefs/

4.3 PERSISTENT IDENTIFIERS FOR DIGITAL ARTEFACTS AND ASSOCIATED ENTITIES

This topic was covered at some length within Deliverable D2.2 of the Linked Heritage project. The report explained the role and importance of such identifiers before describing the primary candidate identifier types for use in the Cultural Heritage arena.

Fundamental features of three general digital identifier standards – URI (Universal Resource Identifier), URL (Universal Resource Locator) and URN (Universal Resource Name) – were presented first and they are included in the table in section 4.1.5 above. The report then described four types of service-associated digital identifier standards: PURL ((Persistent URL) & Handle System), DOI (Digital Object Identifier), OpenURL and ARK (Archival Resource Key). Summary information on these four is presented in the table below.

Identifier	Description/purpose	More information
<i>ARK</i> Archival Resource Key	A URL scheme which can identify both physical and digital objects	http://en.wikipedia.org/wiki/Archival_Resource_Key
<i>DOI</i> Digital Object Identifier	A stored and maintained character string used to uniquely identify any kind of entity, physical, digital or abstract	http://www.doi.org/ The DOI is associated with a prescribed set of metadata. Used in conjunction with the Handle system, the DOI provides an infrastructure for the persistent identification and location of digital resources
<i>OpenURL</i>	A URL with embedded metadata (used by resolver services) to more easily find a resource	http://en.wikipedia.org/wiki/OpenURL
<i>PURL</i> Persistent URL	A URL pointing to a resolver (e.g., a handle) which directs to a current URL	http://purl.oclc.org/docs/help.html#overview

Arguably, the service-associated and maintained identifiers are likely to offer more comprehensive features to Cultural Heritage institutions managing digitized resources, but issues relating to both cost and policy have militated against the widespread adoption of such identifiers in this area.

The example of DOIs may be illustrative here. Although the utility of this identifier system is clear, uptake to date in the DCH environment has been very low. The costs associated with DOI registration may represent one significant barrier. But equally, the absence of a DOI Registration Agency specifically aligned with the DCH community – understanding cultural heritage requirements and able to design registration metadata that is meaningful to that community – may represent an even greater impediment. (Compare for instance the near-ubiquity of DOIs in the area of scientific, technical and medical journals, which are serviced by the Crossref registration agency, itself originally established by the STM publishing industry itself.)

Finally, Linked Heritage also described a number of the proprietary or institution-specific identifier strings currently in use.

4.4 STANDARDS USED IN DIFFERENT COUNTRIES AND INSTITUTIONS

4.4.1 Standards used by the E-CSG

The DCH-RP e-Culture Science Gateway ([eCSG](#)) is a standard-based web 2.0 demonstrative platform. The framework for Science Gateways fully web-based and adopts official worldwide standards and protocols, through their most common implementations. These are:

- The [JSR 168](#) and [JSR 286](#) standards (also known as "portlet 1.0" and "portlet 2.0" standards);
- The [OASIS Security Assertion Markup Language](#) (SAML) standard and its [Shibboleth](#) and [SimpleSAMLphp](#) implementations;
- The Lightweight Direct Access Protocol, and its [OpenLDAP](#) implementation
- The [Cryptographic Token Interface Standard](#) (PKCS#11) standard and its Cryptoki implementation;
- The [Open Grid Forum](#) (OGF) [Simple API for Grid Applications](#) (SAGA) standard and its [JSAGA](#) implementation.

4.4.2 Standards used by the ICCU

The Istituto Centrale per il Catalogo Unico ([ICCU](#)) use a metadata standard [MAG](#) (used instead of METS). MAG intends to promote the collection of a "least common" set for management metadata (in particular, technical metadata about digitalization of images).

4.4.3 Standards used by the KIK-IRPA

The Royal Institute for Cultural Heritage ([KIK-IRPA](#)) has defined a preservation model for their photo library, with more than one million photographs, that is based on the OAIS model.

5 ISSUES AND CHALLENGES; CONCLUSIONS

As implied from the sizes of the tables in Chapter 4, one of the challenges for the DCH community is to choose among the vast number of standards that are already present. This may be especially hard for small DCH institutions without much experience and/or resources. And although there are many “How-To” guides about best practices, the relatively large amount of these guides may in itself be a discouraging.

A related challenge for all who work with digital content is the rapid change in technology. You may have chosen a *de facto* standard that seems stable and widely disseminated, but within a few years that standard may have been abandoned for a new one. Since software developers may stop supporting old standards and/or standards that are seldom used, this is not an issue that can be ignored.

Even if technical e-infrastructures can be built for dissemination of the DCH, there are non-technical issues that also have to be resolved. One such issue is the judicial differences between countries with respect to using e-infrastructures outside your own country for storage of information (especially if the data’s use is restricted by copyright, other rights, or even is classified). Another issue is the actual trust it requires to let go of the immediate control of your DCH information, by storing it not only outside your own institution, but possibly even in another country.

See also section 3.2 (*Key Topics for the Cultural Heritage Area*) for other important issues.

Conclusions

A large amount of groundwork has already been done in sectors other than DCH, but much more needs still to be done for this work to be of essential help for the DCH community. For example, many of the standards, guides, and tools now present would need to be more user-friendly in order to be understandable for non-technical personnel. Furthermore, practical tests made within DCH-RP project have shown that already developed e-infrastructures must be modified and/or improved in order to provide a “pan-European” solution for the DCH community.