

Project no. 269977

APARSEN
Alliance for Permanent Access to the Records of Science
Network

Instrument: Network of Excellence

Thematic Priority: ICT 6-4.1 – Digital Libraries and Digital Preservation

**D13.2 REPORT ABOUT
STANDARDISATION ACTIVITIES:
PROGRESS TO YEAR 3**

Document identifier:	APARSEN-REP-D13_2-01-1_5
Due Date:	2013-12-31
Submission Date:	2014-01-17
Work package:	WP13
Partners:	APA, CERN, STM, FTK, CSC, DNB, ESA, ICT, FORTH, GLOBIT, LTU, UNITN, SBA
WP Lead Partner:	InConTec (ICT)
Document status	Released
URN	urn:nbn:de:101-2014051667

Abstract:

This intermediate report gives an overview about the database, which has been implemented to store the preservation-related standards and projects which have been collected so far in the project and laid out in deliverable D13.1. The document describes the data model and the description of all fields of the database as it is implemented now.

The database with its information will be published in February 2014 on the public website as part of the Virtual Centre of Excellence (VCoE).

Delivery Type Report

Author(s) Hans-Ulrich Heidbrink, Andreas Hundsdörfer (ICT); Holger Brocks (FTK)

Approval Summary David Giarretta/Simon Lambert

Keyword List Standardisation, Preservation, LOTAR

Availability ☒ PUBLIC

Document Status Sheet

Issue	Date	Comment	Author
0.1	2012-12-22	Initial draft	Andreas Hundsdörfer (ICT)
0.2	2013-01-30	Added installation guidelines	Andreas Hundsdörfer (ICT)
0.3	2013-02-10	Initial feedback	Hans-Ulrich Heidbrink (ICT), Holger Brocks (FTK)
1.0	2013-02-	Final version for internal QA	Andreas Hundsdörfer (ICT)
1.1	2013-02-28	Incorporated feedback from internal QA (David Giarretta, Christoph Bruch)	Andreas Hundsdörfer (ICT)
1.2	2013-11-28	Extended structure and content	Hans-Ulrich (ICT), Holger Brocks (FTK)
1.3	2013-12-18	Extended content	Hans-Ulrich (ICT), Holger Brocks (FTK)
1.4	2013-12-23	Extended content	Andreas Hundsdörfer (ICT), Holger Brocks (FTK)
1.5	2014-01-15	Internal review / External review (partially)	Barbara Bazzanella (UNITN) / Karen Colbron (project Presto4U, part of PrestoCentre)

Project information

Project acronym:	APARSEN
Project full title:	Alliance for Permanent Access to the Records of Science Network
Proposal/Contract no.:	269977

Project coordinator: Simon Lambert/David Giaretta

Address:	STFC, Rutherford Appleton Laboratory Chilton, Didcot, Oxon OX11 0QX, UK
Phone:	+44 1235 446235
Fax:	+44 1235 446362
Mobile:	+44 (0) 7770326304
E-mail:	simon.lambert@stfc.ac.uk / david.giaretta@stfc.ac.uk

CONTENT

1	INTRODUCTION	7
1.1	PURPOSE AND SCOPE.....	7
1.2	THE ROLE OF A VCoE FOR STANDARDIZATION	8
1.3	OBJECTIVES.....	8
1.4	METHODOLOGY AND ALIGNMENT.....	9
1.5	TIMETABLE.....	9
2	HARMONIZATION AND ALIGNMENT.....	11
2.1	ALIGNMENT WITH OTHER WORK PACKAGES	11
2.2	COLLABORATION WITH PRESTO4U.....	11
2.3	REVISED DB SCHEMA (UPDATE/NEU)	12
3	ONLINE STANDARDS REGISTRY	22
3.1	RELATED INITIATIVES	22
3.1.1	<i>Library of Congress – Tools Showcase</i>	<i>22</i>
3.1.2	<i>Library of Congress – Sustainability of Digital Formats.....</i>	<i>23</i>
3.1.3	<i>The National Archives – PRONOM.....</i>	<i>23</i>
3.1.4	<i>nestor – Standardisation</i>	<i>24</i>
3.1.5	<i>DPC – Standards and Best Practice Guidelines, File Formats and Standards.....</i>	<i>25</i>
3.1.6	<i>SAA – Standards Portal.....</i>	<i>26</i>
3.1.7	<i>PrestoCentre – PrestoCentre Library</i>	<i>27</i>
3.1.8	<i>DCC – Digital Curation Standards.....</i>	<i>28</i>
3.1.9	<i>The Open Group – Standards Information Base</i>	<i>30</i>
3.2	IMPLEMENTATION	31
3.2.1	<i>Discussion</i>	<i>31</i>
3.2.2	<i>Specification</i>	<i>31</i>
3.2.3	<i>Target Groups</i>	<i>32</i>
3.2.4	<i>Portal Functions.....</i>	<i>32</i>
3.2.5	<i>Implementation Environment</i>	<i>33</i>
3.3	EMBEDDING IN VCoE	33
4	PRESERVATION SOLUTIONS (SOFTWARE APPLICATIONS).....	34
4.1.1	<i>Commercial Sector.....</i>	<i>34</i>
4.1.2	<i>Open Source Sector.....</i>	<i>34</i>
5	PERSISTENT IDENTIFIERS, FILE FORMATS AND OTHERS	34
5.1	PERSISTENT IDENTIFIERS	34
5.2	COMMON FILE FORMAT STANDARDS USED IN PRESERVATION SYSTEMS.....	35
6	CONTEXTUALIZED ACCESS AND USER GUIDANCE.....	36
6.1	CHALLENGES AND SCENARIOS FROM RTD WPs, USE CASES	36
6.2	CONTEXTUALIZED GUIDANCE	37
6.3	ALIGNMENT WITH SCIDIP-ES/APA ORCHESTRATION SERVICE (OPTIONAL).....	38
7	SPECIFICATION OF GAP ANALYSIS METHODOLOGY.....	39
7.1	FRAMEWORK FOR ELICITING GAPS IN TOPICAL COVERAGE	39
7.2	CUSTOMIZATION TO THE ENGINEERING DOMAIN	40
7.3	APPLICATION ON SELECTED COMMUNITIES	41
8	TOWARDS A ROADMAP FOR COMMON STANDARDS FOR INDUSTRIAL DESIGN AND ENGINEERING.....	41
8.1	METHODOLOGY.....	41
8.2	DIALOGUE WITH EXTERNAL STAKEHOLDERS (STEP)	44
	LIST OF FIGURES	45
	LIST OF ABBREVIATIONS.....	45

REFERENCES49

1 INTRODUCTION

The intention in APARSEN is to give an overview of existing common standards within digital preservation and in which contexts these are applicable. Firstly, common standards are surveyed and commented. Next, a number of user categories are listed followed by a decision tree providing help to find which standards are useful or applicable for different user categories. Following, there is a visualization of common standards' coverage and overlaps, and finally a checklist providing guidance for selection of common standards and procurement of digital preservation technology.

The document starts with an introductory part (section 1) highlighting the role of the Virtual Center of Excellence, reiterating the objective of WP13, explaining the methodology on how this document will be expanded over the remaining project duration and finally in this section the timetable for all 4 periods.

In version 1 of this deliverable covering Year 1 of the project section 2 gave an extensive overview about common digital preservation standards, recommendations and initiatives. This overview has already been removed in the last version of this document. Instead section 2 provides information about the alignment with other work packages, the collaboration with the EU-project "Presto4U" and an overview of the widely extended database, which has been implemented in this period, to support the Virtual Center of Excellence in period 4 of the project.

The new chapter 3 briefly describes some related activities of other institutions and organisations, who have tried to make standards available on their respective webpages. Persistent identifiers and file formats are presented unchanged in section 5.

The new chapters 6 presents some use cases, which have to be implemented in year 4 as added value for the VCoE, while section 7 discusses an approach for the planned gap analysis capability and section 8 the initial plans for the roadmap of common standards for the design & engineering domain.

1.1 PURPOSE AND SCOPE

During the APARSEN project this intermediate deliverable will be extended and improved to describe, specify and establish the tasks as they are:

- Collection of preservation standards used in the different communities
- Collection of preservation projects with standardisation activities and their respective results
- Identification of communities applying preservation standards in their repositories
- Assessment of the scope of preservation standards related to the OAIS model
- Build-up of a common standards repository with community oriented search criteria
- Including information about current productive use of the standards in the related communities their purpose of use and related SWOT analysis based positioning
- Certification criteria, procedures and services offered by any member of the community
- Cross-community synergy analysis
- Establishment of a decision tree for the applicability and perspective use of current standards
- Community related gap analysis
- Requirements for missing standards to fill the existing gaps

The following document is an intermediate snapshot of the initial investigation and cannot be seen as complete. Each year this document will be updated and accompanied by structure, describing meta data, links and content that allows to have a common view on established preservation standards in the different communities, feedback from the users, selection criteria along a given decision tree and search capabilities to identify similarities or alternate use models. The final result will give access to the standards repository, clear out existing gaps in the preservation process of the involved communities and will give recommendation and requirements to fill those gaps.

1.2 THE ROLE OF A VCOE FOR STANDARDIZATION

The establishment of a community spanning database can only be performed if all aspects of preservation are concentrated in one single place. Most of the standard related information was established in various research projects and standardization tasks but there is no organization established that combines all communities and networks that have contributed. Therefore the VCoE (Virtual Centre of Excellence) as an open network of networks combines all the standard related knowledge, the applicable validation and certification procedures, represents a common mind for services related to preservation and maintains the infrastructure that APARSEN members are generating. Additionally this VCoE will be open to associations and standardisation bodies that are not yet part of the APARSEN network of excellence but are able and willing to contribute to a common data base containing all known references, links and established open documents related to preservation standards. For this it is mandatory that there are no hurdles or fees to take part in the VCoE activities, also for organisations not participating in the APARSEN network. The VCoE should be open for other associations or potentially contributing standardization bodies and establish in this way an open accessible exchange and information platform for everybody. Therefore the purpose of the standards database is relevant not only to raise awareness about relevant standardization activities within the VCoE but also to facilitate the emergence of new standards where gaps are identified.

The required business models will be evaluated and established in APARSEN work package "Business cases" (WP36). The VCoE is described in detail in the deliverable "VCoE Business Plan".

1.3 OBJECTIVES

The following objectives are taken from the WP13 description in the Description of Work (DoW) document and covers already the DoW-update performed at the end of period 2/beginning of period 3:

To identify common standards - either existing or required new ones - which will enhance the accessibility of information via the interoperability of the systems managed by the partners and the community at large. In the case of new standards this will identify possible routes for standardisation, addressing the cost benefit trade off of the alternatives (see for example international standards versus community standards).

Task 1310 Analysis of current standards

This task will produce an analysis of standards in use for their digital preservation activities by partners, communities and identification of commonalities and differences.

We will take advantage of our links, funded outside the NoE, to on-going global discipline-specific initiatives such as:

- the Group for Earth Observation's Standards and Interoperability Forum (SIF) and in particular to the European branch of it European SIF.*
- Consultative Committee for Space Data Systems - CCSDS which has adopted a reference model for Open Archival Information Systems*
- The ESA HMA project which has standardised via the Open Geospatial Consortium ground segment interfaces, including authorisation services, online data access services*
- The Open Archive Initiative which promotes the interoperability standards for content access*
- The Open Search mass market approach to search sharing and information access*
- SCIDIP-ES for science data preservation*
- Requirements for the preservation of Design and Engineering data resulting from the on-going activities in the SHAMAN project and other European initiatives (e.g. LOTAR (Long Term Archiving and Retrieval) – EN9300; STEP (Standard for the Exchange of Product Model Data))*

Other disciplines represented in the consortium have similar links.

Task 1320 Identification of new standards required

This task will identify areas where common standards would be of benefit. The task will rely on the methodology for interoperability analysis at level of system of systems developed within the GIGAS Support Action with the participation of GEO, GMES initiatives and the INSPIRE directive.

Another discipline-specific study we can take advantage of is based on the High-Energy Physics community which, through the DPHEP initiative, is exploring technical issues in data preservation, and aims to define standards for the field where none exist yet.

We have links to similar discipline-specific studies for example in the domain of enhanced publications.

An important area where we can have an impact is engineering. In order to get industrial acceptance preservation systems need to be structured and built that allow exchange of engineering data in heterogeneous organisations with different Product Lifecycle Management systems and services as data source for preservation activities. Common standards and archiving policies are essential for the wide acceptance and interoperability. Existing design and manufacturing standards need to be extended to allow preservation in all stages of the design and release process.

This activity includes the analysis of on-going standardisation in the LOTAR consortium performed with participation of industrial consortia partners and in organizations like the ProSTEP iViP organization that represents aerospace, automotive and high-tech industries as well as international PLM and CAD tool vendors; the VDA association (Association of the Automotive industries) and the AIA (Aerospace Industries Association) These organizations are key for recommendations to extend the current ISO standards for 10303 (STEP) and 14721 (OAIS) and the CAD interoperability standards in the CAx-IF workgroups as well as to support standardisation of PLM processes in the related PLM-Services group. Outside these European industrial organisations results related to engineering preservation coming out of the SHAMAN project and national projects like the JISC-funded SCARP project will be included.

Concerning LOTAR, a roadmap will be developed for standardisation of long living engineered parts ranging from formats used in PLM applications to formats and processes for long term preservation. We will establish liaisons with standardization bodies and stakeholders, with a view to information sharing and consensus building for the agreement on the roadmap between industry, vendors and research communities

1.4 METHODOLOGY AND ALIGNMENT

The D13.1 document will evolve during 4 years, by incremental additions and improvements. The methodology we will use is based on mainly qualitative methods. The initial planning is to firstly find out what is state-of-art, and use that as well as recent developments to guide/support organizations in their process of getting an overview of as well as later selection of digital preservation standards. Below are the initially planned steps in the methodology to reach the research objectives:

- Literature review and survey of state-of-art + analysis
- Gap analysis
- Case studies
- Development of criteria for selection of standard using a decision tree

1.5 TIMETABLE

The work package 13 started in month 4 and has a duration period until month 48. For each finalized project year an interim report gets generated followed by the final report in month 48. Therefore the entire work packed activities have been split in four sections with specific areas of research and investigations.

Year 1- Review of existing standards, recommendations and projects with standardization activities

In this period desktop based research with the focus on standards had been performed as described in this intermediate deliverable D13.1. The research activities got the focus on generic standards, communities and EU and non EU funded projects or initiatives from preservation related to processes and file formats. Most time had been spent to review the variety of documents, recommendations and initiatives. This ground work which still makes no claim to be complete is the basis for the investigation of the second year period.

Year 2 – Adding more standards especially in the engineering domain; moving the investigation results from document into a database approach

In this reporting period (Y2) we have investigated additional standards, projects etc., which have not been looked at in period 1 and also additional domains (such as publishing, open source and commercial solutions). These activities will continue throughout the project. Significant effort has been spent on standards and project in the design and engineering domain.

The project also started to set-up a taxonomy in order to structure the register of standards identified in the community at large. The taxonomy design is guided by the goal to provide an overview on the continuously growing index and to support practitioners, researchers and students in retrieving standards of relevance to their task. The taxonomy features multiple schemes. Categories are derived from or are mapped onto international classification schemes or common practice in the community at large, respectively.

Furthermore a “database”-like approach has been established and we have already started to load the standards documented in D13.1 into that database, to make that information available in the VCoE and as the preparation for period 3, namely to work out a solution for a user to search e.g. over a kind of “decision tree”).

Year 3 - Adding more standards; migration to MySQL; database extensions; start of user interface development; decision tree;

With the results of the year two investigations and the migration of the findings of the first two years into the database, we migrated the database to MySQL. In September 2013 we began to collaborate with the project Presto4U, part of PrestoCentre, which also has a work package on standardization of audio/visual information. As a result both parties developed an extended data model to fit both parties' needs. Specification of the user interface has been started and it was intended to generate decision trees which summarize the required standards for each domain in each step in the preservation activity following the OAIS model approach. Identifying the domain and the related preservation problem related standards out of a link enabled repository will help the user to find the needs or applicable standards and detailed standard descriptions to get compliant with the already established preservation standards or to gain know-how about communities or projects which are related to the requirement given by the users and collaborating APARSEN partners. This work has not been finished, so that a first version of such decision tree will be available for the soft launch of the VCoE scheduled for end of February 2014.

Year 4 Recommender service and GAP analysis

First public version of the database is scheduled to be available for the soft launch of the VCoE end of February 2014. In the last year also a kind of recommender service will be improved and areas lacking in preservation oriented standards and recommendations for future standardization activities related to the result of a GAP analysis that will be performed during this last period, will be documented. This is the main part of the final deliverable and a prototype version will be made available for the hard launch of the VCoE scheduled for October 2014. More content will also be added by the partners.

2 HARMONIZATION AND ALIGNMENT

As described in previous deliverables, WP13 is embedded in a project context. Hence there are natural synergies across, but also the need for some alignment with other work packages. Addressing the recommendations from the previous review, a cooperation agreement with Presto4U was taken forward. As an immediate result of the cross-project discussions, the WP13 database schema was updated to ensure interoperability w.r.t. the core standards concepts found both, in APARSEN and Presto4U.

2.1 ALIGNMENT WITH OTHER WORK PACKAGES

Common standards in Digital Preservation are naturally linked to Common Tools (WP16) and Preservation Services (WP21) implementing and/or supporting them. Specifically, information about standards associated with digital data can be considered as Representation Information (aggregated/linked to form a Representation Information Network) according to the OAIS reference model [OAIS] which is handled in WP25 Interoperability and intelligibility. Reference Information (WP22 Identifiers and citability) identifies and describes standardized mechanisms for unambiguously identifying digital data. Authenticity and provenance (WP24) also rely on standards for formally representing its required evidence. In particular in D25.1 (Interoperability Objectives and Approaches) you can find an extensive analysis on interoperability initiatives and projects and on interoperability standards, services and solutions which are to be included in the database; for further reading on this topic see D25.1 sections 2.4 and 5 [APARSEN-REP-D25_1-01-1_7]. In general, all APARSEN WPs dealing with relevant standards have contributed input to the WP13 standards database and will further assist in provide adequate classifications for supporting guides access and evidence about usage/implementation in Y4.

Specifically, WP13 and WP16/21 have agreed upon using MySQL as common underlying relational database management system (RDBMS) for managing their information assets. While the long-term goal for the APARSEN VCoE is to provide unified access to Representation Information through a so-called RepInfo Registry (taken over from the SCIDIP-ES project), it was decided to keep the current databases and corresponding schemas modular and independent. However, feasible federation/mediation options will be explored and discussed in Y4.

As stated previously, the WP13 standards database represents a major constituent of the APARSEN VCoE. Interim findings/results and their implications have been continuously fed into the WP11 discussions about the Integrating to a Common vision of digital preservation (Task 1120) and the Creation of Virtual Centre of Excellence (Task 1110). Consolidated WP13 output is and will be disseminated through the Sustainable uptake stream (especially WP44) and gathered input/feedback is fed into work package discussions.

2.2 COLLABORATION WITH PRESTO4U

Presto4U [Presto4U] is a coordination action aiming at broadening the scope of PrestoCentre by establishing nine specific communities of practice in AV digital preservation.

Within the work package on Technology Transfer from Preservation Research, Presto4U prepares a Standards Register for its communities of practice. To exploit potential synergies and avoid duplication of efforts a dialogue between both projects about setting up a cooperation agreement was established¹. The formal document is currently being finalized and expected to be signed by representatives from both projects in early 2014.

¹ thus acknowledging and implementing Recommendations 4 “...discussion with other communities...” and 5 “...contributions from other communities...” from the previous review report.

While negotiating the formal aspects of collaboration, the APARSEN WP13 deliverables, sample database and associated schema were shared and discussed in detail w.r.t. their consistency with ongoing work in Presto4U. Most of the fields turned out to be quite well aligned, mismatches in labels/terminology were harmonized as a result of those discussions. As a result, the “core” standard fields from both projects’ database schemas are aligned and interoperable. However, the actual implementations and associated UI forms remain independent, as disparate communities and project requirements need to be accommodated.

It was agreed that both projects will maintain persistent identifiers for the standards registered respectively, which would allow for cross-referencing and federation.

In the next section, the updated/refactored database model will be presented. The data contained within the OO Base database has been mapped and migrated.

2.3 REVISED DB SCHEMA (UPDATE/NEU)

Figure 1 below is a representation (ER model) of the database, as it was at the end of year with the main goal to have a simple and reusable approach for the collection of standards and projects.

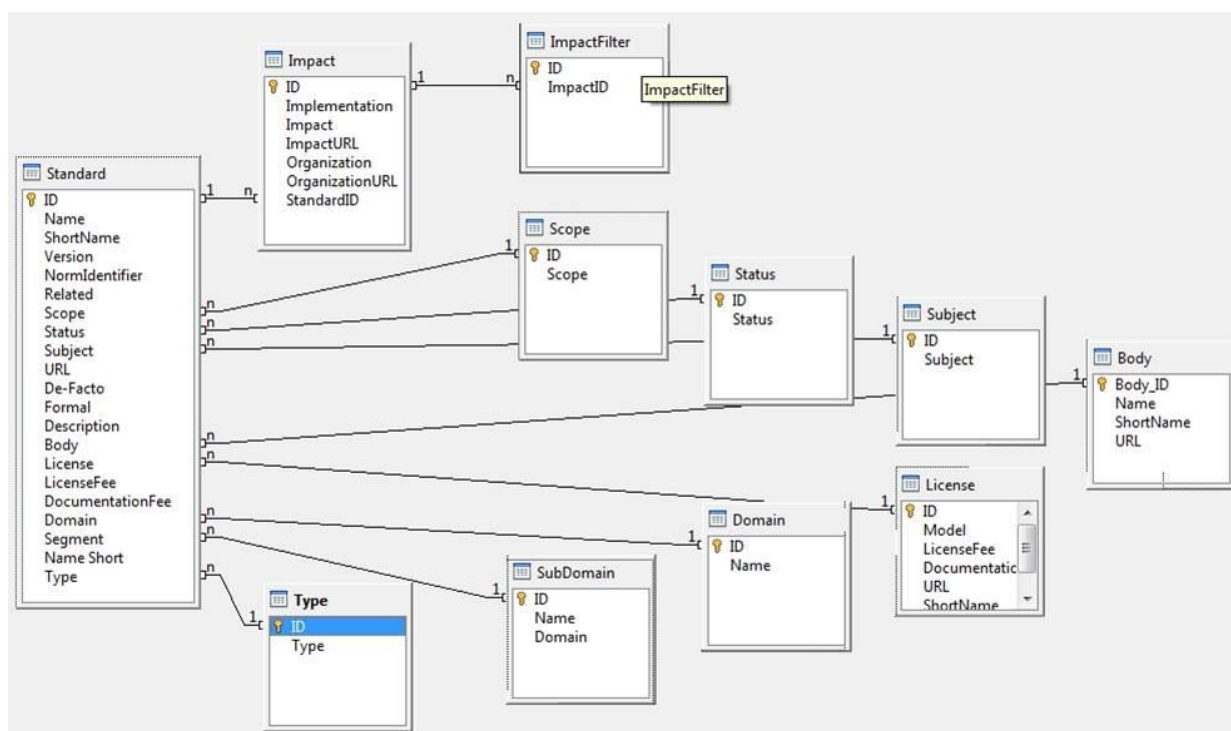


Figure 1: Year-2 database scheme in OpenOffice Base

Figure 2 below represents the data model at the end of year 3 after having it extended throughout the year and aligned with the needs and ideas coming out of extensive discussions with Presto4U.

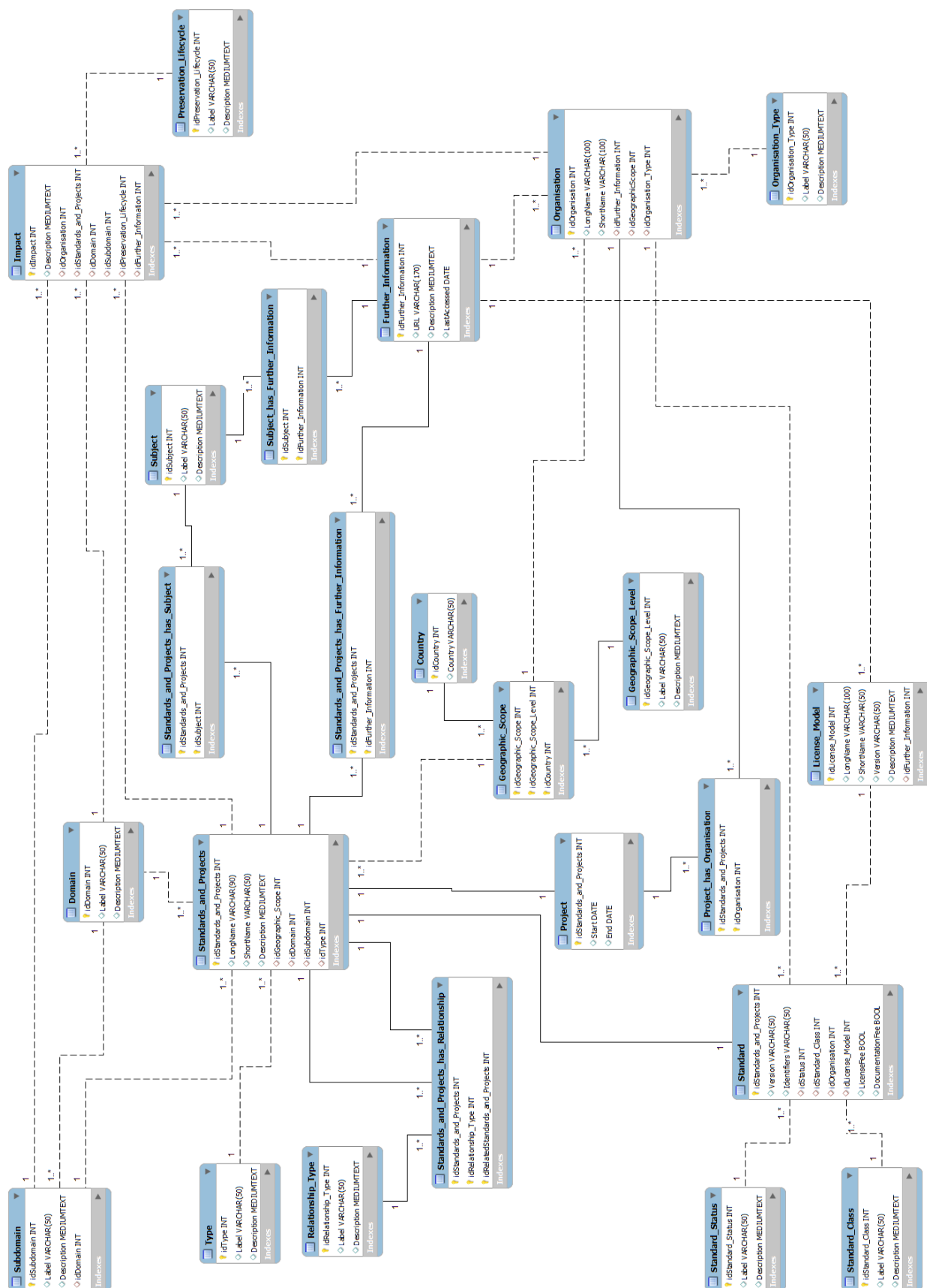


Figure 2: New database in MySQL for VCoE

Standards_and_Projects	
<i>Description</i>	<p>„Superclass“ for <i>Standard</i> and <i>Project</i></p> <ul style="list-style-type: none"> • Combines common fields • Eases the definition of relationships via <i>Relationship_Types</i>, because ID is assigned in Superclass. Prevents the problem, that a <i>Standard</i> and a <i>Project</i> have the same ID
<i>Columns</i>	<ul style="list-style-type: none"> • idStandards_and_Projects: primary key , generated automatically • Long Name: complete name of a <i>Project/Standards</i> • Short Name: Abbreviation of the name • Description: Description of the <i>Standard/Project</i> • idGeographic_Scope: references <i>Geographic_Scope</i> • idDomain: references <i>Domain</i> • idSubdomain: references <i>Subdomain</i> • idType: references <i>Type</i>
<i>N:m Relationships</i>	<ul style="list-style-type: none"> • Standards_and_Projects_has_Relationship • Standards_and_Projects_has_Further_Information • Standards_and_Projects_has_Subject
<i>Entries</i>	

Standards_and_Projects_has_Relationship	
<i>Description</i>	Enables to assign entries of the table <i>Standards_and_Projects</i> several cognate entries of table <i>Standards_and_Projects</i> and to characterize the kind of relationship between each other
<i>Columns</i>	<ul style="list-style-type: none"> • idStandards_and_Projects: references <i>Standards_and_Projects</i> • idRelationship_Type: references <i>Relationship_Type</i> • idRelatedStandards_and_Projects: references <i>Standards_and_Projects</i>
<i>N:m Relationships</i>	
<i>Entries</i>	

Standards_and_Projects_has_Further_Information	
<i>Description</i>	Enables to assign entries of the table <i>Standards_and_Projects</i> several cognate entries of table <i>Further_Information</i>
<i>Columns</i>	<ul style="list-style-type: none"> • idStandards_and_Projects: references <i>Standards_and_Projects</i> • idFurther_Information: references <i>Further_Information</i>
<i>N:m Relationships</i>	
<i>Entries</i>	

Standards_and_Projects_has_Subject	
<i>Description</i>	Enables to assign entries of the table <i>Standards_and_Projects</i> several cognate entries of table <i>Subject</i>
<i>Columns</i>	<ul style="list-style-type: none"> • idStandards_and_Projects: references <i>Standards_and_Projects</i> • idSubject: references <i>Subject</i>
<i>N:m Relationships</i>	
<i>Entries</i>	

Domain	
<i>Description</i>	Rough areas, to which a Standard/Project can be assigned
<i>Columns</i>	<ul style="list-style-type: none"> • idDomain: primary key, created automatically Label: denomination, e.g. Earth Science as generic term for Vulcanology • Description: detailed description of an entry
<i>N:m Relationships</i>	
<i>Entries</i>	

Subdomain	
<i>Description</i>	Details the area to which a Standard/Project can be assigned
<i>Columns</i>	<ul style="list-style-type: none"> • idSubdomain: primary key, created automatically • Label: denomination, e.g. Vulcanology • Description: detailed description of an entry • idDomain: Reference to <i>Domain</i>
<i>N:m Relationships</i>	
<i>Entries</i>	

Subject	
<i>Description</i>	Offers different thematical terms. Several entries in <i>Subject</i> can be assigned to one entry in <i>Standards_and_Projects</i>
<i>Columns</i>	<ul style="list-style-type: none"> • idSubject: : primary key, created automatically • Label: denomination, e.g. Metadata Dictionary or File Format • Description: detailed description of an entry
<i>N:m Relationships</i>	Subject_has_Further_Information
<i>Entries</i>	<ul style="list-style-type: none"> • Communication Protocol • File Format • Process • Certification • Form Factor • Management System • Archiving Format • Ingest (Planning) • Archive • Access • Security • Authenticity • Persistence • Identity • Common Model • Persistent Identifiers • Trust • Other

Subject_has_Further_Information	
<i>Description</i>	Enables to assign entries of the table <i>Subject</i> several cognate entries of table <i>Further_Information</i>
<i>Columns</i>	<ul style="list-style-type: none"> • idSubject: references <i>Subject</i> • idFurther_Information: references <i>Further_Information</i>
<i>N:m Relationships</i>	
<i>Entries</i>	

Type	
<i>Description</i>	Enables the distinction/categorization of the entries in <i>Standards_and_Projects</i> .
<i>Columns</i>	<ul style="list-style-type: none"> • idType: primary key, created automatically • Label: Name of the different categories • Description: detailed description of the respective category
<i>N:m Relationships</i>	
<i>Entries</i>	<ul style="list-style-type: none"> • Standard • Project • External Source • Other

Relationship_Type	
<i>Description</i>	Serves the definition of relationships between Standards/Project. For example, that a Standard A uses a Standard B. A Standard/Project can have more than one relationship to other Projects/Standards
<i>Columns</i>	<ul style="list-style-type: none"> • idRelationships: primary key, created automatically • Label: denomination, e.g. Is Used By • Description: detailed description of an entry
<i>N:m Relationships</i>	
<i>Entries</i>	<ul style="list-style-type: none"> • Not specified • Is earlier version of • Is newer version of • Uses • Is used by

Standard	
<i>Description</i>	„Sub-class“ of <i>Standards_and_Project</i> , which contains entries, which apply only to Standards, not Projects
<i>Columns</i>	<ul style="list-style-type: none"> • idStandards_and_Projects: references the corresponding Entry in <i>Standards_and_Projects</i> • Version: Version of Standard, e.g. 1.9.1 • Identifiers: e.g. Z39.87-2006 • idStandard_Status: references <i>Standard_Status</i> • idStandard_Class: references <i>Standard_Class</i> • idOrganisation: references <i>Organisation</i> • idLicense_Model: references <i>License_Model</i> • LicenseFee: indicates, whether a fee will be collected or not • DocumentationFee: indicates, whether a fee will be collected or not

<i>N:m Relationships</i>	
<i>Entries</i>	

Standards_Status	
<i>Description</i>	It can be specified, which status a Standard has, e.g. whether it has already been published
<i>Columns</i>	<ul style="list-style-type: none"> • idStandard_Status: primary key, created automatically • Label: denomination, e.g. Proposal, Recommendation, White Paper, Requirement, Submission, Published ,... • Description: detailed description of an entry
<i>N:m Relationships</i>	
<i>Entries</i>	

Standards_Class	
<i>Description</i>	Enables to categorise a standard in more detail, e.g. whether it is a formal or a de-facto standard
<i>Columns</i>	<ul style="list-style-type: none"> • idStandard_Class: primary key, created automatically • Label: denomination, e.g. De-Facto, Formal, ... • Description: detailed description of an entry
<i>N:m Relationships</i>	
<i>Entries</i>	<ul style="list-style-type: none"> • DeFacto • Formal • Other

License_Model	
<i>Description</i>	Description of the conditions for the licensing of a Standards
<i>Columns</i>	<ul style="list-style-type: none"> • idLicense: primary key, created automatically • Long Name: complete name • Short Name: Abbreviation • Version • Description: detailed description of an entry • idFurther_Information: references <i>Further_Information</i>
<i>N:m Relationships</i>	
<i>Entries</i>	

Project	
<i>Description</i>	„Sub-class“ of <i>Standards_and_Project</i> , which contains entries, which apply only to Projects, not Standards
<i>Columns</i>	<ul style="list-style-type: none"> • idStandards_and_Projects: references the respective entry in <i>Standards_and_Projects</i> • Start: Start date of the Project • End: End date of the Project
<i>N:m Relationships</i>	<i>Project_has_Organisation</i>
<i>Entries</i>	

Project_has_Organisation	
<i>Description</i>	Enables to assign entries of the table <i>Standards_and_Projects</i> several cognate entries of table <i>Subject</i>
<i>Columns</i>	<ul style="list-style-type: none"> • idStandards_and_Projects: references <i>Projects</i> • idOrganisation: references <i>Organisation</i>
<i>N:m Relationships</i>	
<i>Entries</i>	

Impact	
<i>Description</i>	Describes the implementation of Standards/Projects
<i>Columns</i>	<ul style="list-style-type: none"> • idImpact: primary key, created automatically • Description: description of the respective implementation • idOrganisation: references <i>Organisation</i>, which has implemented standard or project • idStandards_and_Projects: references <i>Standards_and_Projects</i>, which has been implemented • idPreservation_Lifecycle: references <i>Preservation_Lifecycle</i> to indicate an OAIS-Function • idFurther_Information: references <i>Further_Information</i>
<i>N:m Relationships</i>	
<i>Entries</i>	

Preservation_Lifecycle	
<i>Description</i>	Enables the assignment of an impact to a function of the <i>Functional Model</i>
<i>Columns</i>	<ul style="list-style-type: none"> • _Lifecycle: primary key, created automatically • Label: denomination of the different functions • Description: Explanation of the different functions
<i>N:m Relationships</i>	
<i>Entries</i>	<ul style="list-style-type: none"> • Ingest • Archival Storage • Data Management • Administration • Access • Preservation Planning, • Common Services

Geographic_Scope	
<i>Description</i>	geographical assignment of a Standards/Project or an Organization
<i>Columns</i>	<ul style="list-style-type: none"> • idGeographic_Scope: primary key, created automatically • idGeographic_Scope_Level: references <i>Geographic_Scope_Level</i> • idCountry: references <i>Country</i>
<i>N:m Relationships</i>	
<i>Entries</i>	

Geographic_Scope_Level	
<i>Description</i>	Describes different categories of the geographical assignment
<i>Columns</i>	<ul style="list-style-type: none"> • idGeographic_Scope_Level: primary key, created automatically • Label: designation of a Level • Description: detailed description of an entry
<i>N:m Relationships</i>	
<i>Entries</i>	<ul style="list-style-type: none"> • National • Europe-wide • Worldwide • Other

Country	
<i>Description</i>	List of countries, which can be assigned to Geographic_Scope
<i>Columns</i>	<ul style="list-style-type: none"> • idCountry: : primary key, created automatically • Label: Name of country, e.g. Germany
<i>N:m Relationships</i>	
<i>Entries</i>	

Further_Information	
<i>Description</i>	Place for links to Websites with further Information
<i>Columns</i>	<ul style="list-style-type: none"> • idFurther_Information: : primary key, created automatically • URL: URL • Description: Information regarding URL • Last Accessed: Date, on which link was last accessed
<i>N:m Relationships</i>	
<i>Entries</i>	

Organisation	
<i>Description</i>	Further information about Organisations
<i>Columns</i>	<ul style="list-style-type: none"> • idOrganisation: primary key, created automatically • Long Name: complete name of Organization • Short Name: abbreviation of organization name • idFurther_Information: references <i>Further_Information</i> • idGeographic_Scope: references <i>Geographic_Scope</i> • idOrganisation_Type: references <i>Organisation_Type</i>
<i>N:m Relationships</i>	
<i>Entries</i>	

Organisation_Type	
<i>Description</i>	Enables to categorize Organizations, e.g. Body, NGO, Archive/Library, research institution, company,...
<i>Columns</i>	<ul style="list-style-type: none"> • idOrganisationType: ID for every entry • Name: Name of category • Description: Description of an entry
<i>N:m Relationships</i>	
<i>Entries</i>	

3 ONLINE STANDARDS REGISTRY

The (desktop) APARSEN WP13 Standards Database was planned to be migrated into an Online Standards Registry. As preparatory step, existing related initiatives and their offerings were looked upon and assessed before specifying the target audiences and functional specification of the APARSEN Online Standards Registry in more detail. As integral building block of the APARSEN VCoE, the schedule for embedding the Online Standards Registry is also discussed.

3.1 RELATED INITIATIVES

Before specifying the implementation of the APARSEN standards registry, comparable existing online portals/offerings and their access modalities were assessed and informed further specifications/developments.

3.1.1 Library of Congress – Tools Showcase

The LoC Tools Showcase [LoC-TS] provides a “selective list of tools and services” for Digital Preservation. While content-wise being more related to WP16 and WP21, it allows for full text search and filtering according to Tool Type, Usage and Content Domain. The records contain short descriptions of the tool/services itself, together with links to corresponding main (and alternative) web pages. Furthermore, information about audience levels and usage categories are provided.

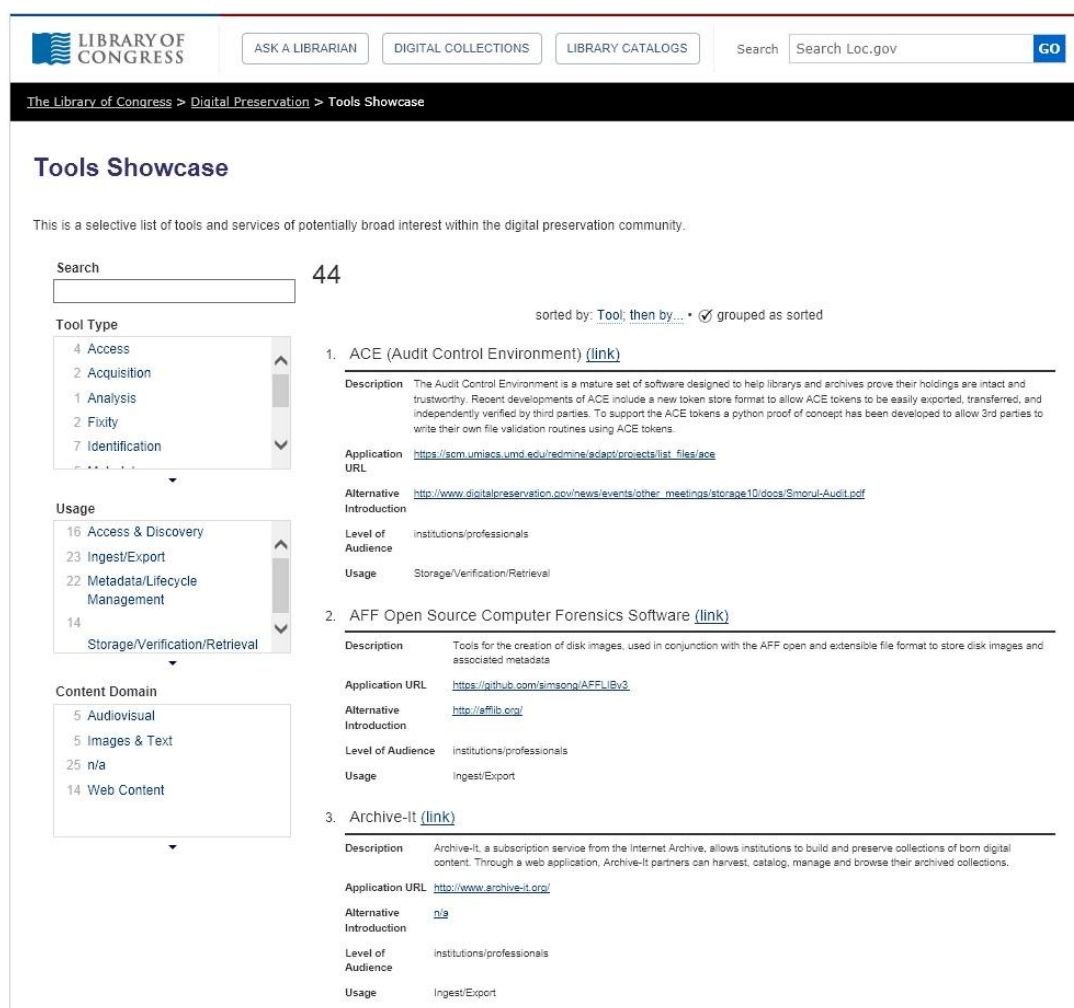
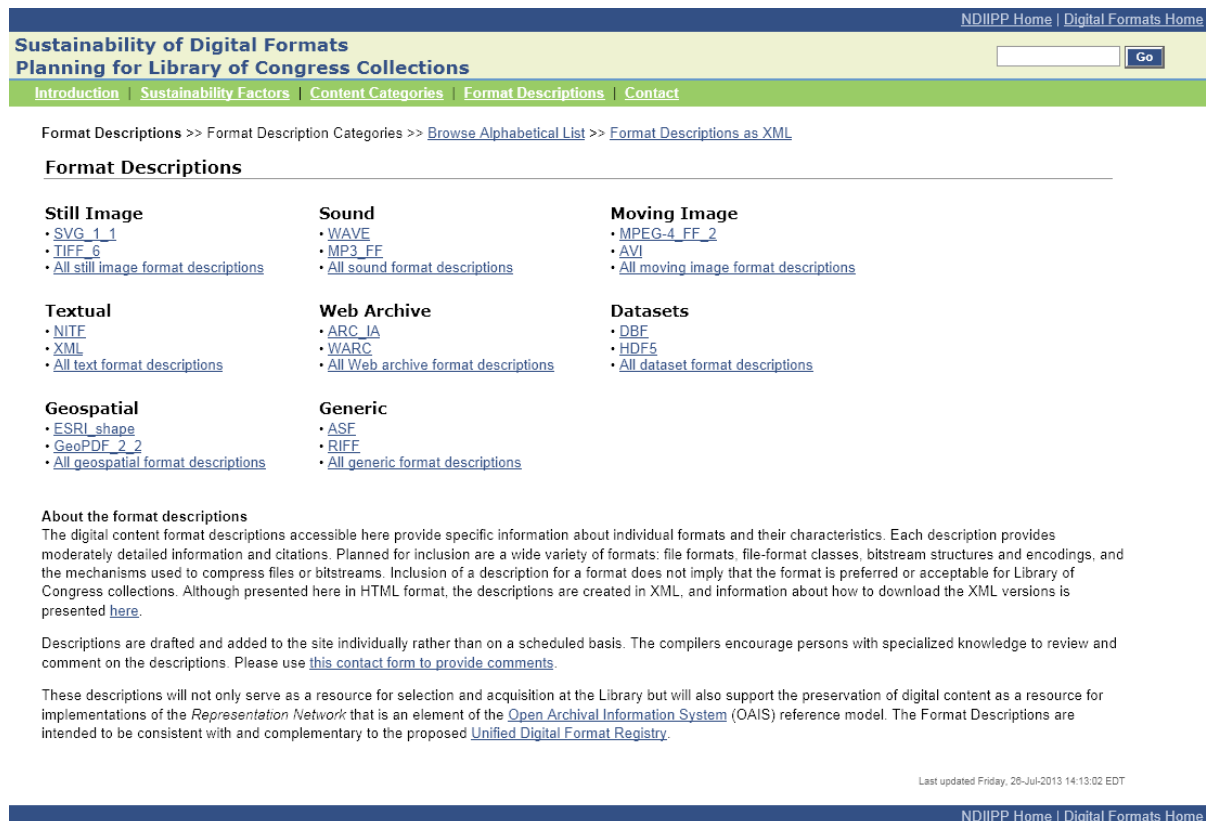


Figure 3: Library of Congress - Tools Showcase screen

3.1.2 Library of Congress – Sustainability of Digital Formats

The LoC Digital Formats [LoC-SoDF] website compiles information about digital content formats. It is organized into five sections: Introduction, Sustainability Factors, Content Categories, and Format Descriptions, which are structured into further subsections. These categorizations constitute views on formats, but also convey much additional background information. The actual format descriptions can be browsed according to (content) category and alphabetically, but can also be downloaded/accessed as XML.



The screenshot shows the 'Format Descriptions' page of the Library of Congress Digital Formats website. The page has a blue header with 'NDIIPP Home | Digital Formats Home' and a search bar. Below the header is a green navigation bar with links: 'Introduction', 'Sustainability Factors', 'Content Categories', 'Format Descriptions', and 'Contact'. The main content area is titled 'Format Descriptions' and lists various digital formats categorized into Still Image, Sound, Moving Image, Textual, Web Archive, Datasets, Geospatial, and Generic. Each category lists specific formats and links to 'All [category] format descriptions'. For example, under 'Still Image', it lists SVG_1_1, TIFF_6, and a link to 'All still image format descriptions'. Under 'Textual', it lists NITF, XML, and a link to 'All text format descriptions'. The page also includes a section 'About the format descriptions' explaining the purpose and availability of the descriptions, and a footer with the date 'Last updated Friday, 26-Jul-2013 14:13:02 EDT'.

Figure 4: Library of Congress - Format Descriptions screen

3.1.3 The National Archives – PRONOM

The PRONOM [PRONOM] registry provides technical information about data file formats, software products and other technical components. It offers diverse search facets (simple, by format, by identifier, by software, by vendor, by lifecycle, by migration path) and provides corresponding (detailed) reports as result.

The screenshot shows the PRONOM Search interface. At the top, there's a search bar and navigation links. The main heading is 'The technical registry PRONOM'. Below this, there's a 'Search : By format' section with several tabs. The 'File format' tab is active, displaying three search categories: '1. File formats', '2. Compatible software', and '3. File format risk'. Each category has a text input field and a 'Search' button. The 'File format' section also includes a 'Simple search' tab. The 'Compatible software' section has dropdown menus for 'all' and 'process in any way'. The 'File format risk' section has a dropdown for 'All' and a 'Search' button.

Figure 5: PRONOM Search screen

3.1.4 nestor – Standardisation

The nestor subsite [nestor2] provides static overviews of current standards for Persistent identifiers, Metadata, File formats: Identification/Registry, Audit & Certification, Reference Models, and Records Management. A list of links pointing to information about standards existing within each category is presented on separate static web pages respectively.

Home // **Standardisation**

- Home
- Working groups
- Publications
- Information services
- Training
- **Standardisation**
- nestor-Seal
- Network
- Events

STANDARDISATION

Many technical standards have yet to crystallise in the relatively new field of digital preservation. nestor provides an overview of existing standards, bundles together standardisation activities and proposes new activities where required. User interests are represented at the national and international levels via the partnership with the DIN organisation.

OVERVIEW OF CURRENT STANDARDS IN DIGITAL PRESERVATION

- Persistent identifiers
- Metadata
- File formats: Identification / Registry
- Audit & Certification
- Reference models
- Records Management

NATIONAL STANDARDISATION

To date, the nestor working groups have initiated three standardisation projects. These are currently being adapted by DIN for adoption as national standards:

- [Z nestor criteria - Catalogue of Criteria for Trusted Repositories Version II](#)
- [Z Catalogue of Criteria for Assessing the Trustworthiness of PI Systems](#)
- [Z Wege ins Archiv \("Into the Archive"\) - A guide for ingesting information into the digital repository](#)

The national standardisation takes place in the working group Records Management and Preservation of Digital Information Objects (NABD 15) in the DIN Standardisation Committee for Libraries and Documentation (NABD). The working group and nestor collaborate closely. The two former nestor working groups "Trusted Repositories" and "Standards for metadata, transfer of objects to digital repositories and object access" have been assimilated as discrete working groups into the Committee for Record Management and Preservation of Digital Information Objects in which their standardisation projects are being expanded into national standards. Further standardisation topics can similarly be handled in nestor working groups and be prepared for national standardisation.

➤ [DIN NABD 15](#)

@ [nestor](#)

Last update: 30.3.2012

Figure 6: nestor standardisation entry screen

3.1.5 DPC – Standards and Best Practice Guidelines, File Formats and Standards

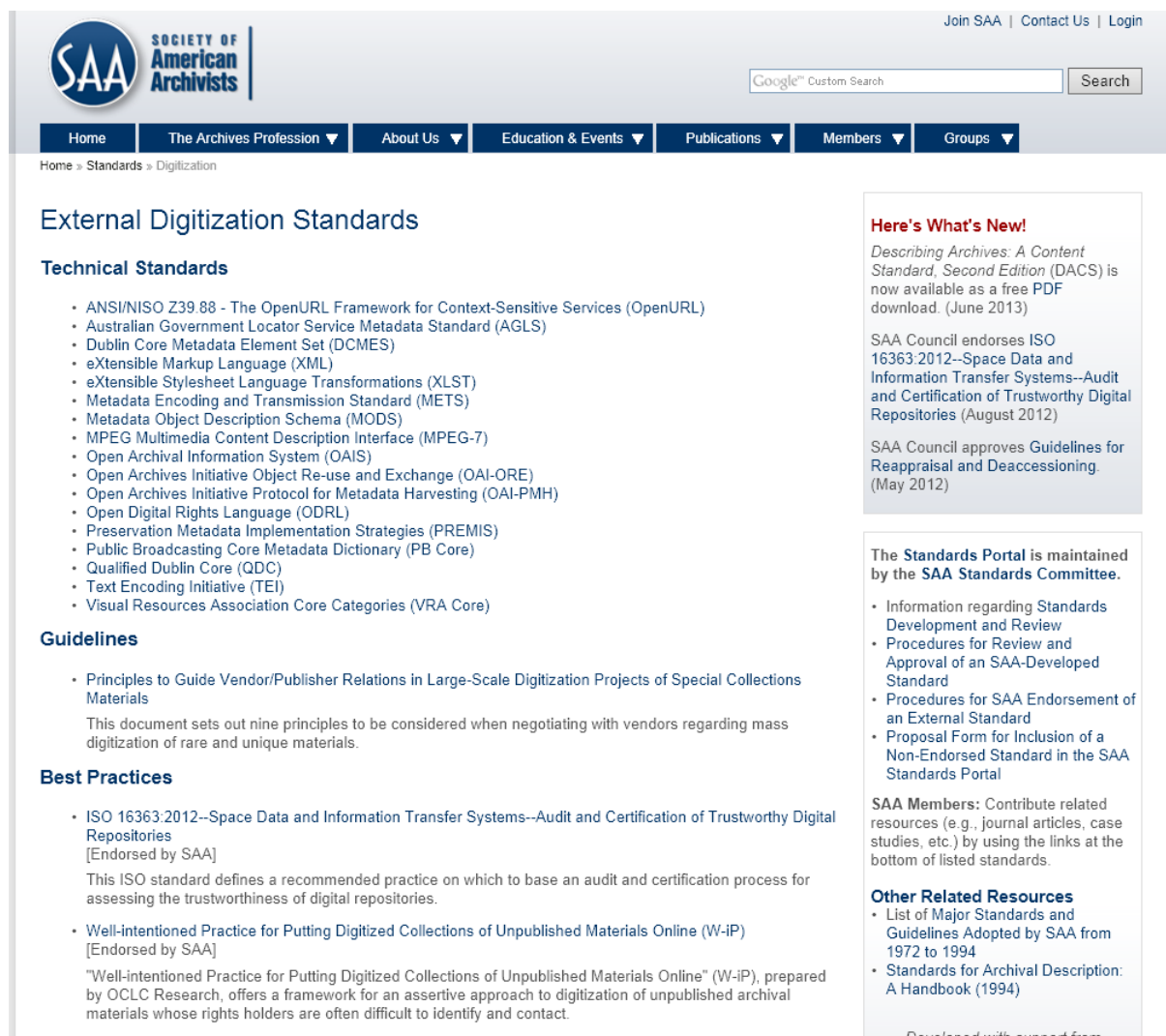
As part of the Institutional Strategies chapter, the Digital Preservation Coalition [DPC] Digital Preservation Handbook provides some basic background information about file format standards in general and associated best practices. The Media and Formats chapter some further high-level information on the subject. A few concrete formats/standards are named as examples within the narrative, but without providing any further detailed information or external references to them (link to "See Exemplars and Further Reading" yields 404 error).

The screenshot shows the Digital Preservation Coalition (DPC) website. At the top, there is a navigation bar with links: Home, Advice, Advocacy, Publications, Training, Events, Members, Newsroom, and About. Below this is a search bar and social media links. The main content area is titled 'Media and Formats - File Format and Standards'. On the left, there is a sidebar with a list of resources under the heading 'In this section'. The main content area contains text about file formats, their evolution, and the importance of backward compatibility. It also mentions the 'window of opportunity' to migrate and the impact of proprietary formats on preservation.

Figure 7: DPC Media and Formats screen

3.1.6 SAA – Standards Portal

The Society of American Archivists [SAA] Standards Portal provides information about SAA-approved standards, guidelines and best-practice documents. The portal front page is structured into 9 sections (Administration and Management, Appraisal and Acquisition, Arrangement and Description, Digitization, Education, Ethics, Values, and Legal Affairs, Preservation, Records and Information Management, and Reference and Access), which are pointing to one or more static web pages with additional links to references/resources.



The screenshot shows the SAA Standards page. At the top is the SAA logo and navigation links: Home, The Archives Profession, About Us, Education & Events, Publications, Members, and Groups. A search bar is also present. The main content area is titled 'External Digitization Standards' and includes sections for 'Technical Standards' (listing various standards like ANSI/NISO Z39.88, AGLS, DCMES, XML, XSLT, METS, MODS, MPEG-7, OAI, OAI-ORE, OAI-PMH, ODR, PREMIS, PB Core, QDC, TEI, and VRA Core), 'Guidelines' (Principles to Guide Vendor/Publisher Relations), and 'Best Practices' (ISO 16363:2012 and Well-intentioned Practice for Putting Digitized Collections Online). On the right, there are sidebars for 'Here's What's New!' (Describing Archives: A Content Standard), 'The Standards Portal is maintained by the SAA Standards Committee' (listing development and review procedures), and 'Other Related Resources' (List of Major Standards and Standards for Archival Description: A Handbook).

Figure 8: SAA Standards screen

3.1.7 PrestoCentre – PrestoCentre Library

The PrestoCentre [PrestoCentre] Library allows to personalize content by community of practice, thus customizing the results which are shown under the respective Library topics (Access, Acquisition, Archival Storage, Cataloging and Metadata, Context, Strategy, Vision, Finance, Ingest, Legal, Preservation Planning). Items within a chosen topic are categorized according to their types (Resources, Tools, and Projects) and can be browsed accordingly. The front page also offers the option to browse all library items by type.

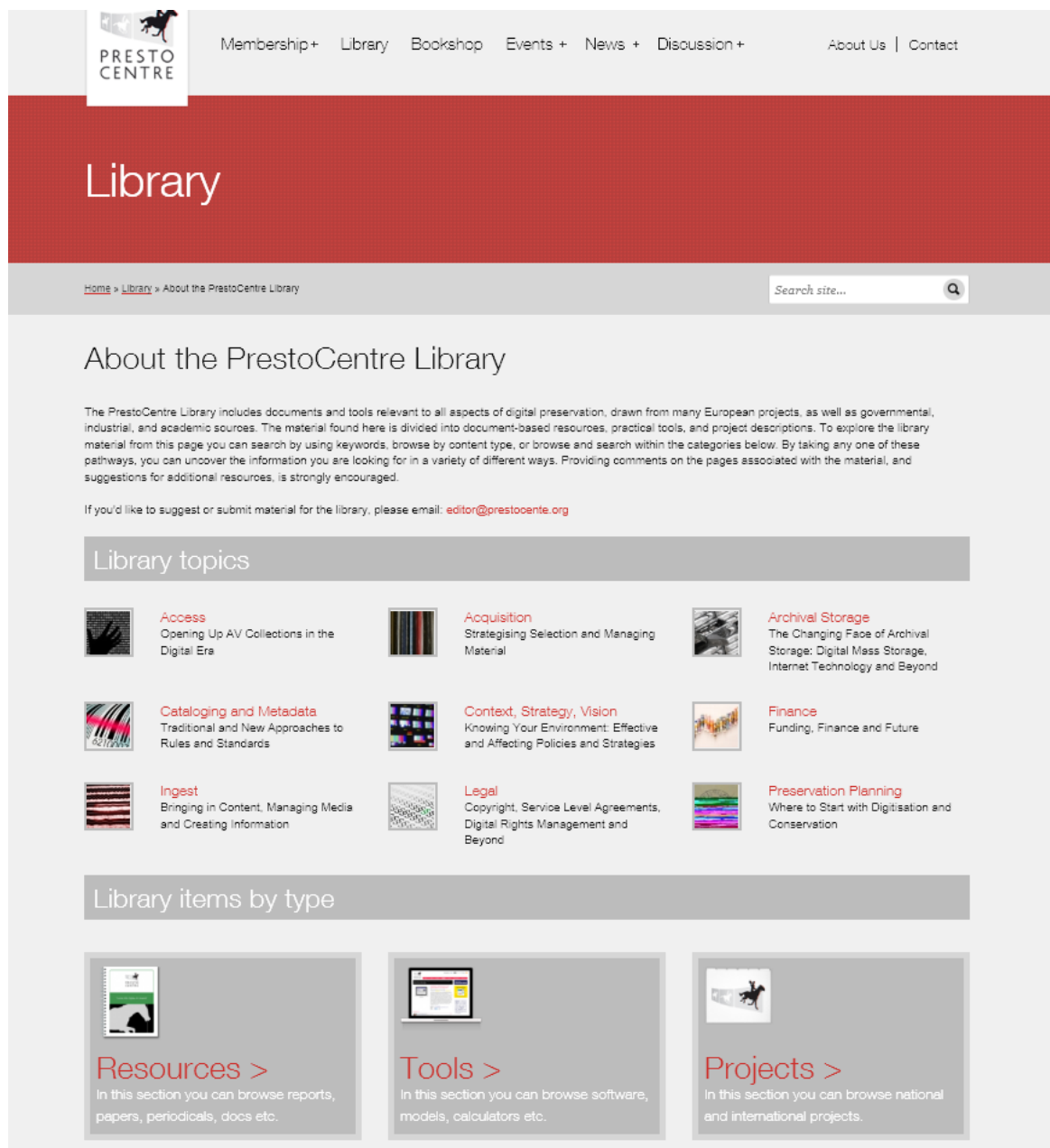
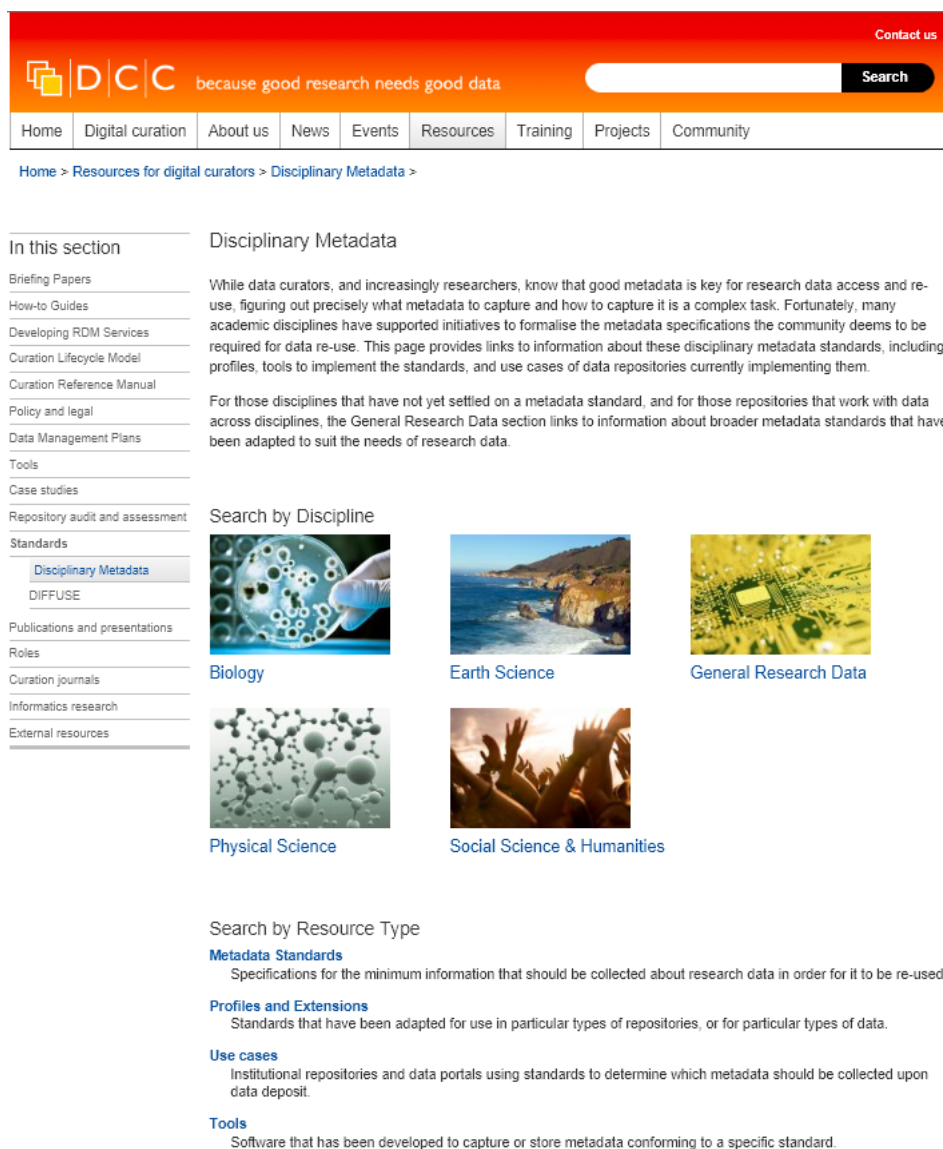


Figure 9: PrestoCenter Library screen

3.1.8 DCC – Digital Curation Standards

The resources section on the Digital Curation Coalition [DCC] web page comprises a subsection on Standards, which is further structured into Disciplinary Metadata (Biology, Earth Science, General Research Data, Physical Science, and Social Science & Humanities) and Standards Frameworks for specific domains originating from the DIFFUSE project (2000-2003). Disciplinary Metadata provides links to information about metadata standards and associated profiles, tools, and use cases depicting their implementation. Access to DIFFUSE Standards Frameworks is performed by either selecting an appropriate Lifecycle Action from the DCC Curation Lifecycle Model or by browsing all standards by Type, Title, or Standards Organisation. Records within the Standards and Specifications Database comprised fields for scope, status, description about development, practical examples, links to official documentation, links to supplemental documentation, and links to database entries concerning

standards bodies. Currently, there appears to be only a single generic framework and very few records available online and the last update to DIFFUSE data was made in 2009.



DCC because good research needs good data

Contact us

Search

Home Digital curation About us News Events Resources Training Projects Community

Home > Resources for digital curators > Disciplinary Metadata >

In this section




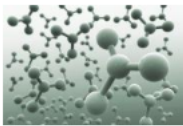

- Briefing Papers
- How-to Guides
- Developing RDM Services
- Curation Lifecycle Model
- Curation Reference Manual
- Policy and legal
- Data Management Plans
- Tools
- Case studies
- Repository audit and assessment
- Standards
 - Disciplinary Metadata**
 - DIFFUSE
- Publications and presentations
- Roles
- Curation journals
- Informatics research
- External resources

Disciplinary Metadata

While data curators, and increasingly researchers, know that good metadata is key for research data access and re-use, figuring out precisely what metadata to capture and how to capture it is a complex task. Fortunately, many academic disciplines have supported initiatives to formalise the metadata specifications the community deems to be required for data re-use. This page provides links to information about these disciplinary metadata standards, including profiles, tools to implement the standards, and use cases of data repositories currently implementing them.

For those disciplines that have not yet settled on a metadata standard, and for those repositories that work with data across disciplines, the General Research Data section links to information about broader metadata standards that have been adapted to suit the needs of research data.

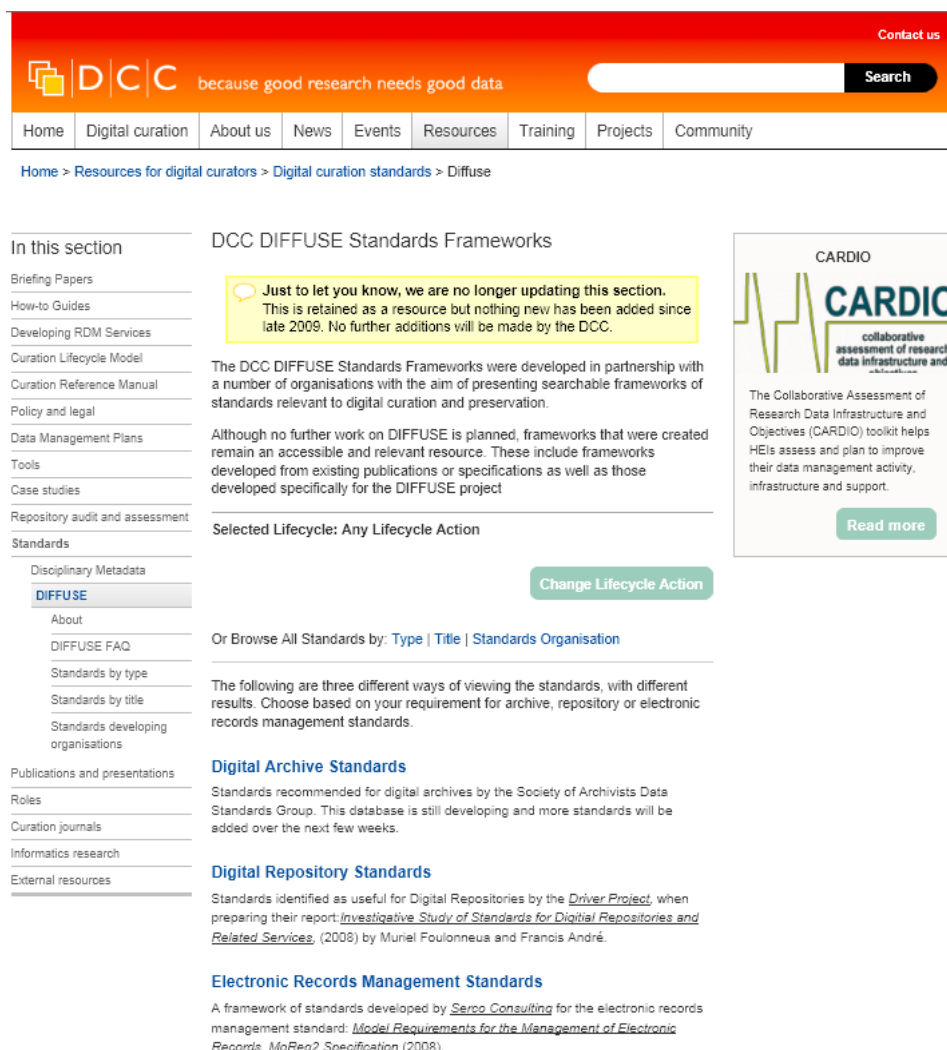
Search by Discipline

-  **Biology**
-  **Earth Science**
-  **General Research Data**
-  **Physical Science**
-  **Social Science & Humanities**

Search by Resource Type

- Metadata Standards**
Specifications for the minimum information that should be collected about research data in order for it to be re-used.
- Profiles and Extensions**
Standards that have been adapted for use in particular types of repositories, or for particular types of data.
- Use cases**
Institutional repositories and data portals using standards to determine which metadata should be collected upon data deposit.
- Tools**
Software that has been developed to capture or store metadata conforming to a specific standard.

Figure 10: DCC Metadata screen



The screenshot shows the DCC DIFFUSE website. The header is orange with the DCC logo and the tagline "because good research needs good data". A search bar is on the right. Below the header is a navigation menu with links: Home, Digital curation, About us, News, Events, Resources, Training, Projects, and Community. The main content area is titled "DCC DIFFUSE Standards Frameworks". A yellow box contains a message: "Just to let you know, we are no longer updating this section. This is retained as a resource but nothing new has been added since late 2009. No further additions will be made by the DCC." Below this, a paragraph explains that the frameworks were developed in partnership with various organizations. A sidebar on the left lists various resources under "In this section". A "Selected Lifecycle: Any Lifecycle Action" section is visible, along with a "Change Lifecycle Action" button. A "CARDIO" banner is on the right. The main content area also includes sections for "Digital Archive Standards", "Digital Repository Standards", and "Electronic Records Management Standards".

Figure 11: DCC DIFFUSE screen

3.1.9 The Open Group – Standards Information Base

The Open Group standards reference database [Open] comprises the current set of Open Group Standards and relevant industry standards which have been adopted. Records comprise fields for Reference (linking either to the Open Group Bookstore or to web sites of standards organizations, Description, Service Category, Type, and Status. Faceted search according to the aforementioned field (+ Title) is possible, in addition to browsing according to Service Categories in the TOGAF Technical Reference Model [TOGAF].

The screenshot displays the 'Standards Information Base' page for 'Information Technology – Digital Compression and Coding of Continuous-tone Still Images: Compliance Testing'. The page is part of 'THE Open GROUP' website. A navigation bar at the top includes links for Home, Sitemap, and a search box. Below this is a menu with categories: Subject Areas, Get Involved, Standards, Certifications, Events, Consortia, Software, Publications, and About Us. On the left, a sidebar lists user options: My Home, Update my Profile, My Roles, SIB, Search, View, Help, and TOGAF®. The main content area shows the title 'Standards Information Base' and details for 'ISO/IEC 10918-2:1995'. It includes a description of the standard, a list of common names (JPEG), status (Adopted), service (Data Interchange Services), category (Raw graphics interface), type (International Standard), and a URL. Below this, 'Bibliographic Details' are listed, including 'Reference ISO_10918-2 Jan 1995'. At the bottom, logos for The Open Group Platinum Members (IBM, PHILIPS, hp, ORACLE) are shown, along with footer links for Home, Sitemap, Privacy, and Legal, and a copyright notice for © 1995-2013 The Open Group.

Figure 12: The Open Group Standards screen

3.2 IMPLEMENTATION

3.2.1 Discussion

Many of the related initiatives which were inspected can be considered as static, manually compiled web pages which were sparsely populated. Specific standards search functions are often not provided, only site-wide full-text search is offered. Some of the initiatives are community-specific (e.g., addressing standards relevant for memory institutions), hence their usefulness/applicability across is limited. The maintenance status was not always obvious (some are no longer maintained, e.g., the DCC Digital Curation Standards has last been updated in 2009), but input/feedback by email was possible in some cases. Overall, evidence about the actual application/implementation/usage... of standards was rare.

3.2.2 Specification

The APARSEN Online Standards Registry is a web-based portal for structured access to standards and projects in Digital Preservation. This includes adequate management and maintenance functions.

The desktop standards database created with Open Office Base represents the baseline for the implementation; its contents were mapped to the revised schema (cf. Section 2.3) and migrated to MySQL RDBMS.

3.2.3 Target Groups

In general, two main target groups of information portals can be identified:

- Information providers – who want to disseminate information through the portal.
- Information seekers – who are looking for specific information in the portal.

D44.3 Report on Interactive Map of Stakeholders in Digital Preservation [GLOBIT] names the following APARSEN stakeholder groups:

- Data creators
- Data holders
- Service providers
- Research institutions
- Networks and multiplier organizations
- Qualification and training
- Accreditation and certification bodies
- Policy makers and funders
- Standardization bodies

Intuitively, Standardization bodies, Accreditation and certification bodies, and also, to some degree, Research institutions can be considered as Information providers. While Data creators and Data owners are potential adopters of Digital Preservation standards (and thus Information seekers), they might not be sufficiently aware of their importance. Hence, groups acting as mediators or brokers between Information seekers and Information providers need to be taken into account. This would correspond to Service providers, Networks and multiplier organizations, Qualification and training, and Policy makers and funders.

In summary, the APARSEN Online Standards registries addresses

- Information seekers
- Information providers
- Information mediators/brokers

3.2.4 Portal Functions

Access to the information contained in the APARSEN Standards database will be provided by a web-based portal with a graphical frontend. The frontend design will be aligned with the APARSEN (VCoE) web site.

The following portal functions will be supported:

- Authentication – Users need to log in for using the portal, which implies registration and access rights management (CRUD). Anonymous access will be discussed in alignment with Creation of Virtual Centre of Excellence (Task 1110).
- Data management – Authorized users can change/update existing records and add new records.
- Feedback functions – Users without proper authorization can request updates/changes, but also suggest new data.
- Redaction functions – Approval of registrations, changes/updates and new records by specific editor role (Optional).
- Search and navigation – Database queries, fulltext/faceted search, but also advanced/contextualized access functions based on use cases/scenarios (cf. Section 6).
- Reporting – Simple (visualized) statistics about usage/contents of the APARSEN Standards Registry.

3.2.5 Implementation Environment

Using JDK 1.7 as programming language (<http://www.oracle.com/technetwork/java/index.html>), Eclipse 4.3.1 (Kepler, <http://www.eclipse.org/>), extended by the Eclipse Datatools platform (DTP) and Eclipse Webtools Platform (WTP), serves as integrated development environment. Subversion (<http://subversion.apache.org/>) is used source version control system and Maven (<http://maven.apache.org/>) is used as build manager.

MySQL Community Edition 5.6 (<http://www.mysql.com/products/community/>) is used as RDBMS, the OO Base database was migrated by using the Eclipse Datatools Platform and then imported into MySQL.

Apache MyFaces Core 2.1 (<http://myfaces.apache.org/>) is used for implementing the web-based user interface, which implements the Java Server Faces 2.1

(<https://jcp.org/aboutJava/communityprocess/mrel/jsr314/index2.html>) specification.

3.3 EMBEDDING IN VCOE

The APARSEN Online Standards Registry is under development. The original OO Base standards database is migrated onto a MySQL server. For embedding the Online Standards Registry within the VCoE, a staged approach is taken:

- VCoE soft launch – Basic search/access functions based on SQL queries realized through WordPress (<http://wordpress.org/>) plugins. Email forms for providing input/feedback. Data management through MySQL Workbench (<http://www.mysql.com/products/workbench/>).
- Standalone evolution – Provision of JSF frontends for data management, provision of authentication functions, feedback and reporting functions and advanced/contextualized search and navigation.
- VCoE hard launch – Embedding in VCoE web site, alignment of authentication functions.

4 PRESERVATION SOLUTIONS (SOFTWARE APPLICATIONS)

4.1.1 Commercial Sector

4.1.2 Open Source Sector

Note: Activities have been performed in WP16 and will continue there throughout the project!

5 PERSISTENT IDENTIFIERS, FILE FORMATS AND OTHERS

Note: This chapter remains unchanged compared with D13.1!

5.1 PERSISTENT IDENTIFIERS

Persistent Identifier are used to address clearly indicated and permanent digital resources. Contrary to normal Web-URL they distinguish between identification and address of a resource. Using resolver technologies it can be ensured that access to resources can still happen even if the physical storage location has changed. Currently no standard is finalized but there several requests for comment visible. Most of them deal with Uniform Resource names (URN) and universally unique Identifiers (UUID)

A UUID is an identifier that is unique across both space and time, with respect to the space of all UUIDs. Since a UUID is a fixed size and contains a time field, it is possible for values to rollover (around A.D. 3400, depending on the specific algorithm used). A UUID can be used for multiple purposes, from tagging objects with an extremely short lifetime, to reliably identifying very persistent objects across a network.

The internal representation of a UUID is a specific sequence of bits in memory. To accurately represent a UUID as a URN, it is necessary to convert the bit sequence to a string representation.

The status of ongoing activities is covered in the document “Overview and studies on persistent identifier infrastructure commissioned by Knowledge Exchange and Prototype development of Meta Resolver Solution commissioned by SURFfoundation Part4”. The PersID project defines it work as follows:

“The PersID project, to successfully establish a reliable and internationally interoperable persistent identification infrastructure, aims to realize a trusting community, non-proprietary ownership of the PID system and the capability to anticipate on relevant future services.

These conditions for satisfactory international performance are best met by the URN-NBN option, because of the good experience with current national solutions, its inclusiveness (other PID systems can be resolved as well) and - being defined by IETF internet standards - its independence from vendors or technology platforms.

The current RFCs concerned with URNs and especially URN-NBNs were established in 1997. Since then the internet has changed and developed as well as function and usage of URNs. Therefore the PersID partners have an interest and in their policy agree to contribute to the updating and refinement of relevant RFCs and to comply to existing and future URN and URN-NBN related standards.”

http://www.persid.org/downloads/finalreports/PersID_Report_Part_4_final.pdf

The mentioned RFCs can be found using the following links:

[RFC 2141 - URN Syntax](#)

[RFC 1737 - Functional Requirements for Uniform Resource Name](#)

[RFC 2169 - A Trivial Convention for using HTTP in URN Resolution](#)

[RFC 2276 - Architectural Principles of Uniform Resource Name Resolution](#)

[RFC 2483 - URI Resolution Services Necessary for URN Resolution](#)

[RFC 2288 - Using Existing Bibliographic Identifiers as Using Existing Bibliographic Identifiers](#)

[RFC 3044 - Using International Serials Number as URN](#)

[RFC 3187 - Using International Standard Book Numbers as URN](#)

[RFC 3188 - Using National Bibliography Numbers as URN](#)

[RFC 3406 - URN Namespace Definition Mechanisms](#)

[RFC 4122 - A Universally Unique IDentifier \(UUID\) URN Namespace](#)

[RFC 4246 - International Standard Audiovisual Number](#)

Note: The contents of this section will be migrated into the database in period 4 and further investigations will be performed.

5.2 COMMON FILE FORMAT STANDARDS USED IN PRESERVATION SYSTEMS

To re-use content information under authenticity aspects it is mandatory to get detailed information about the file formats and codecs which were used to generate the content that has been preserved. The formats can be split in three sections:

- Documents and multimedia files
- Office and data base oriented files
- Engineering and CAD formats

While digital born office formats normally as standard function of the used office program can be converted in standardized file formats like PDF or PDF-A non- digital born document are typically preserved with scanning procedures which in the most cases produce some version of TIFF formats.

In Engineering and design the LOTAR consortium works currently (since 2008) in enabling existing STEP standards to be used for preservation of CAD models and product life cycle information. More information is available in the section 2.2.4 of this document.

For the visualization of the multimedia and office content either format oriented viewers, in some cases also the authoring systems (if still available) or multivalent technologies like the FAB4 browser as established in the SHAMAN project can be used. In the Engineering world STEP or J-T viewers or interfaces for the currently by the LOTAR consortium extended STEP formats will allow permanent access to engineering content.

The US Library of Congress has published a wide list of file formats for preservation artefacts, documents and codes which can be subject for long term preservation.

6 CONTEXTUALIZED ACCESS AND USER GUIDANCE

The APARSEN Online Standards Registry can be queried according to its underlying relational database schema. Several pre-defined queries will be available at the VCoE soft launch. However, Information Seekers might be incapable to express their needs through query forms, whereas Information Brokers/Mediators might need more advanced access functions which give them more control over their searches. All user groups might benefit from configuring the core APARSEN standards concepts within the SCIDIP-ES/APA Orchestration service being deployed at APA, a publisher-subscriber notification services informing about changes/additions to selected news categories.

6.1 CHALLENGES AND SCENARIOS FROM RTD WPS, USE CASES

To access standards and their description directly without long lasting search iterations It was assumed that different types of users will access the standards database. Novices which doesn't know any standard by name but have to get a useful standard with it descriptions and application background as well as preservation experienced users which need to get details about a standard or references how this standard is used and which organization is responsible to maintain or extend the current standard description.

For novices it got assumed that he can describe at the minimum the purpose or subject of the standard search and the domain where he is coming from or where he intends to use the standard.

To support this proceeding the search mask will contain selectable attributes focusing on the usage intend (subject or purpose) of the accessing person. The searcher can select from a scroll down menu bar the subject of the standard he is looking related to his application requirements.

He can select given attributes like e.g.:

- File formats
- Exchange formats
- Process oriented standards
- Meta data dictionaries
- Interface descriptions

and more

In parallel he should also select in a scroll down menu the domain where this standard should be applicable.

- Domains
 - Earth science
 - Industry
 - and more

The search result will give all standards which are meeting the given attributes as a list of standards.

To drill down the number of results which he might be gotten he can return to the search mask and can specify more attributes in the search mask like:

- Sub-Domain (e.g. for Earth Science)
 - Volcanology
 - Oceanography
- and more

- Applicability
 - Global
 - EU-wide
 - National
- License fees
 - Licensed
 - Free of charge
- Documentation fees
 - Charged
 - Uncharged
- Status of the standard
 - Proposed
 - White paper
 - Established
 - De-Facto

All filter or selection menu bars to search in the data base can be reviewed in the data base entry form, which has been part of the deliverable 13.1. The search mask itself is currently in an optimization process to improve the intuitive ease of use for especially those users.

The experienced user with a better know-how background can search with direct input of specific standard names or standard identifier directly. He can also get information about affiliated standards or families of standards especially in the area of meta data dictionaries and file formats.

In each case if there is more than one search result a list of applicable standards appears as result and selecting one item des detailed description of the standard with the related links and all other data which are stored in the data base appear in the equally formatted result document. From there he can identify the different references where this standard has been used, maintained and which impact the standard has generated once this information is available. Links to other standard descriptions like the “Library of Congress”, different versions or to responsible standardization organizations complement the result delivery.

This will give all users regardless if novices or experienced users easy access to applicable, detailed standard descriptions.

6.2 CONTEXTUALIZED GUIDANCE

The needs of Information Seekers might be vague, thus difficult to express through configuration of the advanced access functions which assume some familiarity with standards and the semantic of their attributes. To accommodate those users, the APARSEN Online Standards Registry will provide contextualized guidance, i.e. navigation support based challenges and scenarios derived from the RTD work packages and common vision discussions. The following context will be discussed further w.r.t. to their applicability before implementation:

- Topical navigation – Building on the APARSEN topic areas (cf. D13.1, Section 5.1), users are navigate from topics Trust, Sustainability, Access, and Usability into more specialized themes and are presented with matching records from the standards database. This would require the a priori mapping of standards to topics/themes, with coverage likely to vary across.
- OAIS Functional Model – The seven functions outlined by OAIS (Ingest, Archival Storage, Data Management, Administration, Access, Preservation Planning, and Common Services) might be used as gateways to standards. Again, a priori mappings are a fundamental prerequisite. As many Information Seekers might not be knowledgeable about OAIS, this context might only be appropriate for experts (e.g., Information Brokers/Mediators).

- Lifecycle actions – Lifecycle models (such as the DCC Curation Lifecycle [DCC], the SHAMAN Information Lifecycle Phases [SHAMAN], Product Lifecycle Management ...) might also be used for defining information paths from lifecycle actions to standards. Similar to OAIS above, this requires some familiarity of the Information Seeker with the model and corresponding mappings.
- Implementation evidence – As evidence about the actual usage/application of standards is integral to the APARSEN WP13 approach, Information Seekers are enabled to look for actual implementations within particular settings/domains.
- Use cases – Specific use cases, such as setting up a preservation-aware Product Lifecycle Management system or the evolution of a scientific data archive, can also be used as community-specific gateways to standards. As the communities represented by APA/APARSEN are diverse and heterogeneous, illustrative examples such as from Industrial Design and Engineering will be chosen as proof-of-concepts.

As stated above, further discussion are need w.r.t. to assessing appropriate contextual guidance options. As development resources are limited, some prioritization needs to be defined w.r.t. to actual implementation within the project's lifetime and beyond.

6.3 ALIGNMENT WITH SCIDIP-ES/APA ORCHESTRATION SERVICE (OPTIONAL)

The Orchestration service is a general knowledge broker for exchanging information about events which might impact the long-term usability and/or access of data (see [ENG]). Such events might include: changing technologies (e.g., support for new media and data formats), changing of terminologies/knowledge of specific “Designated Communities”, or change of ownership of data/archives. Specifically, for APARSEN WP13 we are looking into supporting triggers for notifications about changes in existing standards, availability of new standards within a particular domain or with a particular subject, and the availability of evidence illustrating the adoption/implementation of a particular standard within a particular context. Technically, the standards concepts deemed as relevant will be configured as topics in the Orchestration service. The topics will be connected to corresponding triggers defined in the APARSEN WP13 Standards database. Any changes/additions might either be immediately propagated or have to be approved through a specific redaction function.

Note: It has not been decided yet, whether this functionality can be implemented in the given timeframe until the end of year 4!

7 SPECIFICATION OF GAP ANALYSIS METHODOLOGY

The PARSE.Insight project (<http://www.parse-insight.eu/>) identified gaps in the European e-infrastructure. The methodology for APARSEN WP13 is adapted from the PARSE.Insight gap analysis framework. Its application will assess the standards inventory w.r.t. its coverage of the project's topic areas and associated themes for selected communities.

7.1 FRAMEWORK FOR ELICITING GAPS IN TOPICAL COVERAGE

A gap analysis is conducted to determine the difference between the actual state and a desired future state of a system. The frame of reference is defined by the APARSEN Common Vision. Hence, as depicted in D13.1 Section 5.1 the topic areas define the relevant structuring dimensions (Trust, Sustainability, Access, and Usability) and corresponding attributes (themes associated with topic areas) for assessing topical coverage. And allow for identifying (community-specific) areas where common standards would be of benefit (cf. Task 1320 Identification of new standards required, DoW). In consequence, the APARSEN WP13 gap analysis methodology is scenario-based, i.e. it contains generic components (topics, themes) and is to be customized by domain- and community specific lifecycle models and associated views on standards.

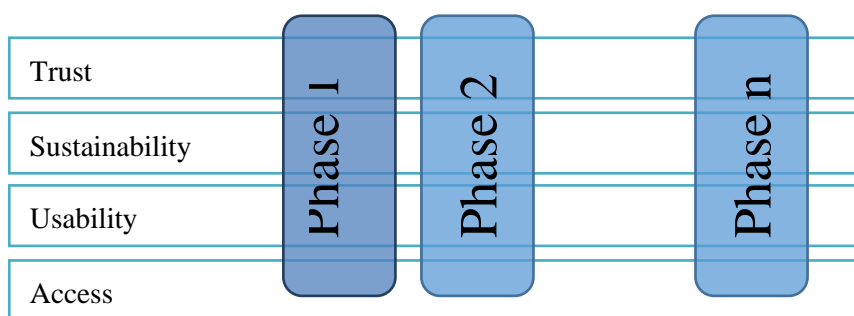


Figure 13: Gap Analysis Framework

Figure 13 illustrates the Gap Analysis Framework derived from PARSE.Insight. Vertically the spectrum of Digital Preservation standards is diffused into Trust, Sustainability, Usability and Access, i.e. the standards contained in the database are mapped to the APARSEN topic areas and themes. Horizontally, the depicted phases contextualize the vertical dimensions by respective domain/community-specific aspects of the chosen lifecycle models. For each lifecycle phase, applicable standards in use are identified and listed, e.g., in a spreadsheet. The lists are then matched against the topics/themes associated with each standard, yielding a quantitative overview of topical coverage for each phase. Insights from these overviews are analyzed/validated qualitatively by domain experts/community members in focused workshops/round-table discussions.

Hence, the APARSEN WP13 gap analysis process encompasses the following steps:

- Select established lifecycle model appropriate for domain/community
- List applicable standards in use for each phase
- Match against associated APARSEN topics/themes
- Analyze/discuss quantitative overviews (possibly iteratively)
- Consolidate findings, report and document

The next section illustrates the customization of the Gap Analysis Framework to the design process in engineering.

7.2 CUSTOMIZATION TO THE ENGINEERING DOMAIN

In the engineering domain, the design lifecycle can be structured into 5 distinct phases (cf. [SHAMAN D12.2]):

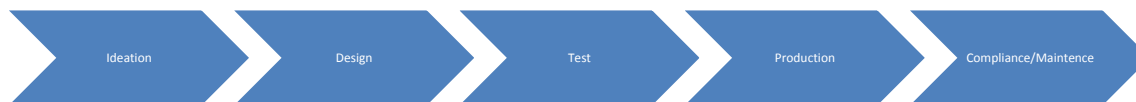


Figure 14: Simplified Product Life Cycle in Engineering

These lifecycle phases are characterized as follows (taken from [SHAMAN D12.2], Section 2):

- **Ideation** – Prior to the entry in the structured collaborative design phase the designer uses his local environment consisting of CAD, simulation, sketching, audio / video recording and office suite software tools to create, change, discard and re-use design documents or artefacts without any limitations given by the process rules of the corporate PLM systems. This phase is called the sandbox phase where data in most cases are not shared with other processes or colleagues in the project.
- **Design** – With the start of the collaborative design phase, engineering documents get shared to initiate new processes or workflows and require updates once data or design objects get changed. From this point forward all design activities are managed by the PLM (Product Lifecycle Management) systems.
- **Test** – One specific release step defines the entry in the prototyping, test and early manufacturing phase where prototypes are built and Zero series performed. With the entry in this phase the data are first time shared with the ERP (Enterprise Resource Planning) Systems. Beside the design files all manufacturing related documents are produced and transferred to the ERP. This includes the numerical control and manufacturing files, handling and application notes. The transfer consists of digital documents and further digital materials such as raw materials, purchased parts or electrical components which have been gone through a successful qualification process.
- **Production** – With the RFP (Release for Production; in some cases Mass Production) the design and test phase is completed. Changes which appear after the entry in this phase are based on existing engineering documents but those are taken as entry point in a new sandbox or design phase with a new part or project number. If such an action appears - may be based on changed marketing requirements, may be due to technological progress or to eliminate failures that have increased the warranty costs - Design Re-Use takes place. That means the documentation of the previous manufactured design is taken as entry point for all required changes. As soon as after production the product hits the market the legal obligations are beginning with warranty, service and maintenance and are finishing as actual design activity in the PLM systems.
- **Compliance/Maintenance** – For mass- or series -production like in the Automotive or in the Medical Equipment Industries the manufacturing phase is finished with the end of production (EOP) of the designed product in that specific version or revision it got originally its RFP. The retention time which needs to be kept starts with this last production day of this specific product revision. Even if other releases or revisions are still in the production process from a legal point of view the legal obligations related to the retention time must start for this designed equipment or part. Depending on the industry segment, there are maintenance

periods from 10 to 30 years defined by law or company contracts. For the automotive industry the VDA (Society of the German Automotive Industry) recommends 10 years after the EOP date out of liability reasons and additional 2 years a safety buffer for any design related data files. This explains the minimum time lines which have to be taken for compliance oriented preservation and archiving.

Re-use takes place within single phases and also among different phases. Two types of re-use can be distinguished: Design Re-use and Knowledge Re-use. With Design Re-use, coherent subsets of finished designs are taken over. Typical uses comprise the troubleshooting of a design or embedding a functional part in a new product. The emphasis is on re-use of data as in the case of e.g. numeric control or logical designs. With knowledge re-use, distinct information that guided a previous design are taken over. Typical uses comprise best practice or comparison with similar designs during the design process.

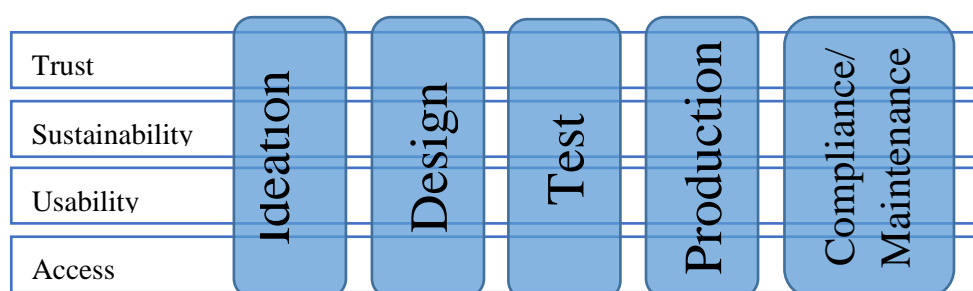


Figure 15: Gap Analysis Framework customized to Design Lifecycle

Figure 15 illustrates an instance of the APARSEN WP13 Gap Analysis Framework which is customized to a design lifecycle from the engineering domain.

7.3 APPLICATION ON SELECTED COMMUNITIES

The customized gap analysis framework from the previous section will be used for analyzing the current topical coverage of Digital Preservation topic areas by existing design and manufacturing standards in use (see Task 1320, DoW).

As preparatory step, the standards database will be extended by relating each record to APARSEN topic areas/themes. This will be done for all standards.

The WP13 partners will then identify applicable standards from the database for each design lifecycle phase, which will yield a quantitative overview of topical coverage for the engineering domain. The overview will then be discussed internally, but also externally by collaborating with relevant stakeholders from the ProSTEP iViP organization. And the overall findings from quantitative and qualitative analyses and discussions will then be fed into the roadmap discussions (cf. Section 8).

8 TOWARDS A ROADMAP FOR COMMON STANDARDS FOR INDUSTRIAL DESIGN AND ENGINEERING

8.1 METHODOLOGY

In the APARSEN project some members of the industry sharing their membership in industrial organizations like STEP, ProStep, EuroStep and others. These organizations are dedicated for setting up standards in the exchange of design and construction data and are also strongly engaged in projects

to allow especially for the airspace and automotive industries data standards for preservation of long-lasting systems like airplanes, cars and space satellites.

APARSEN collaborates with these organizations to enter standard references and content in its data base.

One of the projects under the umbrella of ProStep iViP a nonprofit industrial association in Germany is the LOTAR project which has been initiated by Airbus and other European partners to set standards as part of the STEP manufacturing standard family for the preservation of engineering documents and models to enable manufacturing and reuse years after the documents have been created to fulfill warranty and service obligations

The activities related to „Long Term Archiving & Retrieval“ in the aerospace and defense industry date back to 2002, and currently planned to go on until end of 2015.

- The common goal is the development of the standards series EN / NAS 9300-xxx.
- The multi-part standard covers both the information content and the processes required to ingest, store, administer, manage and access the information.

Since 2005 the activities of the different working groups are harmonized globally by the collaboration with Japanese and American standardization groups from the same industrial domains. Related to the industry and data domains the parts and related standards are structured as follows:

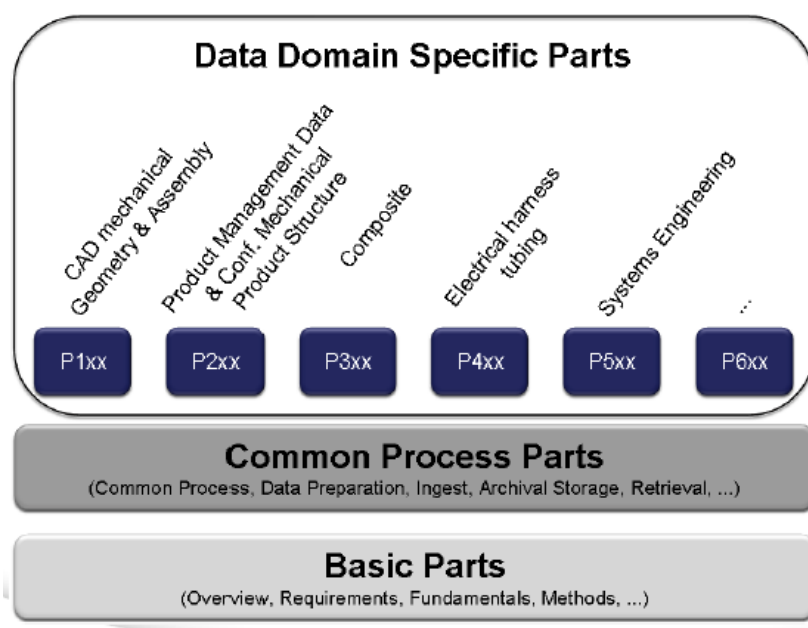


Figure 16: Data Domain specific parts

Beside the establishment of the standards itself, validation procedures are defined to test the accuracy and applicability with various test design to explore the quality of interfaces addressing those standards.

The LOTAR Approach

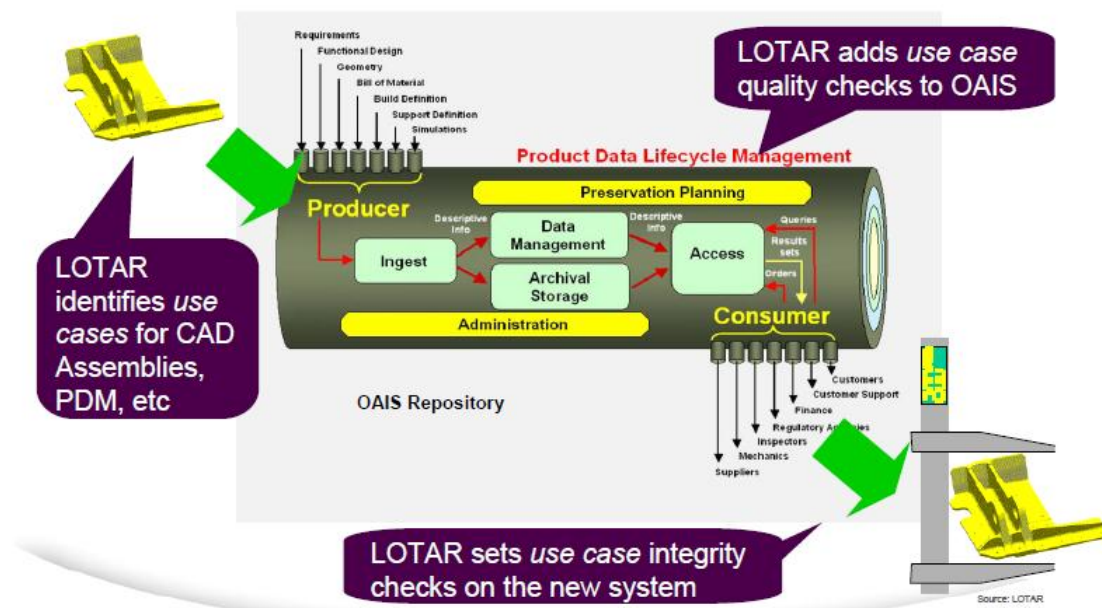


Figure 17: The LOTAR approach

In Collaboration with LOTAR the VDA organization (German Association of the Automotive Industries) adopts preservation standards to the needs of the automotive industry.

- VDA4958-1: Requirements and general recommendations for digital product data retention (legal + technical aspects)
- VDA4958-2: LTA Reference Process to design the processes (data preparation, ingest, archive, access and reprocess the archived data)
- VDA4958-3: Relevant LTA Data and Models defines the minimum of data/documents to be archived (CAD, product structure etc.) and mapping onto standard representations/formats
- VDA4958-4: Certification recommendations for the verification of reliable LTA workflows

The VDA organization works on Preservation use cases along the following schema:

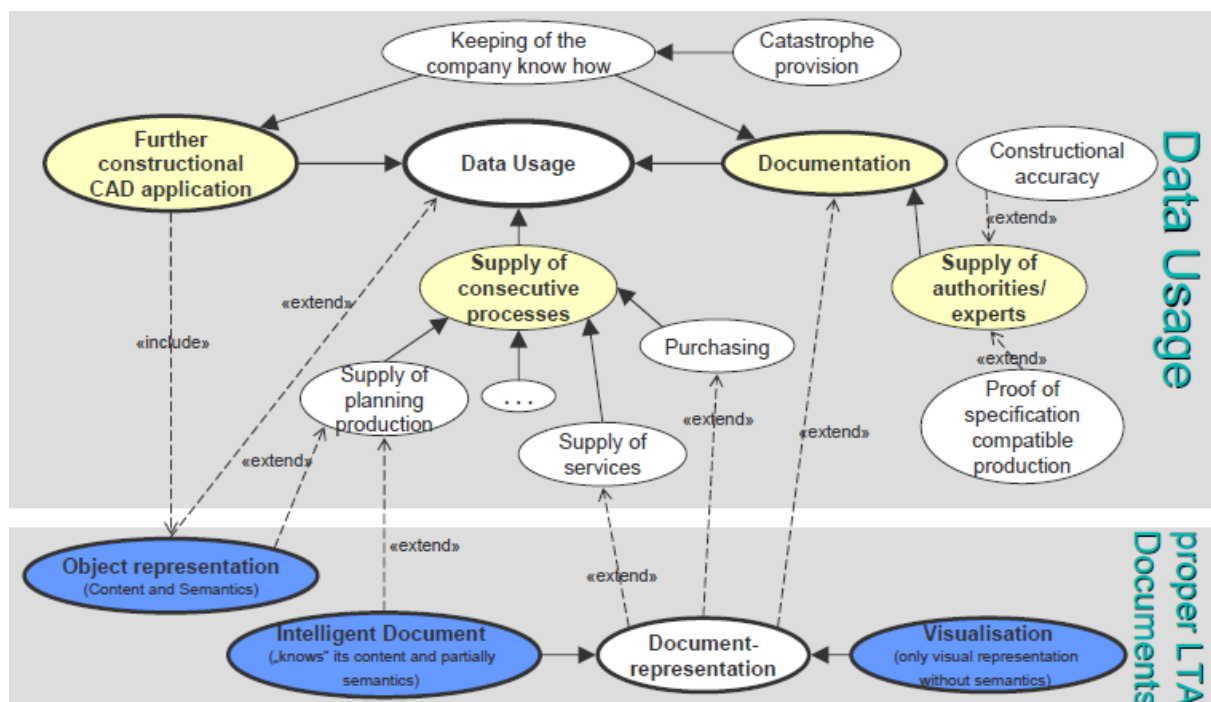


Figure 18: VDA preservation scheme

8.2 DIALOGUE WITH EXTERNAL STAKEHOLDERS (STEP)

To access the results of the standardization activities and to source industrial standard description and references APARSEN collaborates with these stakeholders.

By participation in their workshops and contribution of preservation results especially in the area of light weight visualization and acquisition of PLM (Product Lifecycle Management) oriented product structures and ontologies established during research activities in other EU-Projects like SHAMAN, APARSEN members present those results during some quarterly ongoing workgroup sessions.

Once the DB is fully established and accessible on the APARSEN VCoE web portal the sourcing methods will be opened also to those organizations to widen up the scope of the standard coverage for the entire industry.

LIST OF FIGURES

Figure 1: Year-2 database scheme in OpenOffice Base	12
Figure 2: New database in MySQL for VCoE	13
Figure 3: Library of Congress - Tools Showcase screen	22
Figure 4: Library of Congress - Format Descriptions screen	23
Figure 5: PRONOM Search screen	24
Figure 6: nestor standardisation entry screen	25
Figure 7: DPC Media and Formats screen.....	26
Figure 8: SAA Standards screen.....	27
Figure 9: PrestoCenter Library screen.....	28
Figure 10: DCC Metadata screen	29
Figure 11: DCC DIFFUSE screen.....	30
Figure 12: The Open Group Standards screen.....	31
Figure 13: Gap Analysis Framework.....	39
Figure 14: Simplified Product Life Cycle in Engineering	40
Figure 15: Gap Analysis Framework customized to Design Lifecycle	41
Figure 16: Data Domain specific parts	42
Figure 17: The LOTAR approach	43
Figure 18: VDA preservation scheme	44

List of Abbreviations

2D	two-dimensional
3D	three-dimensional
ABCD	Access to Biological Collections Data
ACI	Approval Content Information
AIA	Aerospace Industries Association
AIP	Archival Information Package
ANSI	American National Standards Institute
AP	Application Protocol (see STEP)
API	application programming interface
CAD	Computer Aided Design
CASPAR	Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval
CAX-IF	CAX Implementor Forum
CGM	Computer Graphics Metafile
CPM	Collaborative project management
CRM	Customer Relationship Management
CWM	Common Warehouse Metamodel
DCC	Digital Curation Centre

DCMI	Dublin Core Metadata Initiative
DDI	Data Documentation Initiative
DIF	Directory Interchange Format
DIP	Dissemination Information Package
DMS	document management systems
DP	digital preservation
DPE	Digital Preservation Europe
DRAMBORA	Digital Repository Audit Method Based on Risk Assessment
DwC	Darwin Core
DXF	Data eXchange Format
EAD	Encoded Archival Description
ECAD	Electronic and Electrical Computer Aided Design
EDA	Electronic Design Automation
EDIF	Electronic Design Interchange Format
e-GIF	eGovernment Interoperability Framework
EML	Ecological Metadata Language
ERP	Enterprise Resource Planning
FGDC	Federal Geographic Data Committee
GD&T	Geometric Dimensioning and Tolerance
GEM	Gateway to Educational Materials
GIS	geographic information system
GML	Geography Mark-up Language
HPGL	Hewlett Packard Graphical language
IAQG	International Aerospace Quality Group
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IGES	Initial Graphic Exchange Standard
IMS	Instructional Management Systems
IP	information package
ISO	International Standards Organization
LMER	Long-term preservation Metadata for Electronic Resources
LOTAR	LOng Term Archiving And Retrieval
MCAD	Mechanical Computer Aided Design
METS	Metadata Encoding and Transmission Standard
MIX	Metadata for Images in XML
MODS	Metadata Object Description Schema

MOF	Meta object facility
NARA	National Archives and Records Administration
nestor	Network of Expertise in Long-Term Storage and Long-Term availability of Digital Resources in Germany
NISO	National Information Standards Organization
NoE	Network of Excellence
OAIS	Open archival information system
OCLC	Online computer Library Center
OMG	Object Management Group
ONIX	ONline Information eXchange
OSD	Object-based Storage Device
PAIS	A Producer-Archive Interface Standard
PCB	Printed Circuit Board
PDF	Portable Document Format
PDI	Preservation Description Information
PDM	Product Data Management
PI	packaging information
Planets	Preservation and Long-term Access through Networked Services
PLM	Product Lifecycle Management
PREMIS	PREservation Metadata Implementation Strategies
PRONI	Public Record Office of Northern Ireland
RA	reference architecture
RDBMS	relational database management system
RLG	Research Libraries Group
SCM	supply chain management
SCORM	Sharable Content Object Reference Model
SIP	Submission Information Package
SOA	Service-Oriented Architecture
STEP	STandard for the Exchange of Product Data (ISO 10303)
TEI	Text Encoding Initiative
TIFF	Tagged Image File Format
TRAC	Trustworthy Repositories Audit & Certification: Criteria and Checklist
TSC	Technical Steering Committee
UML	Unified Modeling Language
URL	Uniform Resource Locator
VCoE	Virtual Center of Excellence
VDA	Verband der Automobilindustrie (German Association of the Automotive Industry)

VIM	vendor interface module
W3C	World Wide Web Consortium
XMI	XML Metadata Interchange
XMP	Extensible Metadata Platform

References

All Web URLs, if not otherwise noted, were last accessed between February 20, 2012 and December 21, 2013.

Allinson, J. (2006)

OAIS as a reference model for repositories: an evaluation

Available: <http://www.ukoln.ac.uk/repositories/publications/oais-evaluation-200607/Drs-OAIS-evaluation-0.5.pdf>

APARSEN-REP-D22_1-01-1_2 (2012)

Persistent Identifiers Interoperability Framework

APARSEN-REP-D25_1-01-1_7 (2013)

Interoperability Objectives and Approaches

Arnaud L. (2005)

XFDU packaging contribution to an implementation of the OAIS reference model.

Available: <http://www.ukoln.ac.uk/events/pv-2005/pv-2005-final-papers/043.pdf>

Austin, T., Mitcham, J. (2007)

Preservation and Management Strategies for Exceptionally Large Data Formats: 'Big Data'. AHDS and English Heritage.

Available: http://ads.ahds.ac.uk/project/bigdata/final_report/bigdata_final_report_1.3.pdf

Barateiro, J., Antunes, G., Cabral, M., Borbinha, J., Rodrigues, R. (2008)

Using a Grid for Digital Preservation, to appear in ICADL 2008.

Barker, S., James, H., Knight, G., Milligan, C., Polfreman, M. (2004)

Long-Term Retention and Reuse of E-Learning Objects and Materials, Report Commissioned by JISC.

CASPAR-D1101-TN-0101-1_0:2007

Review of the State of the Art

Available: <http://admin.casparpreserves.eu/Members/cclrc/Deliverables/review-of-state-of-the-art-1/view>

Cheney J., Lagoze, C., Botticelli P. (2001)

Towards a Theory of Information Preservation. In: Constantopoulos & I.T. Sølvsberg (Eds.): ECDL 2001, LNCS 2163, pp. 340–351, Springer-Verlag Berlin Heidelberg

Currall, J., McKinney, P., Johnson, C. (2006),

Digital Preservation as an Albatross. In Archiving 2006; May 23-26 2006, Ottawa.

Available: <https://dspace.gla.ac.uk/handle/1905/535>

Dappert, A., Enders, M. (2008)

Using METS, PREMIS and MODS for Archiving eJournals. In: D-Lib Magazine September/October 2008, Vol. 14, Number 9/10, ISSN 1082-9873

Available: <http://www.dlib.org/dlib/september08/dappert/09dappert.html>

DCC

Available: <http://www.dcc.ac.uk/resources/standards>

DCC/DPE/nestor/CRL (2007)

DCC/DPE/nestor/CRL Core Requirements for Digital Archives, Center for Research Libraries (CRL), Available: <http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=92>.

DigitalPreservationEurope (2006)

DPE Research Roadmap, DPE-D7.2,

Available:

http://www.digitalpreservationeurope.eu/publications/reports/dpe_research_roadmap_D72.pdf

DLF (2007)

Metadata Encoding and Transmission Standard: Primer and Reference Manual. Digital Library Federation.

Dobratz, S., Schoger, A. (2007)

Trustworthy Digital Long-Term Repositories: The Nestor Approach in the Context of International Developments , LNCS 4675, pp. 210-222.

Dobratz, S., Schoger, A. and Strathmann, S. (2006).

The nestor Catalogue of Criteria for Trusted Digital Repository Evaluation and Certification. In: "Digital Curation & Trusted Repositories: Seeking Success" Workshop in conjunction with the Joint Conference on Digital Libraries, JCDL 2006: Opening Information Horizons, June 11-15, 2006, Chapel Hill, NC, USA. Retrieved 14 December 2006

Available: <http://sils.unc.edu/events/2006jcdl/digitalcuration/Dobratz-JCDLWorkshop2006.pdf>

DPC

Available: <http://www.dpconline.org/advice/preservationhandbook>

DPE Roadmap (2007)

DPE Digital Preservation Research Roadmap. Public Deliverable, DigitalPreservationEurope, 2007

Available: http://www.digitalpreservationeurope.eu/publications/dpe_research_roadmap_D72.pdf

DPHEP Study Group (2009)

Data Preservation in High Energy Physics. Arxiv preprint arXiv: 0912.0255, (December), 1-18.

Retrieved from <http://arxiv.org/abs/0912.0255>

Edelstein et al. (2011)

Orit Edelstein, Michael Factor, Ross King, Thomas Risse, Eliot Salant, and Philip Taylor. Evolving Domains, Problems and Solutions for Long Term Digital Preservation, In Proc. of 8th International Conference on Preservation of Digital Objects (iPRES 2011), Singapore, November 1-4, 2011.

Available: <http://www.l3s.de/~risse/pub/ipres2011.pdf>

Edmondson, R. (2004)

Audiovisual archiving: philosophy and principles. Paris, UNESCO

Available: <http://unesdoc.unesco.org/images/0013/001364/136477e.pdf>

ENG

Available: <http://int-platform.scidip-es.eu/joomla/index.php/orchestrationservicemenu>

Ferreira, M., Baptista, A.A. and Ramalho, J.C. (2006).

A Foundation for Automatic Digital Preservation, Ariadne, Issue 48, July 2006.

Available: <http://www.ariadne.ac.uk/issue48/ferreira-et-al/>

Flouris, G. & Meghini, C. (2007)

Some Preliminary Ideas Towards a Theory of Digital Preservation, 2007, In Proceedings of the 1st International Workshop on Digital Libraries Foundations (DLF1-07)

Available: <http://www.ite.gr/ics/isl/publications/paperlink/DLF107.pdf>

Fresco, M. (2008)

"MoReq2: what it is, and how to use it?", presentation to the 5th DLM Forum Conference, Toulouse, France, December 2008.

Available: http://www.project-consult.net/Files/moReq2_and_how_to_use_it_-_publication_version.pdf

Fresco, M. (2009)

“A European Contribution to Digital preservation”, paper presented at DigCCurr 2009 (Digital Curation Curriculum 2009), April 2009, Chapel Hill, NC, USA.

Available: <http://www.ils.unc.edu/digccurr2009/2b-fresko.pdf>

Fuhr, N. (2005)

Information Retrieval From Information Access to Contextual Retrieval. Designing Information Systems. Festschrift für Jürgen Krause, UVK Verlagsgesellschaft

Available: <http://www.is.informatik.uni-duisburg.de/bib/pdf/ir/Fuhr:05a.pdf>

GLOBIT

APARSEN D44.3 Report on Interactive Map of Stakeholders in Digital Preservation.
<http://www.alliancepermanentaccess.org/wp-content/uploads/2012/04/D44.3-Report-on-interactive-map-of-stakeholders-in-digital-preservation.pdf>

Greenberg, J. (2002)

Metadata and the World Wide Web, Encyclopedia of Library and Information Science, Vol. 72, Supplement 35, pp. 244–261

Heutelbeck, Brunsmann, Wilkes, Hundsdörfer (2009)

"Motivations and Challenges for Digital Preservation in Design and Engineering", a paper presented at the First International Workshop on Innovation in Digital Preservation (InDP), June 19, 2009 in Austin, Texas, USA

Higgins, S., Semple, N. (2006).

OAIS Five-year review: Recommendations for update, Compiled on behalf of members of The Digital Curation Centre and The Digital Preservation Coalition.

Available: http://www.dcc.ac.uk/docs/publications/OAIS_5-Year_Review.pdf

Hockx-Yu, Helen (2006)

Digital preservation in the context of institutional repositories, Program: electronic library and information systems. (40)3.

Available:

<http://www.emeraldinsight.com/Insight/viewContentItem.do?contentType=Article&hdAction=lnkhtml&contentId=1563488>

IASA (2009)

Technical Committee, Kevin Bradley, (Ed.) *Guidelines on the Production and Preservation of Digital Audio Objects* (IASA-TC 04 Second edition) Canberra, IASA.

Available: <http://www.iasa-web.org/audiopreservation-tc04>

IEEE (1990)

Institute of Electrical and Electronics Engineers. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York.

Innocenti, P., Ross, S., Elena, M., Wilson, T., Ludwig, J., and Pempe, W. (2009)

Assessing digital preservation frameworks: the approach of the SHAMAN project. In: International Conference on Management of Emergent Digital EcoSystems, 27-30 October 2009, Lyon, France.

Innocenti, P., Vullo, G., and Ross, S. (2010)

Towards a digital library policy and quality interoperability framework: the DL.org project. New Review of Information Networking, 15 (1). pp. 29-53. ISSN 1361-4576 (doi:10.1080/13614571003751071)

Innocenti, P., Konstantelos, L., Ross, S., Maceviciute, E., Wilson, T., Ludwig, J., and Pempe, W. (2010)

Assessing digital preservation infrastructures: implementing a framework for library, engineering and eScience organisations. In: IS&T Archiving 2010: Preservation Strategies and Imaging Technologies for Cultural Heritage Institutions and Memory Organisations, 1-4 June 2010, Den Haag, The Netherlands.

ISO 15225:2000

Nomenclature -- Specification for a nomenclature system for medical devices for the purpose of regulatory data exchange. 19 pp.

ISO 15489-1:2001(E)

Information and Documentation – Records Management – Part 1: General.

ISO/IEC 11179

Information Technology -- Metadata registries (MDR).

ISO/TR 18492:2005(E)

Long-term Preservation of Electronic Document-based Information. Technical report ISO/TR 18492. ISBN 0 580 47827 0, 18 pp.

JISC Digital Media (2009)

Introduction to Digital Video

Available: www.jiscdigitalmedia.ac.uk/movingimages/advice/introduction-to-digital-video

Kashyap V., Sheth, A. (1996)

Schematic and Semantic Similarities between Database Objects: A Context-Based Approach, Very Large Databases J., vol. 5, no. 4, 1996, pp. 276–304.

Klas C.-P., Fuhr, N., Schaefer, A. (2004)

Evaluating Strategic Support for Information Access in the DAFFODIL System. In: Research and Advanced Technology for Digital Libraries. Proc. European Conference on Digital Libraries (ECDL 2004), LNCS 3232, Springer, pp. 476-487.

Available: <http://www.springerlink.com/content/ba6wuqcygj634y42/>

Knight, G. (2006)

A minimal preservation metadata element set for the SHERPA DP Project, AHDS, 21 pp.

Available: <http://www.sherpadp.org.uk/documents/wp44-preservation-metadata.pdf>

Knight, S. (2008)

From theory to practice: digital preservation at the National Library of New Zealand Te Puna Mātauranga o Aotearoa. IFLA, 2008

Available: <http://www.ifla.org/IV/ifla74/papers/084-Knight-en.pdf>. Accessed 30 May 2009

Lavoie, B.F., ed. (2001)

Preservation metadata for digital objects: A review of the state of the art (A white paper by the OCLC/RLG working group on preservation metadata). OCLC Research publication.

Available: http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf

Lavoie, B.F., ed. (2002)

Preservation metadata and the OAIS information model: A metadata framework to support the preservation of digital objects (OCLC/RLG working group on preservation metadata). OCLC Research publication. Available: http://www.oclc.org/research/projects/pmwg/pm_framework.pdf

Lavoie, B.F. (2004a).

Implementing metadata in digital preservation systems: The PREMIS activity. D-Lib Magazine 10, no. 4.

Available: <http://www.dlib.org/dlib/april04/lavoie/04lavoie.html>

Lavoie, B. (2004b)

The Open Archival Information System Reference Model: Introductory Guide. DPC Technology Watch Report 04-01, 19 pp.

Available: http://www.dpconline.org/docs/lavoie_OAIS.pdf

Lavoie, B., Henry, G., Dempsey, L. (2006)

A Service Framework for Libraries, D-Lib Magazine, 12(7/8) ISSN 1082-9873

Available: <http://www.dlib.org/dlib/july06/lavoie/07lavoie.html>

LoC-TS

Available: <http://www.digitalpreservation.gov/tools/>

LoC-SoDF

Available: <http://www.digitalpreservation.gov/formats/>

McHugh, A., Innocenti, P., Ross, S. (2008)

"Assessing risks to digital cultural heritage with DRAMBORA". *International Documentation Committee of the International Council of Museums (CIDOC) 2008, Athens, Greece, 15–18 September 2008*

McHugh, A., Ross, S., Ruusalepp, R., and Hofman, H. (2007)

The Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) (Glasgow, DCC/DPE)

Miles, S., Wong, C., Fang, W., Groth, P., Zauner, K.-P., Moreau, L. (2007)

Provenance based validation of e-science experiments. *Web Semant.*, 5(1), pp 28-38.

Moore, R. (2008)

Towards a theory of digital preservation. *The International Journal of Digital Curation*, V3 (1) 2008.

<http://www.ijdc.net/index.php/ijdc/article/viewFile/63/42>. Accessed 31 May 2009.

Moore, R. & Smith, M. (2007)

Automated Validation of Trusted Digital Repository Assessment Criteria, *Journal of Digital Information* 8(2).

Available: <http://journals.tdl.org/jodi/article/view/198/181>

Moore, R. W. and Smith, M (2006)

Assessment of RLG Trusted Digital Repository Requirements. In: "Digital Curation & Trusted Repositories: Seeking Success" Workshop in conjunction with the Joint Conference on Digital Libraries, JCDL 2006: Opening Information Horizons, June 11-15, 2006, Chapel Hill, NC, USA.

Available: http://sil.unc.edu/events/2006jcdl/digitalcuration/Moore_Smith-JCDLWorkshop2006.pdf

NEDLIB (2000)

Metadata for Long Term Preservation / by C. Lupovici and J. Masanès, Bibliothèque Nationale de France. Den Haag: Koninklijke Bibliotheek, 2000. (NEDLIB Report series; 2). ISBN 90-62-59-1469

Available: <http://nedlib.kb.nl/results/NEDLIBmetadata.pdf>.

nestor (2006)

Kriterienkatalog vertrauenswürdige digitale Langzeitarchive, Version 1 (Draft for public comment).

Available: <http://www.nbn-resolving.de/?urn:nbn:de:0008-2006060710>

nestor2 (2012)

http://www.langzeitarchivierung.de/Subsites/nestor/EN/Standardisation/standardisierung_node.html

OAIS 2012

Reference Model for an Open Archival Information System (OAIS) CCSDS 650.0-M-2 MAGENTA BOOK

Available: <http://public.ccsds.org/publications/archive/650x0m2.pdf>

OAIS Review DCC (2006)

Five-year review Recommendations for update (compiled on behalf of members of The Digital Curation Centre and The Digital Preservation Coalition by Sarah Higgins and Najla Semple)

Available: http://www.dcc.ac.uk/docs/publications/OAIS_5-Year_Review.pdf

OCLC and RLG (2005)

Data Dictionary for Preservation Metadata, Final Report of the PREMIS Working Group. OCLC and RLG.

OpenOffice.org

Apache OpenOffice (AOO) is an open-source office software suite.

Available: <http://www.openoffice.org>

Open

Available: <https://www2.opengroup.org/ogsys/jsp/publications/sibMainPage.jsp>

PAIS (2004)

Recommendation for a Producer-Archive Interface Methodology Abstract Standard, CCSDS 651.0-B-1, BLUE BOOK, May 2004.

Planets (2010)

Project website <http://www.planets-project.eu>

PREMIS (2008)

Data Dictionary for Preservation Metadata: PREMIS version 2.0,

Available: <http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>.

Presto4U

Project website: <http://www.presto4u.eu/>

PrestoCentre

Available: <https://www.prestocentre.org/library>

Presto (2001)

D2 Archive preservation and exploitation requirements.

Available: <http://presto.joanneum.ac.at/outcomes.asp>

PrestoCentre (2011)

Available: <http://prestocentre.eu/> and <http://blogs.ec.europa.eu/digital-agenda/prestocentre-anice-example-of-european-cooperation-for-preserving-and-making-available-our-cultural-heritage/#more-33>

PrestoPRIME (2011a)

Metadata Mapping, Verification And Visualization Tools.

Available: <http://prestoprime.joanneum.at/>

PrestoPRIME (2011b)

PrestoPRIME Modelling Tools by IT Innovation.

Available: <http://prestoprime.itinnovation.soton.ac.uk/>

PrestoSpace (2004)

D2-1 User Requirements Final Report.

Available: http://prestospace.org/project/deliverables/D2-1_User_Requirements_Final_Report.pdf

PrestoSpace (2005)

The History of the 'Preservation Factory' Concept.

Available: http://prestospace.org/PF_concept.pdf

PrestoSpace (2006)

The Preservation Guide Wiki

Available: <http://wiki.prestospace.org/>

The Temporary Archiving

Roadmap available at: <http://wiki.prestospace.org/pmwiki.php?n=Main.Roadmap>

PRONOM

Available: <http://www.nationalarchives.gov.uk/PRONOM>

Rabinovici and al. (2008)

Rabinovici-Cohen, S., Factor, M. E., Naor, D., Ramati, L., Reshef, P., Ronen, S., Satran, J., Giaretta, D. L. Preservation DataStores: New storage paradigm for preservation environments, IBM Journal of Research and Development, (issue on Storage Technologies and Systems), Vol. 52, Number 4/5, p p. 389-400.

Available: <http://www.research.ibm.com/journal/rd/524/rabinovici.pdf>

RLG-OCLC (2002)

Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report

Available: <http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>

Ross, S., Kim, Y. & Dobрева, M. (2007)

Preliminary framework for designing prototype tools for assisting with preservation quality metadata extraction for ingest into digital repository, Pisa, DELOS NoE, December 2007, ISBN 2-912335-39-6.

Ross, S., McHugh, A. (2006)

The Role of Evidence in Establishing Trust in Repositories, D-Lib Magazine, July/August, v.12, n7/8 (Also published in Archivi e Computer, August 2006)

Available: <http://www.dlib.org/dlib/july06/ross/07ross.html>.

SAA

Available: <http://www2.archivists.org/standards>

SCAPE Research (2011)

Stephan Strodl, Petar Petrov, Andreas Rauber. Research on Digital Preservation within Projects co-funded by the EU in the ICT Programme, SCAPE Public Report, June 2011.

Available: http://www.scape-project.eu/wp-content/uploads/2012/01/SCAPE_digpres_research_ict.pdf

SHAMAN

Holger Brocks, Alfred Kranstedt, Gerald Jäschke and Matthias Hemmje (2010). Modeling Context for Digital Preservation, Studies in Computational Intelligence 260/2010, Smart Information and Knowledge Management, Springer, Heidelberg, pp 197-226).

SHAMAN D12.2

“Report Investigating Demand For Rule-Oriented Data Management Technologies For The Design And Engineering Domain”, SHAMAN Project Deliverable, September 30, 2010

Sireman, B. (2006)

Report on Comparison of Planets with OAIS, Planets deliverable PP/7-D1.

Available: http://www.planets-project.eu/docs/reports/Planets_PP7-D1_ReportOnComparisonOfPlanetsWithOais.pdf

Smyth, Z. (2006)

Developing a pre-ingest strategy for digital records, Presentation at the Joint DCC/LUCAS Workshop, 30.11– 1.12.2006, Foresight Centre, Liverpool (2006). Available: <http://www.dcc.ac.uk/events/archives-2006/preingest.pps>.

Steinke, T. (2006).

Universal object format: An archiving and exchange format for digital objects. Federal Ministry of Education and Research.

Strodl, S., Rauber, A. (2007),

Preservation Planning in the OAIS Model, iPRES2007, October 11-12, 2007, Beijing, China.

Available: http://www.ifs.tuwien.ac.at/~strodl/paper/Strodl_IPRES07.pdf

Theodoridou et al., 2010

Theodoridou, M., Tzitzikas, Y., Doerr, M., Marketakis, Y. & Melessanakis, V.: Modeling and querying provenance by extending CIDOC CRM. Distributed and Parallel Databases 27(2): 169-210 (2010)

Thomaz, K. (2006)

Critical Factors for Digital Records Preservation. Journal of Information, Information Technology, and Organizations, Vol.1, pp. 21-41.

Available: <http://www.jiito.org/articles/JIITOV1p021-041Thomaz12.pdf>

TOGAF

Available: <http://www.opengroup.org/togaf/>

TRAC (2007)

Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist, RLG-NARA, v.1.

Available: <http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91>.

van Wijngaarden, H. (2007)

Long-term preservation and permanent access: How to ensure the long-term reuse value of your digital assets. Journal of Digit Asset Management, Vol. 3, pp. 102- 109, ISSN 1743-6540

Available: <http://dx.doi.org/10.1057/palgrave.dam.3650064>

Watry, P. (2007)

Digital Preservation Theory and Application: Transcontinental Persistent Archives Testbed Activity. International Journal of Digital Curation 2(2). Available <http://www.ijdc.net/ijdc/article/view/43/0>

Westhead, M., Wen, T., Carroll, R. (2003)

Describing data on the grid. In GRID, pp. 134 140.

Wheatley, P. (2004)

Institutional Repositories in the context of Digital Preservation. DPC Technology Watch Series Report 04-02, Digital Preservation Coalition.

<http://www.dpconline.org/docs/DPCTWf4word.pdf>.

Wilkes, Brunsmann, Heutelbeck, Hemmje, Hundsdörfer, Heidbrink (2009)

"Towards support for long-term digital preservation in product life cycle management" a paper presented at The Sixth International Conference on Preservation of Digital Objects (iPRES), October 5-6, 2009 in San Francisco, California, USA.

Woods, K., Brown, G. (2008)

Migration Performance for Legacy Data Access. International Journal of Digital Curation, vol. 3(2), issn 1746-8256,

Available: <http://www.ijdc.net/ijdc/article/view/88>