
reUSE!
reUSE Digital Master Files of Printed Material!
eContent Project No. 11173
http://reuse.uibk.ac.at/

# *White Paper on Digital Repositories*

Andreas Aschenbrenner, Max Kaiser
andreas.aschenbrenner@researchstudio.at
max.kaiser@onb.ac.at

with a chapter on submission agreements by Christian Recht
and contributions by Sven Bahnik, Susanne Dobratz, Liina Enok, Michael Horvath, Bettina Kann, Krista Kiisa, Marco Köttstorfer, Günter Mühlberger, Matthias Schulz

March 2005

## Executive Summary

reUSE is a collaborative project of national and university libraries from a variety of European countries. Their shared mission is to collect, archive, and make available publications from various sources within their area of influence, including public sector institutions, non-profit organisations, higher education, and other organisations disseminating textual documents into the public domain. Formerly the mere collection of printed copies was sufficient to fulfil this mandate, however the domain moves increasingly into digital processes. Publications are originally created in digital form, and an increasing number of publications are only available digitally. reUSE partners aim to extend their activities to digital master files and thereby tie into the traditional publishing process in order to cause minimum disruption and additional effort for publishers. reUSE has already sparked interest at potential data providers for the possibility to preserve and raise the profile of their publications, and future users are likely to be attracted by the comprehensive collections of publications that are openly accessible for re-use.

The *reUSE White Paper on Digital Repositories* is based on two main pillars, an environmental scan of current developments and open issues in the international arena, and a status check and cross-analysis of the ongoing implementation work at the reUSE demonstrators. The paper aims to feed into demonstrator implementation, guide the development of the evaluation framework (Work Package 3 in the reUSE project), and contribute to internal discussion and exchange of experience.

International digital repositories research achieved major advancements in the last couple of years. Fundamental theoretical works nourished international discussion and shaped thinking about digital repositories, and wide-spread implementation efforts further contributed to research and raised the standing of digital repositories in society. While in the late 1990s digital repository initiatives struggled mainly with technicalities, the focus has now moved to establishing ‘trusted' digital repositories that are organisationally viable and sustainable in the long term. reUSE builds heavily on the work on 'trusted' digital repositories and the OAIS (Open Archival Information System) model. The White Paper introduces both, and further highlights other relevant issues with references to relevant literature for detailed discussion.

The second pillar of the White Paper presents the status quo at reUSE demonstrators. The analysis is underpinned by interviews and surveys at reUSE demonstrator organisations. reUSE demonstrator organisations stem from a range of organisational settings. Yet all follow the common goal of tying into the traditional publication process and collecting digital master files of public sector publications along with their print counterparts. Besides establishing the digital repository infrastructure at reUSE demonstrators, other core challenges are of an organisational nature including negotiations with information producers and the creation of feasible workflows for accessioning publications into the repository. In the rapidly ongoing implementation work, this analysis is of course only a snapshot of a transient situation. Yet, this analysis enables self-assessment and furthers discussion among the reUSE partners.

# Table of Contents

# 1 Introduction

reUSE is a collaborative project of national and university libraries, all committed to extending their services into the digital domain.[1] reUSE partners have the mission to collect, preserver, and make available print publications by public sector institutions, including national, regional and local administrative units, cultural and scientific associations, universities and educational organisations, as well as non-governmental organisations. These publications are documents based on text and images such as books, journals, reports, newspapers, working papers, and studies. Instead of only collecting the printed publications, reUSE partners aim to collect, maintain, and provide access to the respective digital master files as well. They thereby – and this is novel among ongoing digital preservation initiatives – tie into the traditional publishing process in order to cause minimum disruption and additional effort for the publishers and to attain a comprehensive digital collection. In addition to this, many public sector institutions are switching from print publications to e-only,[2] thereby making the collection and preservation of electronic publications an imperative task.

To this end, reUSE partners are implementing four demonstrator services at their institutions, which are meant to be extended towards trusted digital repositories and adopted for permanent operation after the reUSE project has finished. Each demonstrator reUSE service is integrated in the very institution's business environment and other digital activities, and the trusted repository is tailored to the respective organisational and technological environment. The service provides open access through online gateways and via the established library channels such as electronic catalogues.

The reUSE partners continuously seek to extend their digital holdings through active promotion of the service at public institutions. The goal is the acquisition of a critical mass of digital content that allows the creation of novel value added services. reUSE therefore not only improves accessibility of public sector information, but also it is a building block of a market place for digital content. The functionality of the reUSE service may – as an additional service feature – be opened to private publishers and other producers of print publications with a commercial stake in their products. While in most cases these publications cannot be made freely available, they could be incorporated in reUSE processes.

reUSE partners provide and update the organisational and technical framework best suitable for capturing the publication in tandem with necessary metadata. Close collaboration with the data providers will be the key in establishing efficient workflows and high quality services. reUSE partner institutions cooperate in the reUSE project in improving both accessibility and preservation of their digital collections.

With its unique opportunities and objectives, reUSE is funded through the *eContent* Programme of the European Commission.[3] reUSE partners are libraries and universities from Austria, Estonia, Germany and Slovenia. Demonstrator repositories are being set up at the National Library of Estonia, the Media and Library Center of the Humboldt University Berlin, the Austrian National Library, and the Austrian Literature Online (ALO) consortium consisting

[1] http://www2.uibk.ac.at/reuse/.
[2] The OCLC Five-Year Information Format Trends (2003) report identifies a clear trend towards the dissemination of documents in digital form only, and their publication on the public web.
OCLC (Online Computer Library Center): Five-Year Information Format Trends (2003), http://www.oclc.org/reports/2003format.htm.
[3] European Commission eContent Programme, http://www.cordis.lu/econtent/.

of the University Libraries Innsbruck and Graz, and the University Linz.[4] The National and University Library of Slovenia, Die Deutsche Bibliothek and the University of Ljubljana Faculty of Civil and Geodetic Engineering will evaluate these demonstrators in the course of the reUSE project.[5]

This White Paper provides both, an overview of current practical experiences and trends regarding digital repositories in international research, and a status check and cross-analysis of the ongoing implementation work at the reUSE demonstrators. The analysis is underpinned by interviews and surveys at reUSE demonstrator organisations. References to them as well as to external publications are featured throughout this paper. The White Paper aims to feed into demonstrator implementation, guide the development of the evaluation framework, and contribute to internal discussion and exchange of experience.

[4] See chapter 3 for a profile of the reUSE demonstrator repositories.
[5] An introduction of the reUSE partner institutions can be found in the Annex.

# 2 Digital Repositories

'Digital repositories' are an inclusive topic that embraces a variety of communities, methodologies, and technologies. This part of the White Paper aims to give a general overview of the field, offer starting points for further study, and highlight some selected issues of particular relevance to reUSE.

The concept of 'trust' and what it entails has gained primary importance recently. It bridges all communities that delve into digital repositories, and influences all approaches in their fundamental conception. The first chapter 'International Trends' aims to establish a picture of international initiatives, and to develop the concept of trust and what it means to reUSE. It is the objective of the reUSE partners to establish *trusted* repositories, and hence the concept is important for the next steps in the reUSE project and fundamental to this paper.

Similarly pivotal is the Open Archival Information System (OAIS), and much of the analysis in this paper is based on this standard reference model. Subsequent to the description of the OAIS, a synopsis of metadata presents a key building block of digital repositories. Other selected issues pertaining to digital repositories with a special view on interoperability and preservation complete this more theoretical environmental scan on issues and international developments surrounding repositories. Relevant literature and initiatives highlighted throughout this paper are useful starting points for going into more depth on any of the issues.

## 2.1 International Trends

Digital repositories come in many different forms and sizes. Moreover, the term is used with different meaning depending on community and situation. The concept of trust affects all initiatives that are committed to establishing a repository organisation that is successful in the long term. The following chapter develops reUSE's notion of the term 'repository' in the sections 'institutional repositories' and 'trusted repositories', and provides reUSE's perspective and terminology usage after those two sections.

### 2.1.1 via Institutional Repositories ...

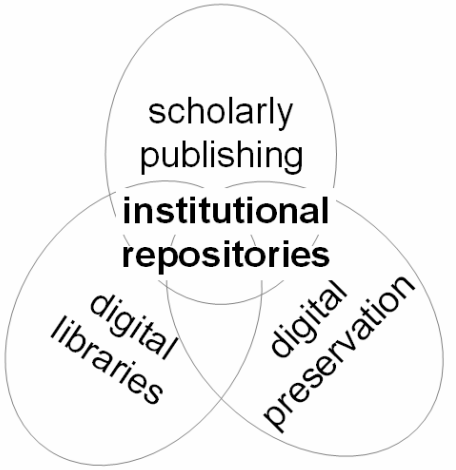
The novel information and communication technologies of the late 20th century, and the ensuing information revolution resulted in a massive surge in information production that is still ongoing.[6] Information is produced in education, research, government, business, and generally all walks of life. Companies and institutions have a stake in the information produced under their auspices, and therefore interest in systems for capturing and managing information soared in the last decade and is likely to rise further in the next.

Over the years the buzzword 'information management' grew into 'content management' and 'knowledge management'. Among numerous benefits these concepts promise organisations better control of their activities, close-knit structures through information exchange, and generally increased production at lower costs. Organisations all over the world took the promise and deployed electronic Document or Records Management Systems, and later

[6] Cf. Peter Lyman, Hal R. Varian: How much information? 2003 (October 2003), http://www.sims.berkeley.edu/how-much-info-2003.

Content or Knowledge Management Systems.[7] Despite a considerable overlap in functionality, these various systems all have their specific purpose and are tailored to the organisation's requirements.

'Institutional repositories' are another offspring of this development. The concept was coined in the higher education and research arena, where institutions such as the nuclear research centre CERN[8] or the Humboldt University[9] established repositories to capture, manage, and disseminate their institutional intellectual output. An institutional repository is typically installed at an organisation that consists of rather independent departments or programmes.[10] It is an independent service of an institution or a consortium of institutions, in contrast to Document Management Systems and others that are tightly integrated into the organisational workflow.




CERN and Humboldt University were among the early adopters, building tailor-made systems themselves. In the late 1990s, when the first institutions started their initiatives, institutional repositories were still rare and by and large had to be constructed from scratch. Therefore, the creation of a repository demanded a considerable input of software development work by skilled computer scientists, which only large institutions were able to afford. Only a few years later various initiatives embarked on developing general-purpose repositories in customisable software packages, and institutions around the world increasingly pick those packages up to establish their own repository. Today, the total number of active repository services at institutions world-wide surely is considerable.[11]

One of these open source institutional repository packages is the OPUS software[12] by the University of Stuttgart, which was first released in 1998, and which is used at about two dozen institutions in Germany. Probably the most widely spread institutional repository software is GNU EPrints[13], which was first released in the year 2000[14], and which is installed

[7] Richard E. Barry: Managing Distinctions: Enterprise Information, Document, Records, Knowledge and Content Management. In: Records and Information Management Review, 18:2 (2002), http://www.mybestdocs.com/barry-r-im-rm-distinctions.htm.
[8] CERN – European Organization for Nuclear Research, http://www.cern.ch. CERN Document Server (CDS), http://cds.cern.ch/.
[9] Humboldt University, http://www.hu-berlin.de/. edoc – Dokumenten- und Publikationsserver der Humboldt-Universität zu Berlin, http://edoc.hu-berlin.de/.
[10] Note the inclusive definition of 'institutional repositories' assumed here, in contrast to other definitions that confine the use of institutional repositories to universities. Cf. Clifford Lynch: Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. ARL Bimonthly Report 226 (February 2003), http://www.arl.org/newsltr/226/ir.html, and Raym Crow: The Case for Institutional Repositories. A SPARC Position Paper (August 2002), http://www.arl.org/sparc/IR/ir.html.
[11] DOAR – the Directory of Open Access Repositories, a collaboration between the University of Nottingham (UK) and the University of Lund (Sweden) will provide a comprehensive list of institutional and subject-based repositories, http://www.opendoar.org/. See also the Institutional Archives Registry at http://archives.eprints.org/ and the list of repositories conforming to the Open Archives Protocol for Metadata Harvesting (OAI-PMH) at http://www.openarchives.org/Register/BrowseSites/.
[12] OPUS – Online Publikationsverbund der Region Stuttgart, http://elib.uni-stuttgart.de/opus/.
[13] EPrints.org – Self-Archiving and Open Access (OA) Eprint Archives, http://www.eprints.org/.

at almost 150 institutions by early 2005. Two very powerful software packages with a wide area of applicability, also outside the institutional repository arena, are Fedora[15] and DSpace[16]. Fedora is developed cooperatively by the University of Virginia and Cornell University, and is most renowned for its open and flexible architecture. DSpace by the Massachusetts Institute of Technology (MIT) had a massive impact when it was launched in 2002 for its powerful customisation and community functionality. The DSpace Federation is probably the most active open source developer community among current digital repository software, and it potentially serves as a digital preservation forum that goes beyond technological issues.[17]

Overviews of these and other institutional repository software packages include the ones provided by the Open Society Institute[18], NESTOR[19], and the Open Archives Forum[20].

SURVEY DATA

(cf. chapter 3) The reUSE demonstrator at the National Library of Estonia builds on Fedora software, whereas the Austrian National Library has a commercial system, and the systems at Humboldt University and by the ALO Consortium are self-made.

Institutional repositories come in a range of different flavours. In most cases the repository service is targeted at the staff or members of an organisation. Even in consortia consisting of multiple organisations such as the University of California eScholarship Repository[21] the target community is clearly delimited. With the specific needs of its community in mind, the repository service may restrict its collections to specific topical areas, to specific data formats, or may impose other restrictions to facilitate repository management and increase accessibility. Some repositories have access restrictions for copyright or other reasons,

[14] Robert Tansley and Stevan Harnad: Eprints.org Software for Creating Institutional and Individual Open Archives. In: D-Lib Magazine – In Brief, 6:10 (October 2000), http://www.dlib.org/dlib/october00/10inbrief.html#HARNAD.
[15] The Mellon FEDORA (Flexible Extensible Digital Object and Repository Architecture) Project, University of Virginia and Cornell University, http://www.fedora.info/. Cf. Thornton Staples, Ross Wayland, Sandra Payette: The Fedora Project. An Open-Source Digital Object Repository Management System. In: D-Lib Magazine 9:4 (April 2003), http://www.dlib.org/dlib/april03/staples/04staples.html.
[16] DSpace, Massachusetts Institute of Technology (MIT), http://www.dspace.org/. Cf. MacKenzie Smith et al.: DSpace. An Open Source Dynamic Digital Repository. In: D-Lib Magazine 9:1 (January 2003), http://www.dlib.org/dlib/january03/smith/01smith.html.
[17] Julie Walker: DSpace, digital preservation, and business models. Presentation at the ERPANET Seminar "Business Models related to Digital Preservation", Amsterdam (September 20–22, 2004), http://www.erpanet.org/events/2004/amsterdam/presentations/erpaTraining-Amsterdam_Walker.pdf, and Jim Downing: DSpace@Cambridge. Planned Digital Preservation R&D. Presentation at the BL/CURL DPC Forum, London (October 19, 2004), http://www.dpconline.org/docs/events/041019downing.pdf. Cf. the DSpace Wiki at http://wiki.dspace.org/.
[18] Open Society Institute (OSI): A Guide to Institutional Repository Software. 3rd Edition, August 2004, http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf.
[19] Uwe M. Borghoff et al.: Vergleich bestehender Archivierungssysteme [Comparison of existing archiving systems (in German)] (= nestor - materialien 3 – 2005), http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-20050117016.
[20] Open Archives Forum: Information Resource Database: List of Repositories, http://www.oaforum.org/oaf_db/list_db/list_repositories.php.
[21] University of California, California Digital Library: eScholarship Repository, http://repositories.cdlib.org/escholarship/.

though overall institutional repositories associate with the Open Access movement[22] and support free access to their resources. Ready-made repository software can accommodate these and similar policy decisions. The open access movement is also key driver of interoperability between the individual repositories, and – as a further evidence of this – a majority of institutional repository software incorporate modules for OAI[23] functionalities.

While a lot has been achieved in the last couple of years, technologies are all but stable and best practices are only beginning to emerge. After an initial focus on technology in early institutional repository initiatives, managerial issues have come to the fore. In the SHERPA[24] project, for example, a range of British universities cooperate to investigate policy, business models, intellectual property rights, quality control and other managerial issues. This broad cooperation among major British universities is exemplary to other initiatives in any field.

SURVEY DATA

(R 3.7.1) All reUSE demonstrator organisations accept responsibility for the **long-term maintenance of digital resources** on behalf of its depositors and for the benefit of current and future users.

With the focus on managerial issues long-term strategic planning and considerations regarding the stability of the collections came along. Resources collected in institutional repositories are of long-term value, and measures for their preservation are increasingly urgent.[25] A study[26] commissioned by the JISC[27] in the UK found that the preservation challenges posed by the static, mainly textual, content of institutional repositories are rather of an organisational than a technological nature. Current projects such as eSPIDA[28] and SHERPA DP[29] are dedicated to preservation issues in repositories, and indeed their focus is on the development of strategy and business models for moving institutional repositories from project funding to sustainability. Both projects alongside a multitude of other British projects in this area are part of the current JISC

[22] The Open Access Initiative pushes towards free public access to research. Cf. Peter Suber: Open Access Overview, http://www.earlham.edu/~peters/fos/overview.htm (constantly updated) and Charles W. Bailey, Jr.: Scholarly Electronic Publishing Bibliography, http://info.lib.uh.edu/sepb/sepb.html (constantly updated). Cf. Open Society Institute: Budapest Open Access Initiative, http://www.soros.org/openaccess/.

[23] OAI – Open Archives Initiative, http://www.openarchives.org/.

[24] SHERPA – Securing a Hybrid Environment for Research Preservation and Access, http://www.sherpa.ac.uk/, funded by the JISC FAIR Programme: Focus on Access to Institutional Resources, http://www.jisc.ac.uk/index.cfm?name=programme_fair. Cf. Stephen Pinfield: Open Archives and UK Institutions. An Overview. In: D-Lib Magazine 9:3 (March 2003), http://www.dlib.org/dlib/march03/pinfield/03pinfield.html.

[25] See section 2.4.2 on Preservation for a more extensive discussion.

[26] Hamish James, Raivo Ruusalepp, Sheila Anderson, Stephen Pinfield: Feasibility and Requirements Study on Preservation of E-Prints. Report Commissioned by the Joint Information Systems Committee (JISC), October 2003, http://www.jisc.ac.uk/uploaded_documents/e-prints_report_final.pdf.

[27] Joint Information Systems Committee (JISC), http://www.jisc.ac.uk/.

[28] eSPIDA: An effective Strategic model for the Preservation and disposal of Institutional Digital Assets, http://www.gla.ac.uk/espida/index.shtml. Project under the JISC Programme "Supporting Digital Preservation and Asset Management in Institutions"; cf. William Nixon: From ePrints to eSPIDA: Digital Preservation at the University of Glasgow. Presentation at the DPC Forum: Digital Preservation in Institutional Repositories; London (19 October 2004), http://www.dpconline.org/docs/events/041019nixon.pdf.

[29] SHERPA Digital Preservation: Creating a Persistent Preservation Environment for Institutional Repositories, http://ahds.ac.uk/about/projects/sherpa-dp/. Project under the JISC Programme "Supporting Digital Preservation and Asset Management in Institutions".

programme "Supporting Digital Preservation and Asset Management in Institutions"[30] and the forthcoming JISC Digital Repositories programme[31], which both support conceptual research or active implementation on institutional repositories and digital preservation. Technology development pursues digital preservation as well, and for example repository software developers including DSpace[32] and Fedora[33] are working towards increased preservation capabilities of their systems.

Notable initiatives outside Great Britain include DARE[34] and DiVA[35]. Both are consortia of university libraries and other institutions in the Netherlands and the Nordic countries respectively, which establish an interoperable repository infrastructure for capturing the intellectual output of the institutions. Besides cooperating amongst themselves on all levels, they cooperate with their national libraries on the preservation of their collections.[36] Another notable initiative is DINI[37] in Germany, which establishes standards for institutional repositories, fosters interoperability, and facilitates discussion and collaborative development.

In these and other current initiatives a general trend towards cooperation is clearly discernable, and after a proliferation of numerous small initiatives there now is a concentration towards larger consortia. Increased cooperation may thrust the initiatives forward on their way towards long-term strategic planning and sustainability. The Canadian Association of Research Libraries formulated the vision of a "larger global system of distributed, interoperable repositories"[38], and this is the shared goal of most institutional repository initiatives.

### 2.1.2 ... towards Trusted Repositories

The notion of a "Trusted Digital Repository" became widespread with the report "Trusted Digital Repositories: Attributes and Responsibilities"[39], which was prepared in 2002 by the

[30] Supporting Digital Preservation and Asset Management in Institutions. JISC programme, October 2004 – September 2006, http://www.jisc.ac.uk/index.cfm?name=programme_404.
[31] JISC Digital Repositories Programme, cf. http://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind0412&L=digital-preservation&T=0&F=&S=&P=183.
[32] Cf. Fn 17.
[33] Cf. Carl Lagoze, Sandy Payette, Edwin Shin, Chris Wilper: Fedora: An Architecture for Complex Objects and their Relationships (Preprint, 7 Jan 2005), http://www.arxiv.org/ftp/cs/papers/0501/0501012.pdf. See also the proposal to the Andrew W. Mellow Foundation (Fedora Phase 2), http://www.fedora.info/documents/fedora2_final_public.shtml.
[34] Digital Academic REpositories (DARE), http://www.surf.nl/en/themas/index2.php?oid=7.
[35] Digitala Vetenskapliga Arkivet [Digital Scientific Archive], http://www.diva-portal.se/.
[36] Annemiek van der Kuil, Martin Feijen: The Dawning of the Dutch Network of Digital Academic REpositories (DARE): A Shared Experience. In: Ariadne 41 (30 October 2004), http://www.ariadne.ac.uk/issue41/vanderkuil/intro.html; Eva Müller, Uwe Klosa, Peter Hansson, Stefan Andersson: Archiving Workflow between a Local Repository and the National Archive. Experiences from the DiVA Project (August 2003), http://bibnum.bnf.fr/ecdl/2003/proceedings.php?f=muller.
[37] DINI – Deutsche Initiative für Netzwerkinformation [German Initiative for Networked Information], http://www.dini.de/.
[38] Canadian Association of Research Libraries: Access to Research Information: A Critical Component for Canada's Innovation Strategy. Brief on the Government of Canada's Innovation Strategy (September 2002), http://www.carl-abrc.ca/projects/innovstrat/access.htm.
[39] RLG Working Group on Digital Archive Attributes: Trusted Digital Repositories: Attributes and Responsibilities. RLG, OCLC Report. Mountain View CA 2002, http://www.rlg.org/longterm/repositories.pdf.

RLG[40]/OCLC[41] Working Group on Digital Archive Attributes[42]. Library institutions from the USA, Europe, and Australia contributed to this report which originally was targeted at cultural organizations with large-scale, heterogeneous collections. Today it permeates all types of organisations, and influences international thinking regarding digital preservation.

Digital repositories generally aim to provide reliable, long-term access to managed digital resources. A viable technological approach tailored to the specific situation needs to be implemented to achieve this primary goal. Long-term digital preservation, however, is as much an organisational challenge as a technological one. The "Trusted Digital Repositories" report establishes attributes and responsibilities of long-term reliable digital repositories. Thereby, attributes of a Trusted Digital Repository as articulated in the report include administrative responsibility, organisational viability, financial sustainability, technological and procedural suitability, system security, and procedural accountability. Responsibilities are a number of high-level organisational and curatorial responsibilities, as well as operational accountability. The authors of the report argue that only an adequate organisational and policy environment can ensure the chance for successful long-term preservation.

In addition to the primary goal of establishing a long-term reliable digital repository, the repository organisation needs to establish trust in order to achieve its goals. Information producers should have the confidence to entrust their digital resources to the repository for long-term preservation, as well as users should be able to trust in the reliability of the repository and the authenticity of its content. Repository management in turn needs trust in its peers for cooperation and in third party providers for outsourcing certain functions. Overall, the successful operation of a digital repository and its anchoring in the target community is inextricably tied to trust.

But how can trust be attained? Traditional libraries and other cultural heritage institutions had decades and perhaps even centuries to prove their trustworthiness and to become trusted in the analogue realm. Such a long time is not available to digital archives. Instead, the "Trusted Digital Repositories" report suggests a certification programme as means to establish the trustworthiness of an organisation. Such a certification scheme would allow to scrutinise the policy framework of the digital repository, and the adequate implementation of the trusted repository attributes and responsibilities. Preservation initiatives picked up this suggestion: research projects work towards establishing certification criteria and procedures, some large organisations aim to establish certification authorities, and many digital repositories plan to certify their repository as soon as possible. To view some of these initiatives, for example the German DINI initiative hosts a certificate[43] for institutional repositories in higher education, which has been welcomed by relevant initiatives in German speaking countries. The competence network NESTOR builds on DINI's experience in its workgroup on certification of trusted repositories, which aims to establish transparency between stakeholders in the German preservation

SURVEY DATA

(R 3.9/10) All reUSE demonstrator organisations plan to **certify** their repository as a **trusted digital repository** as soon as such a certification scheme is available. The Humboldt Document and Publication Server already is DINI certified.

[40] RLG – Research Libraries Group, http://www.rlg.org/.
[41] OCLC – Online Computer Library Center, http://www.oclc.org/.
[42] RLG/OCLC Digital Archive Attributes Working Group, http://www.rlg.org/longterm/attribswg.html.
[43] Arbeitsgruppe Elektronisches Publizieren: DINI Zertifikat Dokumenten- und Publikationsserver [DINI Certificate for Open Access Document Servers], http://www.dini.de/documents/Zertifikat.pdf.

landscape.[44] Also the newly established Digital Curation Center (DCC) in the UK is developing auditing and certification processes for trusted repositories.[45] RLG together with NARA[46] convened a Digital Repository Certification Task Force[47] that works towards general certification requirements applying to any type of digital repository. This effort is part of ongoing work with the Open Archival Information System (OAIS) model and intended to go into an ISO[48] standardization process.

All certification initiatives face the challenge of defining requirements that may be elusive to objective metrics, which encumbers an unbiased audit as well as comparison between distinct archives. Since trusted repository requirements are primarily on a policy level and since they must fit a range of different repository types, the certification requirements and the audit process may need to be adapted to the specific situation. These and similar issues remain to be dealt with on the way towards trusted digital repository certification.

While it will take some more time until a certification scheme is provided for any type of digital repository and sufficiently known by relevant stakeholders, the concept of a 'trusted digital repository' has impressed the importance of organisational issues in digital preservation and guides many initiatives towards sustainable structures and an adequate policy framework already now.

### 2.1.3 The reUSE Perspective

reUSE partners are from a range of organisation types and include university libraries and national libraries. National libraries serve the whole of a country's society, and their targeted user group is more diverse than that of a university library. Also the relationship with information producers differs to the situation in higher education (as will be described in the following chapter).

For these reasons the paper in the following uses the term 'repository', and it only uses 'institutional repository' in the case of a repository focused on a specific topic or organisation. The terms 'archive' and 'repository' are used interchangeably.

In the broad field of digital preservation 'repository' has various meanings. For example, 'storage repository' and 'trusted repository' reflect totally different concepts, the former concentrated on hardware and related functions, and the latter reflecting a conceptually high and holistic view. reUSE partners aim to establish 'trusted repositories' that cover organisation-specific responsibilities starting at the production and collection of digital resources to their preservation and access. The paper therefore uses the term in this holistic and rather general sense of an organisation intending to steward and preserve its (digital) resources.

[44] NESTOR Working Group on Trusted Repositories Certification, http://nestor.cms.hu-berlin.de/tiki/tiki-index.php?page=Working+Group+on+Trusted+Repositories+Certification+%28nestor%29.
[45] Cf. http://www.dcc.ac.uk/research.html.
[46] NARA – National Archives and Records Administration, USA, http://www.archives.gov/.
[47] RLG/NARA Digital Repository Certification Task Force, http://www.rlg.org/en/page.php?Page_ID=367.
[48] ISO – International Organization for Standardization, http://www.iso.org/.

## 2.2 OAIS – reference and common ground

One of the key attributes of a trusted digital repository according to the RLG/OCLC report is "Compliance with the Reference Model for an Open Archival Information System (OAIS)".[49] The OAIS is fundamental in the digital repository arena as a conceptual framework, a plug-in model and reference. Upcoming chapters repeatedly refer to the model as a means to establish structure and common ground, which all relevant stakeholders can identify with. The OAIS is therefore briefly introduced here below. Obviously only a short abstract of the complex standard can be given for the purposes of this paper. More details can be found in the OAIS standard itself[50] or in digests such as the excellent introductory guide[51] commissioned by the Digital Preservation Coalition[52] in the UK.

The *Open Archival Information System* reference model was developed by the CCSDS, the Consultative Committee for Space Data Systems[53]. While the standard originates from the space data community, it is not confined to the requirements of and the situation at space agencies. The CCSDS created the OAIS starting from the early 1990s in an open consultation process that encouraged comments from other communities and is committed to maintaining the standard in the same open manner. (This is also the origin of the 'open' in the name of the standard.) As part of this open process, also the library community contributed to the OAIS, and particularly NEDLIB[54] – a collaborative project of eight European national libraries from 1998 to 2000 – gave extensive feedback to early versions of the OAIS.[55] In 2002 the OAIS has been adopted as ISO standard 14721[56].

An OAIS is understood to as an organisation of people and systems charged with the task of preserving information over the long-term and making it accessible to a specific class of users (known as the “designated community”). The OAIS reference model describes the activities in an archival system and identifies both internal and external interfaces to the archive functions. It is a conceptual framework that is applicable to any type of archival object (including physical artefacts in a hybrid archive), and it does not pre-empt any specific implementation or system design. A single organisation may have multiple OAIS archives with different missions and/or distinct designated communities. While each archive needs to provide all functions as specified in the OAIS standard, specific functions could be shared between archives or even between different organisations following the modularity of the OAIS reference model.

---

[49] Trusted Digital Repositories (Fn. 39), p. 13.

[50] Consultative Committee for Space Data Systems: Reference Model for an Open Archival Information System (OAIS), http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html. In the following, this White Paper refers to the version of the OAIS model which has been adopted as ISO 14721:2003: CCSDS 650.0-B-1: Reference Model for an Open Archival Information System (OAIS). Blue Book, Issue 1. January 2002, http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf.

[51] Brian F. Lavoie: The Open Archival Information System Reference Model: Introductory Guide. DPC Technological Watch Report (January 2004), http://www.dpconline.org/docs/lavoie_OAIS.pdf.

[52] DPC – Digital Preservation Coalition, http://www.dpconline.org/.

[53] CCSDS – Consultative Committee for Space Data, http://public.ccsds.org/default.aspx.

[54] NEDLIB – Networked European Deposit Library, http://www.kb.nl/coop/nedlib/.

[55] Cf. NEDLIB Contribution to the Review of the OAIS, http://www.kb.nl/coop/nedlib/results/OAISreviewbyNEDLIB.html.

[56] ISO 14721:2003. Space data and information transfer systems – Open archival information system – Reference model, http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=24683&ICS1=49&ICS2=140&ICS3.

One of the OAIS' core achievements is the concepts and terminology it develops. This facilitates communication and comparison of approaches to long-term digital preservation across domains and organisation-specific boundaries. Therefore the OAIS guides research and implementation in the whole digital preservation community.

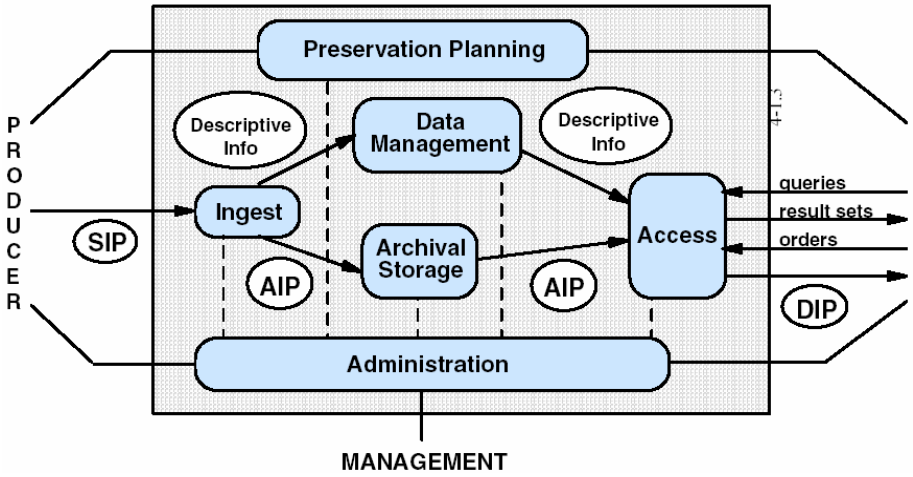



**Figure 1: OAIS Functional Model**

This well-known graphic from the OAIS model[57] describes the core components of the OAIS and defines their respective tasks and responsibilities. These six high-level functional entities include the external interfaces to the data producer in "Ingest" and to the user in "Access", as well as the OAIS-internal entities "Data Management" and "Archival Storage", together with the strategic entities "Preservation Planning" and "Administration".

In addition to this functional model the OAIS also provides an information model that defines the information objects which are managed by the archive. An archival item is accessioned into the archive by the producer in the form of a "Submission Information Package (SIP)", preserved in the archive as an "Archival Information Package (AIP)", and delivered to the user community as a "Dissemination Information Package (DIP)". An Information Package consists of a digital object bound together with different types of metadata required for preservation and access.

Numerous initiatives were guided by the OAIS reference model, and some current repository implementations even claim to conform to the OAIS[58]. Yet, how OAIS compliance can be objectively measured remains to be established. Similar to the concept of a "trusted digital repository" the OAIS is sufficiently high-level to lack clear-cut criteria for measuring compliance. In fact, there are numerous conceivable ways for implementing OAIS functions and in the end an OAIS implementation is highly situation-specific. Projects are in progress to evaluate OAIS compliance at specific organisations,[59] and an agreed evaluation procedure for any type of digital repository will be a valuable component of trusted digital repository certification.

This White Paper was guided by the OAIS reference model in various ways. It adopts some OAIS-specific terminology, such as "designated community" or "ingest", without assuming

[57] OAIS (Fn. 50), p. 4-1.
[58] A respective short-list is available from the RLG website "OAIS-Modelled Repositories", http://www.rlg.org/en/page.php?Page_ID=377.
[59] For example: Assessment of UK Data Archive and The National Archives compliance with OAIS/METS, http://www.jisc.ac.uk/index.cfm?name=project_oais.

that all readers are familiar with the respective concepts. Furthermore, system descriptions in this paper are structured according to the OAIS functional model for ease of comparison.

Standards including the OAIS are essential components in building trusted digital repositories. The OAIS models on a high conceptual level, yet standards on a procedural and technological level are equally needed. Metadata is one of the building blocks of any digital repository, and the upcoming section outlines the relevant standards and international developments in this field.

## 2.3 Metadata – repository building block

The previous sections have shown that initiatives do not need to start from scratch when establishing their digital repositories. In addition to the experiences of the numerous initiatives in the field, repository development can build on fundamental building blocks including international standards, base technologies, and metadata. This section introduces relevant concepts and developments with regard to metadata, and – as all other sections in this paper – references international standards and base technologies relevant to the field.

Metadata, chiefly defined as (structured) data about data, have been used in libraries for centuries.[60] Traditional library catalogues bear the concepts and roots of what is called descriptive metadata today. In addition to catalogues that index library collections and paper documentation attached to collection items, a variety of meta information is inherent in the organisation of libraries. Perhaps a library building has specific sections for different subject areas, or confidential areas that are not open for public access. Also these structures and the way library business models are enforced carry a variety of administrative and rights metadata.

In the digital world, metadata has become a central paradigm. Existing types of metadata were augmented, new metadata types gearing at the specifics of the digital environment were adopted, and the concepts were applied in other fields creating community-specific metadata sets. In digital collections metadata fulfil a variety of tasks, including:

- describing collection items (e.g. author, creation date);
- supporting retrieval (by subject browsing and metadata search, as well as full text search) and identification;
- managing collection items within a digital library system;
- establishing their context (e.g. reference to accompanying documents, documentation of the circumstances of their creation);
- preserving their authenticity (e.g. document the provenance and context of the publication, record the history trail of the item in the archive);
- protecting the integrity of archival items (against external manipulation or unintentional corruption);
- enforcing a business model and rights management (e.g. only permit single parallel access for specific user groups at a certain location);
- facilitating interoperability (between independent, distributed digital archives);

[60] Jessica Milstead and Susan Feldman: Metadata: Cataloguing by Any Other Name ... In: ONLINE Magazine, January 1999, http://www.onlinemag.net/OL1999/milstead1.html.

- identification of the technical nature of the digital object for preservation purposes.

All these tasks are supported by three main types of metadata: descriptive, structural, and administrative metadata.[61] Descriptive metadata obviously describe a resource for discovery and identification. Structural metadata indicate the composition of items, for example chapters and annexes in books, or the aggregation of images and textual elements in a hand-out. Administrative metadata facilitate the management of a resource; rights management metadata, technical metadata and preservation metadata are part of this category. Metadata can describe resources at any level of aggregation, for instance on a data file level, on the level of a compound object that may consist of several files, or on a collection level.

Metadata schemas specify the elements and structure of a metadata set, and they may be limited to a specific application in a specific organisation or span a whole community. Metadata schemas can be defined on different levels. A slim framework may only list a number of mandatory or voluntary elements without any further specifications. Thereby it is essential to clearly describe the meaning and scope of each element, since different communities may associate inherently different meanings with the same word; for example a 'creator' may be an author or, in a different context, a composer or even an estate agent who creates a customer profile in a database. On another level, a metadata schema could define the metadata syntax and structure. The METS[64] standard, for instance, is a plug-in model for encoding descriptive, administrative, and structural metadata, and its syntax is encoded in XML Schema[65]. METS does not define metadata elements but remains on a high structural level. Organisational metadata schemas would then comply with METS and plug-in other metadata standards with a suitable element set.

REALITY CHECK

(cf. OAIS mapping, Preservation Description Information; chapter 3.5.2.2)

All reUSE partners acknowledge the METS standard to be a useful metadata framework that is influential in the digital repository arena, and they all comply with METS. Other metadata standards, however, vary greatly between the partners. For example, with regard to technical metadata for images the ALO consortium applies DIG35[62] whereas the Austrian National Library uses MIX[63], and the other two demonstrators have not yet defined their respective metadata requirements. However, metadata schemes at reUSE demonstrators are still open for refinement.

In addition to a range of new tasks and areas of applicability, digital processes of course brought a revolution in the creation and management of metadata. In the digital world metadata can be copied, exchanged and linked, inherited between different resources, they can even be automatically transcribed between different forms of representation, and some metadata types can be created automatically in the course of a specific action on a digital resource and retained for future reference.

[61] National Information Standards Organization (NISO): Understanding Metadata. Bethesda: NISO Press 2004, http://www.niso.org/standards/resources/UnderstandingMetadata.pdf.
[62] DIG35 Specification. Metadata for Digital Images, http://www.i3a.org/i_dig35.html.
[63] MIX – NISO Metadata for Images in XML Schema (Library of Congress), http://www.loc.gov/standards/mix/. The MIX Schema is based on NISO Z39.87, Data Dictionary – Technical Metadata for Digital Still Images, http://www.niso.org/standards/standard_detail.cfm?std_id=731.
[64] METS – Metadata Encoding & Transmission Standard (Library of Congress), http://www.loc.gov/standards/mets/.
[65] World Wide Web Consortium: XML Schema, http://www.w3.org/XML/Schema.

Despite powerful processing techniques, metadata creation and management continue to draw a lot of effort. The next two sections will dwell on issues including standards, interoperability, workflow and automation from a metadata perspective. While the next chapter puts an emphasis on descriptive metadata and the one thereafter on preservation metadata, the abstract issues are relevant to all types of metadata. Moreover, the issues are essential to digital repositories in general terms as well, and they are also discussed in other contexts in the chapters below.

### 2.3.1 Metadata convergence, from the viewpoint descriptive metadata

As mentioned above, descriptive metadata is the type most similar to traditional library cataloguing. It includes elements such as author and title, and it is essential for resource discovery and identification. Best known among relevant standards is the Dublin Core Metadata Element Set (DC)[66], which was developed by a global forum in an open discussion and consensus process. Its main goal is interoperability and broad applicability, and it is hence designed to be lean and simple: it consists of only 15 loosely defined elements. A bibliographic metadata set that is geared at the library sector is MARC (Machine Readable Cataloguing)[67]. MARC (MARC21) is a widely spread international standard maintained by the US Library of Congress, though mostly derivates of the standard are in use all over the world. MODS (Metadata Object Description Schema)[68] and MARCXML[69] further define an XML syntax for MARC21 records.

DC, MARC, MODS, and MARCXML only highlight some of the descriptive metadata schemas out there. Others may be more or less comprehensive, dedicated to a specific application or a specific community, such as the Encoded Archival Description (EAD)[70] that stems from the archives community, or CSDGM[71] for digital geospatial metadata.

Standards are essential in the metadata landscape. Obviously, ready-made metadata schemas that fit the local requirements take from that organisation the burden to design a new schema. The combined expertise put into the definition of international standards is thereby of enormous value. Beyond being a valuable reference, standards compliance is imperative for facilitating interoperability. Distributed digital repositories may exchange resources in a meaningful way, if one system understands the metadata of the other. Moreover, they may exchange only metadata without the actual resources in order to link distinct digital collections.

However, metadata 'standards' are proliferating, and they come in many forms and sizes. Organisations rarely find a single standard that exactly fits their requirements. A perfect fit is thereby important, since a metadata scheme should obviously satisfy at least all corporate requirements, but at the same time it should be rather small for the inherent cost of metadata creation and stewardship.[72] The result is the birth of yet another metadata schema or the

[66] Dublin Core Metadata Initiative, http://www.dublincore.org/.
[67] MAchine Readable Cataloguing (MARC), http://www.loc.gov/marc/marc.html.
[68] Metadata Object Description Schema (MODS), http://www.loc.gov/standards/mods/.
[69] MARC 21 in XML, http://www.loc.gov/standards/marcxml/.
[70] Encoded Archival Description (EAD), http://www.loc.gov/ead/.
[71] Federal Geographic Data Committee: Content Standard for Digital Geospatial Metadata (CSDGM). FGDC-STD-001-1998 (June 1998), http://www.fgdc.gov/metadata/contstan.html.
[72] ERPANET: Getting what you want, knowing what you have and keeping what you need – Metadata in digital preservation. Final Report of the ERPANET Training Seminar: Metadata in Digital

extension of an existing standard. This proliferation of metadata schemas obviously dwarfs interoperability aspirations, as metadata need to match on a syntactic, structural, as well as semantic level to be interoperable.

To reconcile the rift between standards compliance, interoperability, and corporate requirements, organisations assume a metadata schema that fits their needs and provide for automatic mapping of their metadata onto a standard representation. This is possible if the corporate metadata is sufficiently detailed to satisfy target metadata standards, and if the metadata system is defined adequately for machine processing. Thereby, for instance, a MODS metadata set can be automatically mapped to DC (obs: though not the other way round).[73] However, automatic mapping between specific metadata sets should be accounted for from the outset of creating a metadata framework, as otherwise cross-walks between different metadata sets may lack quality or be impossible to achieve without the investment of resource-intensive manual input.

In another practical application DC metadata may be automatically gleaned from a corporate metadata framework and shared with other digital repositories. Sharing of descriptive metadata in this way has become particularly popular in the context of the open access movement. Various protocols including the OAI-PMH[74], Z39.50[75], or as of late SRW/SRU[76] have been developed to support the exchange of metadata, distributed searching, and the linkage of repositories in this way.

SURVEY DATA

(O 7.3) For reUSE partners the URN is the persistent identifier of choice. The Austrian National Library and Humboldt University Library already use an URN identification scheme based on the NBN (National Bibliography Number), and ALO and the National Library of Estonia plan to deploy it.

When sharing resources and linking between distributed digital repositories, the unique identification of these resources needs to be sorted out. Identification schemes such as an URL enable direct access to digital resources at their source. However, digital resources have proven to be extremely volatile as they may easily change their location or vanish completely, and particularly the URL turned out to be unreliable.[77] For this reason the e-publishing industry as well as the (digital) library community initiated Persistent Identifier schemes, which promise to facilitate resource identification in the long-term and despite the possible move of digital resources other locations.[78] A variety of persistent identifiers and their

Preservation. Marburg (3–5 September 2003), http://www.erpanet.org/events/2003/marburg/finalMarburg%20report.pdf.

[73] Günter Waibel: Like Russian Dolls: Nesting Standards for Digital Preservation. In: RLG DigiNews 7:3 (June 15, 2003), http://www.rlg.org/preserv/diginews/v7_n3_feature2.html.

[74] Open Archives Initiative Protocol for Metadata Harvesting, http://www.openarchives.org/.

[75] Z39:50 Information Retrieval: Application Service Definition and Protocol Specification. Library of Congress International Standard Maintenance Agency, http://www.loc.gov/z3950/agency/; cf. http://www.niso.org/z39.50/z3950.html.

[76] SRW/SRU - Search and Retrieve Web Service / Search and Retrieve URL Service, http://www.loc.gov/z3950/agency/zing/srw/.

[77] Wallace Koehler: A Longitudinal Study of Web Pages Continued: A Consideration of Document Persistence. In: Information Research 9:2 (January 2004), http://informationr.net/ir/9-2/paper174.html.

[78] Giuseppe Vitiello: Identifiers and Identification Systems. An Informational Look at Policies and Roles from a Library Perspective. In: D-Lib Magazine 10:1 (January 2004), http://www.dlib.org/dlib/january04/vitiello/01vitiello.html. Diana Dack: Persistent Identification Systems. Consultancy report for the National Library of Australia, May 2001, http://www.nla.gov.au/initiatives/persistence/PIcontents.html. For the perspective of the publishing

resolution services with different technical characteristics and terms of use have emerged since, including the URN[79], the Handle System[80], the DOI[81], the PURL[82], or the ARK[83]. While being part of descriptive metadata, persistent identification is of great importance in the context of digital preservation.[84] Further metadata requirements with regard to digital preservation are discussed in the following.

### 2.3.2 Joining the ends through cooperation; preservation metadata

Recent years have seen rising activity in the area of preservation metadata, with a myriad of proposals and lively open discussion.[85] Although preservation metadata above was pigeonholed as a sub-category of administrative metadata, it interacts strongly with all other metadata and it is therefore not easy to classify. Sometimes preservation metadata merely signifies the technical descriptions of a specific digital file, including the file format version or the resolution of an image. This kind of metadata is often labelled as 'technical metadata'. On its most general however, preservation metadata is considered any metadata schema that is geared at supporting the management of digital resources over time, and as such it may be an overall concept that influences other types of metadata as well. A variety of initiatives move within this continuum with their proposals and design issues. The area is still very young and very dynamic with few (stable) preservation metadata schemas operational.

Pioneering initiatives regarding preservation metadata have been active from the year 2000, including the NEDLIB project[86], Cedars[87], and – on a more theoretical level – the first OCLC/RLG Working Group on Preservation Metadata[88]. These initiatives all independently chose the OAIS model as a general reference framework for the creation of their metadata

industry cf. Steve Sieck: Using the DOI to Improve Profitability in Publishers' E-Commerce Operations, December 2003, http://doi.contentdirections.com/eps/sieck1.pdf.

[79] URN – Uniform Resource Names, http://www.ietf.org/html.charters/OLD/urn-charter.html, see also: World Wide Web Consortium: Naming and Addressing: URIs, URLs, ..., http://www.w3.org/Addressing/. For the namespace URN:NBN see http://www.persistent-identifier.de/?lang=en.

[80] The Handle System, http://www.handle.net/.

[81] DOI – Digital Object Identifier System, http://www.doi.org/. Cf. Norman Paskin, DOI. A 2003 Progress Report, In: D-Lib Magazine 9:6 (2003), http://www.dlib.org/dlib/june03/paskin/06paskin.html

[82] PURL – Persistent Uniform Resource Locator, http://purl.org/

[83] ARK – Archival Resource Key, http://www.cdlib.org/inside/diglib/ark/. Cf. John A. Kunze: Towards Electronic Persistence Using ARK Identifiers, http://www.cdlib.org/inside/diglib/ark/arkcdl.pdf.

[84] Cf. the resources available at the ERPANET Seminar on Persistent Identifiers (Cork, June 17–18, 2004), http://www.erpanet.org/events/2004/cork/index.php.

[85] Michael Day: Preservation metadata initiatives: practicality, sustainability, and interoperability. (October 2004) In: Frank M. Bischoff, Hans Hofman, Seamus Ross (ed.): Metadata in Preservation: Selected Papers from an ERPANET Seminar at the Archives School Marburg, 3–5 September, 2003, (Veröffentlichungen der Archivschule Marburg, Institut für Archivwissenschaft, 40), Marburg: Archivschule Marburg, pp. 91–117, http://www.ukoln.ac.uk/preservation/publications/erpanet-marburg/day-paper.pdf.

[86] Catherine Lupovici, Julien Masanès: Metadata for Long Term Preservation. Amsterdam: NEDLIB Consortium 2000 (NEDLIB Report Series, 2), http://www.kb.nl/coop/nedlib/results/NEDLIBmetadata.pdf.

[87] Cedars Guide To Preservation Metadata (March 2002), http://www.leeds.ac.uk/cedars/guideto/metadata/guidetometadata.pdf.

[88] Preservation Metadata Framework Working Group, http://www.oclc.org/research/projects/pmwg/wg1.htm. Cf. OCLC/RLG Working Group on Preservation Metadata: Preservation Metadata and the OAIS Information Model. A Metadata Framework to Support the Preservation of Digital Objects, June 2002, http://www.oclc.org/research/pmwg/pm_framework.pdf.

schema. The OAIS defines a conceptual Information Model, yet it does not impose atomic elements or any other definitions at a detailed level. The OAIS Information Model distinguishes between Content Information with Reference Information, Preservation Description Information, Packaging Information, and Descriptive Information.[89] Thus, the general OAIS Information Model is a very inclusive description of metadata requirements within a digital repository for long-term digital preservation, with all its impacts on descriptive, administrative, and other metadata types. Its inclusiveness illustrates the breadth of preservation-related issues regarding metadata, and it provides a useful framework for establishing a dedicated preservation metadata schema.

The National Library of New Zealand (NLNZ) designed a preservation metadata schema "designed to strike a balance between the principles expressed in the OAIS Information Model and the practicalities of implementing a working set of preservation metadata".[90] The NLNZ model is most renowned for its modular structure and the way it allows to plug-in metadata components. This is particularly relevant when it comes to describing single file format types. As such it is possible to specify the resolution of an image or an audio file, yet not for a text; or to indicate the dimensions of an image and the duration of an audio file.[91] The NLNZ model has greatly influenced the preservation scene, and younger preservation metadata schemas were modelled after it, as for example LMER of Die Deutsche Bibliothek[92].

Following the work of the first OCLC/RLG Working Group on Preservation Metadata mentioned above, OCLC/RLG convened a new expert group called PREMIS[93]. In contrast to the first working group, PREMIS has slightly opened to organisations outside the library community. They follow two main issues, which are also covered by two separate sub-groups. The Core Elements sub-group aims to define generic preservation metadata elements which are applicable in any context[94]; and the Implementation Strategies sub-group focuses on the translation of recommendations to actual digital preservation systems including issues like the encoding, storage, and management of preservation metadata[95]. The PREMIS Data Model also addresses the issue of time in metadata, and how actions on collection items by various agents over time can be recorded in the schema.[96] The issue of

[89] A closer description of the OAIS model is provided in the previous chapter 2.2, and a an analysis of the OAIS Information Model from a reUSE perspective is given in chapter 3.5.2 below.
[90] Sam Searle, Dave Thomson: Preservation Metadata: Pragmatic First Steps at the National Library of New Zealand. In: D-Lib Magazine 9:4 (April 2003), http://www.dlib.org/dlib/april03/thompson/04thompson.html.
[91] National Library of New Zealand: Metadata Standards Framework – Preservation Metadata (revised, 2003), http://www.natlib.govt.nz/files/4initiatives_metaschema_revised.pdf.
[92] LMER – Langzeitarchivierungsmetadaten für elektronische Ressourcen [Long-Term Preservation Metadata for Electronic Resources], http://www.ddb.de/standards/lmer/.
[93] OCLC/RLG Working Group: PREMIS – PREservation Metadata: Implementation Strategies, http://www.oclc.org/research/projects/pmwg/.
[94] Cf. Rebecca Guenther, PREMIS – Preservation Metadata Implementation Strategies Update 2: Core Elements for Metadata to Support Digital Preservation, RLG DigiNews 8:6 (2004), http://www.rlg.org/en/page.php?Page_ID=20492#article2.
[95] Cf. Priscilla Caplan, PREMIS – Preservation Metadata Implementation Strategies Update 1: Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community, RLG DigiNews 8:5 (2004), http://www.rlg.org/en/page.php?Page_ID=20462#article2.
[96] After the finalisation of this White Paper the PREMIS group published its final report and its data dictionary for preservation metadata, see http://www.oclc.org/research/projects/pmwg/.

time was first explored in the ABC Ontology[97], and is particularly relevant for recording history trails of digital resources with the documentation of any preservation actions performed on the resource.

As discussed above, the generation of adequate and sufficient metadata may be one of the most costly tasks of digital repositories. Still metadata generation requires considerable manual input, although well-designed workflows, the reuse of metadata from other sources, and automation could greatly streamline efforts and reduce costs. The broad discussion of preservation metadata in the digital preservation community is certainly conducive to developing converging solutions that facilitate the exchange of technology, as well as the exchange of metadata through interoperable systems. One practical example for this is the initiative to build an international format registry[98], which essentially aims to concentrate efforts in creating Representation Information – documentation of data formats and the preservation of their technical specifications.

### 2.3.3 Relevance for reUSE

Metadata is a core building block of digital repositories, and certainly of great relevance for reUSE demonstrators. reUSE partners have recognised that an environmental scan of international activities with regard to metadata in digital repositories potentially saves them considerable effort while raising the quality of their metadata frameworks. This refers to the adoption of external metadata frameworks, but also the consideration of implementation issues debated in the international community. Tackling these issues from the outset rather than letting a metadata framework evolve "naturally" forestalls fundamental changes in the system architecture and workflow design, and reduces changes in the metadata schema when objects have already been accessioned into the archive. After all, these issues may illicit significantly more costs when performed on a running system than when considered early on. The participation of the reUSE partners in international co-operations and constant review of international developments ensure quality processes and state-of-the-art metadata systems.

REALITY CHECK

(cf. OAIS mapping, Preservation Description Information; chapter 3.5.2.2)

The Austrian National Library is moving towards an XML-based metadata architecture in their production system, which is capable of flexibly accommodating any metadata schema that has an XML representation. This approach promises flexibility for future extensions, though linkage and mappings between different metadata schemas remain to be resolved.

Apart from metadata developments in the digital repository community, metadata in the publishing sector are obviously of high importance to reUSE demonstrators.[99] The metadata systems of publishers may contain a wealth of data that could be extracted and reused, which could significantly reduce costs of metadata creation at the reUSE demonstrators.

[97] Carl Lagoze, Jane Hunter: The ABC Ontology and Model. In: Journal of Digital Information, 2:2 (November 2001), http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Lagoze/lagoze-final.pdf.
[98] Global Digital Format Registry (GDFR), http://hul.harvard.edu/gdfr/. Cf. Stephen L. Abrams, David Seaman: Towards a Global Digital Format Registry. 69th IFLA General Conference and Council (Berlin, August 1–9, 2003), http://www.ifla.org/IV/ifla69/papers/128e-Abrams_Seaman.pdf.
[99] Amy Brand, Frank Daly, Barbara Meyers: Metadata Demystified – A Guide for Publishers. The Sheridan Press / NISO Press 2003, http://www.niso.org/standards/resources/Metadata_Demystified.pdf.

Cooperation with data publishers must therefore go beyond the digital publications themselves and include metadata.

Metadata techniques are still relatively young and continue to evolve. Also, the embedding of publishers in the workflow at reUSE demonstrators has not been tackled on such a large scale before. The lack of long-term experience in this evolving field and the inherent instability of solutions, hence, make the creation of flexible metadata systems imperative, as these systems will change over time.

## 2.4 Interoperability, Digital Preservation, and Related Issues

Previous sections described international developments surrounding digital repositories and the guiding principles of building an OAIS that is trusted. However, this field is still very young and many challenges remain to be tackled. Various documents address current challenges and developments, and these issues are not reiterated here.[100] Rather, this White Paper highlights some selected issues that are of particular relevance to reUSE and to the following chapters of this paper.

Two key issues in digital repositories today certainly are interoperability and digital preservation. (It is, after all, no curious coincidence that the OAIS model addresses both issues in prominent position, its sections "preservation perspectives" and "archive interoperability", and that current research programmes in the field are often aligned along these two issues.) The two fields are related to a variety of other topics including cooperation, strategy planning and automation, and some of those issues relevant for reUSE are discussed in the following. While all of these topics are significant at all levels and form components of a digital repository, they are discussed here in the context of interoperability and preservation.

### 2.4.1 Interoperability

Interoperability between digital repositories can be achieved in various ways and on various levels. More than a technological challenge, it is an organisational and political one that builds on agreements and sound cooperation between two or more parties.[101] When focusing on the exchange of digital objects between repositories, the parties first of all need to provide data and/or metadata to their partners which they can understand. To achieve this, the parties need to agree on the structure, the syntax and the semantics of the data and/or metadata objects, and they need to either comply with common formats or to provide a means for metadata mapping and/or for object conversion. In addition to the content of the objects, the partners have to arrange the actual transfer and to agree on issues including selection criteria (which digital objects, or even which components of a digital object), the transfer channel (e.g. permanent online connection or regular transfer of offline media) as well as the interface and protocol (i.e. the transfer format and schedule).

Interoperability varies in how close different repositories are linked together. At its closest, a group of repositories may be federated to a single repository space. Current initiatives in this

[100] See for example: Paul Wheatley: Institutional Repositories in the context of Digital Preservation. DPC Technology Watch Report 04-02 (March 2004), http://www.dpconline.org/docs/DPCTWf4word.pdf.
[101] Cf. Paul Miller: Interoperability. What is it and Why should I want it? In: Ariadne 24 (June 2000), http://www.ariadne.ac.uk/issue24/interoperability/.

area include ARROW[102] in Australia, the JISC Information Environment[103] in the UK, DLF Aquifer[104], NDIIPP[105], and NSDL[106] in the United States.

Interoperability considerations within reUSE are rather at the other end of this continuum. However, there may be agreements for exchanging specific digital objects between partners, where this fits the profile of the respective repositories and where this is deemed beneficial. Thereby, the partners' commitment to the use of standards such as the OAI protocol or the METS metadata schema described earlier is most conducive to achieving interoperability.

Apart from cooperation between repositories, interoperability between a repository and its external stakeholders – including data producers and also users – is integral for ensuring a smooth transfer of objects between the parties involved. At best the respective data formats and workflows plug into each other, and fit the target's requirements with minimum effort at the source. Interoperability in this sense is discussed in more detail in the discussion of workflows below (cf. chapter 2.4.3).

To reiterate, interoperability is more than merely suitable technology interfaces to plug separate systems together. It is also an organisational issue in which the workflow for data exchange needs to be carefully designed, and the parties need to share data models for a common understanding of the underlying syntax and semantics. Encumbering simple and long-term stable interoperability mechanisms are the dynamicity of the digital repositories field due to its young age, and the range of stakeholders with their varied interests. To meet the intricacy of the issues involved, an inclusive forum to identify opportunities and trends is essential. Only cooperation between all stakeholders, peers, and interest groups with their varied objectives, backgrounds, and origins will enable effective interoperability in the long-term. And such an inclusive forum and open discussion are also – and perhaps even more than in interoperability – the pillars of the digital preservation community, for preservation related activities affect and interact with all areas of a digital repository.

[102] ARROW – Australian Research Repositories Online to the World, http://arrow.edu.au.
[103] JISC Information Environment, http://www.jisc.ac.uk/index.cfm?name=ie_home.
[104] DLF Aquifer: the Distributed, Open, Digital Library. Digital Library Federation (DLF), http://www.diglib.org/aquifer/.
[105] NDIIPP – National Digital Information Infrastructure and Preservation Program, http://www.digitalpreservation.gov/.
[106] NSDL – National Science Digital Library, http://www.nsdl.org/.

### 2.4.2 Digital Preservation

> Digital preservation consists of the processes aimed at ensuring the continued accessibility of digital materials. To do this involves finding ways to re-present what was originally presented to users by a combination of software and hardware tools acting on data. To achieve this requires digital objects to be understood and managed at four levels: as physical phenomena; as logical encodings; as conceptual objects that have meaning to humans; and as sets of essential elements that must be preserved in order to offer future users the essence of the object.
>
> UNESCO Guidelines[107], p. 34

This abstract of the "UNESCO Guidelines for the Preservation of Digital Heritage" epitomises the goals of digital preservation. It is by now a widely known phenomenon that the rapid development of our technological and organisational environment puts much of the data we create at immediate risk of being corrupted or even lost. Various high profile cases[108], where digital information was irretrievably lost as the media it was stored on or its software formats became obsolete, first pointed organisations involved in the stewardship of information resources to the new challenges of the digital age. In 1996 the report "Preserving Digital Information"[109] triggered much of the massive interest and manifold initiatives regarding digital preservation that we see today. The steep rise in preservation related activities notwithstanding, active counter measures against digital preservation challenges remain with some spearheading organisations and individuals, and stable solutions are still to be developed.

As the UNESCO guidelines point out, long-term accessibility of digital resources is hindered by both, hardware and software evolution and ultimately technological obsolescence. The rapid technology pace has given birth to a myriad of data carriers for storing digital information that all had a reasonable market saturation at their high times. Hardly two decades ago the 5¼-inch floppy discs were ubiquitous data carriers. Superseded by 3½-inch discs in the mid 1990s and later by various types of optical CDs, nobody uses 5¼ discs any longer. Ascending the spiral further, hardly anybody still uses 3½ discs nowadays.[110] So apart from the enormous fragility of magnetic and optical carriers, which leads to their physical deterioration and the loss of the data they hold in a fairly short time, it may be difficult to find suitable, still operational readers for those discs, or systems from which the data can be transferred to current systems. To tackle these hardware related problems, most

[107] UNESCO, Information Society Division: Guidelines for the Preservation of Digital Heritage. Prepared by the National Library of Australia (Colin Webb), March 2003, http://unesdoc.unesco.org/images/0013/001300/130071e.pdf.

[108] Model examples for the urgent necessity of digital preservation include the often cited NASA Viking Mars mission, which lost twenty percent of the data it collected in 1976, or the BBC Domesday Project from the late 1980s; in both cases huge amounts of money were invested in the creation of digital information resources, which were later conceded to the ongoing technological advancement. Cf. Jeffrey Darlington, Andy Finney, Adrian Pearce: Domesday Redux: The Rescue of the BBC Domesday Project Videodisks. In: Ariadne 36 (July 2003), http://www.ariadne.ac.uk/issue36/tna/. See also: CAMiLEON: BBC Domesday, http://www.si.umich.edu/CAMILEON/domesday/domesday.html.

[109] Preserving Digital Information. Report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and The Research Libraries Group, May 1, 1996, http://www.rlg.org/en/page.php?Page_ID=20442.

[110] Cf. Wikipedia: Floppy Disk, http://en.wikipedia.org/wiki/Floppy_disk.

institutions assumed an approach of refreshing information on their data carriers at least once a year, and transferring data to current media every five to ten years.

Perhaps even more pressing than the fragility of hardware and the obsolescence of both carriers and readers, however, is the rapid evolution of software. Digital information is tied to the software environment it was created in, or other software applications capable of reading and rendering those information objects. The mere retention of software applications is obviously insufficient, due to the inherent dependence of software on the operating system, which in turn is tied to a specific hardware environment. Focussing on textual documents, for example, new version releases of the Microsoft Word application come at least every two years. Of course, the borders between new versions and major bug-fixes are blurry, yet at least five versions of the text-processing application can be identified since the release of "Microsoft Word 2.0" in 1993.[111] Microsoft ceases to support a product after five years.[112] While new versions of the same software are usually backwards-compatible, each step bears the danger of small quirks being introduced in documents: altered layout, re-numbered footnotes, lost top-lines, a paragraph shifted to the next page, lost comments are just some potential document corruptions. As the difference between software versions increases or documents are moved between entirely different applications (for example WordStar to WordPerfect to Microsoft Word to OpenOffice), the risk of corruptions and essentially loss of information rises. These problems notwithstanding, the successive migration of documents from one format to a superseding format is the most widely applied preservation strategy, at least as far as static, textual documents are concerned. The wide range of available formats that may be hugely complex and proprietary types, and the huge and heterogeneous collections of many digital libraries aggravate the laborious tasks involved in a migration strategy for digital preservation.

REALITY CHECK

reUSE demonstrators expect that Adobe PDF will be the most prevalent **format** supplied by data providers from public sector organisations, perhaps with a share of way over two thirds. Others may include the proprietary formats Microsoft Word and QuarkXPress.

One mechanism to relieve the strain posed by successive format migration is the transfer of a digital resource to a suitable preservation format at its submission to the repository. For example, an author may write her text in the proprietary Microsoft Word format and submit it as is to the repository. The repository in turn considers Word unsuitable for long-term maintenance and converts the text to an Adobe PDF (Portable Document Format), which has a publicly available format specification and which promises immutable representation into the future. Such an early conversion may yield a format that is more stable at a time when the source format is not yet obsolete and the necessary conversion tools may be readily available. Alternatively, the repository may bring forth format recommendations and guidelines for data providers, and may reject non-conformant formats. Formats may be judged along several lines as to their suitability for preservation. Important criteria thereby include openness (whether the repository has access to the format specification), portability (dependence on external hardware or software), and quality (features including simplicity, robustness, and market saturation).[113]

[111] A complete account of the Microsoft Word versions since the release of Word for Windows in 1989 includes Microsoft Word 2 (1991), 6 (1993), 95 [=7] (1995), 97 [=8] (1997), 2000 [=9] (1999), XP/2002 [=10] (2002), and 2003 [=11] (2003).
[112] Microsoft Support Lifecycle, http://support.microsoft.com/lifecycle/.
[113] Lars R. Clausen: Handling File Formats, May 2004, http://www.netarchive.dk/website/publications/FileFormats-2004.pdf.

Obviously, formats cannot be transferred into any other formats. A video can hardly be transferred into a PDF, or a text illustrated with colourful images should not be transferred into a text-only file encoded in ASCII. Thus, while the criteria listed before are useful to assess preservation suitability of a particular format, the repository also needs to assess the significant properties of its resources, i.e. those aspects of a resource it intends to preserve.[114] Important aspects thereby include readability, comprehensibility, appearance, functionality, and look & feel. This task is purely up to the repository. For example, one repository with regard to its particular mandate may consider the colourful illustrations of a text document to be no significant property, and may consequently preserve only the ASCII text, which is more stable than a PDF with embedded images. Another repository may consider the text to be editable and other MS Word specific functionality to be essential significant properties, which would be corrupted by a transfer to PDF.[115]

For the preservation of textual documents a transfer to a format like PDF may turn out to be sufficient for the requirements of many organisations from a significant properties viewpoint. When analysing the preservation suitability, however, many preservation initiatives consider PDF as unsuitable for long-term digital preservation due to its complexity. Specifically, PDF may contain dynamic content, embed proprietary formats, depend on external resources, and include other features which are inapt for preservation. For this reason the PDF/A format[116] is currently being specified, a subset of PDF that disallows its complex features and is meant to be suitable for preservation purposes.

Apart from PDF/A and best suitable for the preservation of textual documents from the current perspective would be an XML-based format. However, due to the difficulties in ensuring an adequate transfer to XML, this approach is in many situations perceived as infeasible.[117]

REALITY CHECK

(cf. OAIS Preservation Planning, chapter 3.5.3.5)

All reUSE demonstrators are in the process of evaluating PDF/A. Humboldt University Library and the National Libraries of Austria have recommendations with categories for formats that are supported or unsuitable for preservation. The Austrian National Library is working on best practice recommendations for the adequate creation of formats, with a first set of **guidelines** available for PDF. Humboldt imposes strict format restrictions to enable the conversion to an XML-based preservation format.

[114] Cf. Carl Rauch, Andreas Rauber: Preserving Digital Media: Towards a Preservation Solution Evaluation Metric, http://www.ifs.tuwien.ac.at/~rauch/Rau04_Utility_Analysis.pdf.
[115] Andreas Aschenbrenner: File Format Features and Significant Properties. Presentation at the Chinese-European Workshop on Digital Preservation (Beijing, China, July 14–16, 2004), http://www.csdl.ac.cn/meeting/cedp/schedule.html.
[116] PDF/A – Portable Document Format / Archive. Submitted to ISO for accreditation as an international standard. A joint activity by the Association for Suppliers of Printing, Publishing and Converting Technologies (NPES), and the Association for Information and Image Management (AIIM International), http://www.aiim.org/standards.asp?ID=25013. ISO Standard under development: ISO/DIS 19005-1 – Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1), http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=38920&scopelist=PROGRAMME. Cf. the information on PDF/A at the Digital Formats for Library of Congress Collections web site, http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml.
[117] Cf. Andreas Aschenbrenner: The Bits and Bites of Data Formats – Stainless Design for Digital Endurance. In: RLG DigiNews 8:1 (Feb. 2004), http://www.rlg.org/en/page.php?Page_ID=13201#article2.

As one component of a migration strategy, many preservation initiatives are currently systematically documenting data formats, the software they are created with, and collecting their format specifications where available. The Library of Congress is currently building up a website which provides information about digital content formats.[118] The UK National Archives created the first operational format registry, PRONOM[119], and the international preservation community is working towards the creation of an international data format registry[120], for this is a widely shared concern. However, an operating format registry that offers a variety of services including format identification, verification, and conversion may still be a number of years away, due to technical but also, and even more than that, due to organisational issues.

While the search of remedies against technological obsolescence remains at the core of digital preservation, the changing **organisational** environment in which digital resources are being produced is recognised as a considerable threat to the resources' long-term stability as well. Organisational change that triggers modifications to the technological environment is only one aspect of this. Away from challenges, however, the activities related to facilitating digital longevity are more than merely technological processes[121], and they embrace among others procedural, social and cultural, economic, and legal questions[122].

A long-lasting approach to digital preservation needs to span multiple generations of systems and technologies, as it needs to bridge organisational change and new staff. Thus, digital longevity can only be achieved on a strategic level with programmes that address responsibility, viability, sustainability, technical suitability, security, and accountability.[123] Furthermore, digital preservation cannot be ascribed to an isolated organisational entity. Actions at all lifecycle stages of a digital object may impact on its 'preservability', and therefore it is vital to incorporate digital preservation in the organisation's workflows and to account for it already from the outset and at all stages of the digital repository lifecycle. A preservation **policy** framework is therefore an essential tool in guiding the organisation towards a consistent approach regarding digital preservation, and it assigns the responsibility and provides the necessary authorization.[124]

To sum up briefly here at this point, digital preservation permeates all aspects of the repository, from ingest to access and from strategy down to technology and procedures, and a digital preservation approach can only be successful if it involves a comprehensive strategy supported by people and organisation as well as technology. The following closes up on financial issues and the people dimension, two points that further underline the breadth and strategic nature of preservation programmes.

As preservation is touching on so many different areas, communication across departmental borders and the exchange of experience across sectors is essential. Institutions increasingly demand structured **training** to mitigate the lack of long-term experience and skills in the field.

[118] Digital Formats for Library of Congress Collections, http://www.digitalpreservation.gov/formats/index.shtml.
[119] UK National Archives: PRONOM, http://www.records.pro.gov.uk/pronom/.
[120] Global Digital Format Registry (GDFR), http://hul.harvard.edu/gdfr/.
[121] Brian Lavoie, Lorcan Dempsey: Thirteen Ways of Looking at ... Digital Preservation. In: D-Lib Magazine, 10:7/8 (July/August 2004), http://www.dlib.org/dlib/july04/lavoie/07lavoie.html.
[122] See also chapter 3.3 for a discussion of legal aspects with regard to the cooperation between data providers and the digital repository.
[123] Trusted Digital Repositories (Fn. 39).
[124] ERPANET Guidance: Digital Preservation Policy (September 2003), http://www.erpanet.org/guidance/docs/ERPANETPolicyTool.pdf.

Organisations and projects including ERPANET[125] and DELOS[126] offer relevant workshops on a range of issues including information management and repository management, and they thereby address this requirement as far as possible at this young age of the field. Apart from staff training, repository interfaces are proving hard for users to operate and respective assistance needs to be provided. Most importantly, information producers need guidance in accessioning their resources into the repository.

SURVEY DATA

(O5.5/6) The reUSE demonstrator at the Humboldt University Berlin offers **training** for data producers.

(O5.2) reUSE partner institutions follow various paths for staff training.

Insufficient training of internal or external stakeholders needs to be compensated at other stages of the workflow and may impact on the quality of the digital objects or cost considerably. For example, if data producers are not capable or not willing to provide data in a requested format and/or to supply adequate metadata, the repository may need to follow up on this and either go back to the data producer or prepare the data and metadata itself.

**Costs** of a digital repository are hard to calculate due to the lack of hands-on data from other initiatives to compare with. Only homogeneous collections that have some level of influence over data producers and a clearly delimited designated community exhibit stable costs related to digital preservation over the years of their existence. Some data is transferable from other processes, however the lack of experience with digital preservation costs obstructs a complete picture. In general, digital preservation costs are assumed to be significant, and even greater in the digital environment than for paper.[127]

REALITY CHECK

**Cost** calculations are essential for digital repository management.

The Austrian National Library carried out a five year cost prognosis as part of their digital preservation strategy.

Humboldt University Library builds on their experience and has found costs to be relatively stable since the creation of their digital repository edoc in 1997.

While clear numbers are missing, some factors impacting on digital preservation costs can be identified. The costs of a digital repository and particularly those of digital preservation may be considerable in the long term, since the digital resources have to be continually perpetuated from one technology generation to the next. Exacerbating this is the impossibility to predict further technology evolution. It is, however, necessary to incorporate a predictive component into a comprehensive cost calculation, considering questions such as the development of hardware costs and the future demand for storage space. Consequently, cost calculation for digital preservation has to be predictive, yet credible.

[125] ERPANET – Electronic Resource Preservation and Access Network, http://www.erpanet.org/.
[126] DELOS – Network of Excellence on Digital Libraries, http://www.delos.info/.
[127] Maggie Jones, Neil Beagrie: Preservation Management of Digital Materials. A Handbook (2002), Chapter 2.1: Strategic Overview. How much does it cost?, http://www.dpconline.org/graphics/digpres/stratoverview.html#how2.

What else becomes clear when viewing the cost factors are the tight interrelations between the stages of the information lifecycle. Digital material demands a pro-active approach that starts at the creation of the objects and demands perpetual attention over the whole retention period. More effort put in the quality and the documentation of the objects at an early stage lowers costs for corrective measures at a later stage disproportionately. Therefore digital preservation affects all actors involved in the digital object lifecycle. Particularly data providers including commercial publishers are therefore concerned about the potential costs arising from e.g. supplying metadata that go beyond what is ordinarily supplied, and the adequate transfer of those to the digital archive.[128]

Economies of scale and cooperation with other archives could lower costs significantly. The creation of software tools to support and automate necessary tasks is a dominant expenditure that can be shared between archives. Even digital objects and the responsibility for their preservation can be shared to raise efficiency, increase safety, and to lower costs.

REALITY CHECK

With regard to **costs** of digital archive functions, reUSE partners are convinced that the administrative stages prior to ingest (including negotiation of the submission agreement) and the ingest stage consume significant financial and staff resources. Thereby, contacting and establishing cooperation with potential depositors demands considerable expenditures for all reUSE partners, and ongoing expenditures rise proportional to quality requirements for the digital objects and their metadata.

While a variety of initiatives[129] embarked on formulating cost models for digital preservation, they all remain on a rather abstract level. Only slowly and with rising experience some hands-on numbers start to emerge.[130] Cost models are key management tools and the open exchange of such numbers may be defining for some initiatives. With rising experience it may also become possible not only to calculate costs but also to juxtapose them to the actual benefits of the repository and to thereby impress a more positive slant on cost/benefit calculation.

### 2.4.3 Towards Trusted Digital Repositories

Many of the issues discussed in the previous sections can and should not be pigeonholed. Indeed, an action taken for the sake of interoperability may actually be conducive to preservation and may simultaneously affect other features of the digital repository at the same time. This is why a holistic strategy is integral to give the repository a general direction and to ensure consistent development. All the contained issues including cooperation, policy, training, and costs discussed above of course have a bearing on the whole organisation. Above they were discussed from the perspective of interoperability and preservation respectively. Below a number of issues are approached from a general standpoint spanning

[128] Cf. The Impact of the Extension of Legal Deposit to Non-Print Publications. Assessment of Cost and Other Quantifiable Impacts. Study report prepared for the Joint Committee on Voluntary Deposit by Electronic Publishing Services Ltd (1 October 2002), http://www.alpsp.org/2004pdfs/LegalDepositofNon-PrintPublications.pdf.

[129] Cf. and also refer to the bibliography in: ERPANET Guidance: Costing Orientation (September 2003), http://www.erpanet.org/guidance/docs/ERPANETCostingTool.pdf.

[130] Anne R. Kenney: Digital Preservation Management: Identifying and Securing the Requisite Resources. Presentation at the ERPANET Seminar on Business Models related to Digital Preservation (Amsterdam, September 20–22, 2004), http://www.erpanet.org/events/2004/amsterdam/presentations/erpaTraining-Amsterdam_Kenney.pdf.

the whole organisation. Again, the addressed issues are not an exhaustive listing. Rather those issues of particular importance to reUSE and to the following chapters of this White Paper have been selected for discussion.

The issue of **cooperation** has been discussed in the context of interoperability as well as preservation, and beyond that cooperation permeates all aspects of a digital repository. Networks such as the European project BRICKS[131] are essential for fostering technological advancement and convergence with increased interoperability. On a strategic level, openness and cooperation generates a multiplicity of benefits to all stakeholders involved, including a greater pool of skills and experience, more stable business models, more robust preservation approaches, potentially synergies in workflow management, and many others. With this in mind it is obvious that the use of open standards, accounting for interoperability, and engaging in other international collaborations is also key for the reUSE network and its individual partners.[132]

**Business models** and advocacy are among the issues where digital repositories could benefit greatly from cooperation and exchange of experience. A repository combines a variety of stakeholders with different interests. Often institutions are struggling to publicise the benefits of depositing resources in digital repositories. To draw an analogy with EPrints software, in early 2005 EPrints repositories had an average of just above 200 collection items.[133] It has to be noted that there are only a couple of huge repositories vis-à-vis a majority of small repositories with only a few dozens of collection items. This may be since a considerable number of EPrints instalments are pilot projects or have just been started up. It may also be the case, however, that only a small number have been truly accepted by the respective community.

SURVEY DATA

(T2.10/11) reUSE partners are cautious in estimating the **size of their collections** in five years time. The Humboldt University expects to boost its collection from 2.000 to 5.000 items. ALO with its direct information sources and national libraries hope for thousands of collection items by then.

In any case, the average number of collection items in EPrints repositories is far below the number of items targeted by reUSE partners. Often a critical mass of content decides over success or failure. Therefore, reUSE partners need to develop a suitable business model to maximise interest into their service from both, information producers and users. Successful strategies for institutional repositories include integration of the repository into the business environment and daily work, and pro-active promotion of the repository and the benefits of participation. reUSE partners have clear advantages over many other repository initiatives: they already established a working relationship with data suppliers in the print domain, and the reUSE host institutions are trusted, which should rub off on the new reUSE service. Furthermore the reUSE partners will put effort in pro-active collection and promotion of the reuse service. Valuable is also the tight cooperation among reUSE partners for discussion of

[131] BRICKS – Building Resources for Integrated Cultural Knowledge Services. Project of the European Commission in the IST Sixth Framework Programme, http://www.brickscommunity.org/.
[132] Apart from cooperation with peers, good working agreements with data providers are essential for creating comprehensive collections of good quality. In the absence of sufficient legal deposit regulations for digital material any cooperation between data providers and the reUSE demonstrators is voluntary. For a discussion see the reUSE analysis chapter 3.2.1 on external stakeholders / data providers, and the treatise on submission agreements in chapter 3.3.
[133] On January 6th 2005 statistics at www.eprints.org announced 31.688 total records in 148 known archives running EPrints software worldwide.

strategies and exchange of experiences including for example cost figures and strategy implications.

Another area where repositories can share experiences and which awaits urgent development is workflow management. Each functional entity of a repository has its tasks and responsibilities, and performs a number of (routine) processes. **Workflows** describe and formalise these processes. A sound workflow design has numerous benefits, including its potential to maximise quality and control, and to facilitate transparency and communication. Eventually the employment of a workflow management system allows to efficiently control the coordination between activities, applications, and process participants.[134] Workflows are essential components of trusted digital repositories, and additionally a smooth workflow has the potential of reducing costs and time input.

However, workflow design has not yet gained ground in this field, due to the lack of experience in digital repository design[135] and the initial creation costs. After all, proficiency in workflow design is essential for achieving modular and flexible processes that can change over time.[136] Various initiatives are currently creating abstract functional models related to repository management, which could be the basis for workflow design. These initiatives include the emerging OAIS "Producer-Archive Interface Methodology Abstract Standard"[137] that focuses on the ingest of resources into the repository, and the preservation function models by InterPARES[138]. Eventually, however, workflow design is dependant on the internal organisational structure and the specific relationships with external stakeholders, and it hence needs to be tailored to the individual organisation.

REALITY CHECK

(cf. reUSE workflows, chapter 3.4)

reUSE demonstrators are designing their **workflows** carefully and in an iterative process. All found the OAIS Producer-Archive Interface to be a useful reference. Innsbruck University Library undertook to translate and tailor the document to their specific needs, an effort which sparked interest among the other partners and was discussed extensively. reUSE demonstrators continue to refine their workflows and push towards automation wherever possible.

Explicit workflow descriptions and careful workflow design is the basis for streamlining processes and for repository **automation**. Many tasks in a digital repository have to be implemented for each and every collection item, which entails an enormous workload. The

[134] Michael zur Muehlen: Workflow-based Process Controlling. In: Foundation, Design and Application of workflow-driven Process Information Systems. Berlin: Logos 2004.
[135] Note that experience in digitisation is hardly transferable to entirely digital processes. For relevant experience with workflows in digitisation see for example: Assessing the Costs of Conversion. Making of America IV. A Handbook created for the Andrew W. Mellon Foundation. The University of Michigan Digital Library Services, July 2001, http://www.umdl.umich.edu/pubs/moa4_costs.pdf.
[136] Stephan Heuscher: Workflows in Digital Preservation. Presentation at the ERPANET Workshop on Workflows (Budapest, October 13–15, 2004), http://www.erpanet.org/events/2004/budapest/presentations/Workflows_in_Digital_Preservation_2004-10-13.pdf.
[137] Consultative Committee for Space Data Systems (CCSDS): Producer-Archive Interface Methodology Abstract Standard. CCSDS 651.0-B-1, Blue Book, May 2004, http://www.ccsds.org/CCSDS/documents/651x0b1.pdf.
[138] InterPARES Preservation Task Force: A Model of the Preservation Function, June 2002 (=The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project, Appendix 5), http://www.interpares.org/book/index.cfm.

DSEP process model[139] developed in the NEDLIB project distinguishes eleven workflows within the digital archive of a deposit library, from selection and acquisition to preservation and access. Workflow management systems that, for example, allow a largely automatic ingest procedure at the archive are key for reducing costs and maximising the quality of repository collections. Therefore, automation is aspired where possible, and tools such as JHove[140] for file format validation and automatic metadata extraction have a high impact in the preservation landscape. The National Library of New Zealand estimates that about 80–90 percent of metadata could be created automatically.[141] However, reliable automation requires a pragmatic approach with adequate workflow design and quality control. Any consistent repository workflow geared at automation will be a combination of automation and machine-assisted, semi-automatic procedure. Maximising automation and the integration of all processes with various degrees of automation into a tightly-knit repository workflow are vital requirements for the further evolution of digital repositories. Without these repository initiatives may never be able to achieve operational viability, doomed to remain rather short-term projects and unsustainable for the long-term.

Workflow design and automation is not only essential within the repository, but also at the interfaces to external stakeholders. Organisational as well as technological challenges encumber a smooth workflow, which minimises the data producer's effort and ensures a complete data transfer into the repository. One possible approach to this is introduced by the ADAPT project's Producer-Archive Workflow Network (PAWN).[142] PAWN surpasses the mere transfer of data – e.g. via FTP or offline transfer of CDs – by enabling the producer to create complete information packages containing digital objects as well as metadata. It thereby plugs directly into the repository's ingest workflow. However, PAWN demands the close cooperation of the data producer and assumes various characteristics of the producer's systems, which limit PAWN's applicability. Nevertheless, PAWN and various other initiatives[143] working on data ingest are most valuable for their contribution to workflow and repository design. A smooth workflow between the information producer and the repository is one of the key elements of the reUSE project, and it is discussed in the next chapter.

[139] Titia van de Werf: The Deposit System for Electronic Publications. A Process Model. Amsterdam 2000 (= NEDLIB Report 6), http://www.kb.nl/coop/nedlib/results/DSEPprocessmodel.pdf.
[140] JHOVE – JSTOR/Harvard Object Validation Environment, http://hul.harvard.edu/jhove/.
[141] ERPANET Chat on Metadata for Digital Preservation (November 2003). In succession to the ERPANET Seminar: Metadata in Digital Preservation. Marburg, Germany (September 3–5, 2003), http://www.erpanet.org/events/2003/marburg/.
[142] ADAPT – An Approach to Digital Archiving and Preservation Technology (University of Maryland Institute for Advance Computer Studies, http://www.umiacs.umd.edu/research/adapt/. Cf. Mike Smorul, Joseph JaJa, Yang Wang, and Fritz McCall: PAWN: Producer-Archive Workflow Network in Support of Digital Preservation. Technical Report, 2004 (CS-TR-4607, UMIACS-TR-2004-49), http://www.umiacs.umd.edu/research/adapt/papers/UMIACS-TR-2004-49.pdf.
[143] See also, for example, PRESERV – PReservation Eprint SERVices, http://www.jisc.ac.uk/index.cfm?name=project_preserv.

Overall, the creation of trusted digital repositories evokes multiple interdependencies between all life-cycle stages of a digital object and the various stakeholders involved, and intricately links together strategy with technology and people. The last sections underlined the necessity for cooperation, a strategic framework, as well as suitable technology and procedure. The demanding objective of building a trusted digital repository may perhaps only truly be accomplished, if efforts are embedded in an adequate national policy environment and relevant collaborations.

REALITY CHECK

(O5.3) All reUSE partners engage in a variety of collaborations as part of their institutional digital preservation development plans. Humboldt University Library and Die Deutsche Bibliothek have positive experiences with their participation in the national digital preservation forum NESTOR[144], which was modelled after the UK Digital Preservation Coalition[145]. Austria is working towards a similar national network, which was announced in the resolution of a notable event at the Austrian National Library with massive participation from all types and sizes of Austrian

[144] Kompetenznetzwerk Langzeitarchivierung NESTOR – Network of Expertise in Long-Term Storage of Digital Resources, http://www.langzeitarchivierung.de/.
[145] Digital Preservation Coalition, United Kingdom, http://www.dpconline.org/
[146] Long-term preservation in the digital age. The UNESCO Charta on the Preservation of Digital Heritage, and an Austrian national strategy. An event of the Austrian UNESCO Commission and the Austrian National Library, Vienna, 9 March 2005, http://www.onb.ac.at/about/lza/veranstaltungen/unesco/.

# 3 reUSE demonstrators

After the previous chapter provided a scan of ongoing initiatives in the international repository arena and especially the reUSE-related issues and trends, this chapter presents the reUSE demonstrators. The section is informed by a variety of surveys and analyses among the four reUSE demonstrators, including user scenarios and use case analyses, OAIS mappings, overview surveys, and interviews. While the documents produced in the course of these analyses remain reUSE internal, this chapter highlights the key approaches and distinctive features of reUSE demonstrators.

reUSE demonstrator organisations feature a variety of organisational settings. Just to highlight some salient organisational features before the introduction of the demonstrators in the next section: the ALO demonstrator is implemented by a consortium of two Austrian university libraries and one university institute; the National Library of Estonia is installing an open source repository, whereas the Austrian National Library acquired a commercial software product; and the system at Humboldt University is the only repository with a thematic collection. Despite this range of approaches, reUSE demonstrator organisations follow the common goal to collect, preserve, and make available publications that were formerly collected in print form only, as outlined in the introduction to this paper.

Each of the four organisations undertook careful analysis before constructing the demonstrators, and analysis and review is ongoing. Two key steps in this analysis are the identification of stakeholders and their needs, and the subsequent specification of functionalities and requirements before system deployment. This chapter synthesises these two steps at the reUSE demonstrators by first identifying the relevant stakeholders with a special focus on data producers and agreements between them and the repository, and then presenting an overview of workflow and system functionalities analysed along OAIS concepts.

## 3.1 reUSE Demonstrator Profiles

### 3.1.1 ALO – Austrian Literature Online

The ALO digital repository has been set up in March 2002. Originally designed for digitised documents it also contains electronic documents in different formats (XML, PDF, RTF,..). Currently more than 6.500 books, journals, and manuscripts are online. ALO is maintained by a working group of Austrian libraries: The university libraries from Innsbruck (UBI) and Graz (UBG) and the university department i3s3 which is responsible for the technical development. The ALO system is strongly based on standards: All (meta-)data are available in an XML file which is assembled according to METS (Metadata Encoding and Transmission Standard). Descriptive data are gathered according to the Dublin Core standard and the MAB2[147] standard. The workflow within ALO is designed in a way that the documents are made available on a special website of the digital repository as well as via the local and national electronic library catalogues (ALEPH) in Austria. A direct link is included in the MAB2 record which allows the user to access an electronic document of the digital repository. ALO is designed as an open-source package available for free. The technical

[147] MAB2 – Maschinelles Austauschformat für Bibliotheken [Automated Library Exchange Format], http://www.ddb.de/professionell/mab_e.htm.

basis is a mySQL database and JAVA servlets carrying out several features. Via the SOAP[148] standard from the W3 Consortium, information interchange between different systems is facilitated. SOAP is also especially important to support web-services for added value services, such as print on demand. The ALO system is completed with a client programme used for decentralised creation of METS objects which are than uploaded to the central server via the SOAP interface. This interface is written in Borland Delphi and runs under Microsoft systems. The University Library Graz is partner of the ALO consortium and one of the main users.

### 3.1.2 National Library of Estonia

The National Library of Estonia (NLE) is setting up a trusted digital repository to collect, preserve and make available publications from public sector publishers. Digital material from NGOs and private publishers is accepted if they follow the archive's submission guidelines and open access policy. The creation, maintenance and ongoing maintenance of the repository are the collaborative effort of various NLE departments, including Technology Services for technical issues, Collection Development Department for collecting and cataloguing digital publications, and the Marketing Manager for external relations to information producers. The whole initiative is supervised by the Director of Technology Services and approved by the Director General. These organisational arrangements help ensure the sustainability of the archive and its services.

From a technical perspective, the repository system is being built upon the Fedora Digital Repository Management System[149]. The repository is accessible through an online interface. Also, relevant records will be linked to the OPAC and other bibliographical databases.

In implementing this repository the National Library of Estonia builds on extensive experiences from a previous pilot-project called ARES.[150] During the years 1999 to 2001 digital publications of four scientific publishers were collected providing valuable insights into the publisher's attitudes and workflows in electronic publishing.

### 3.1.3 Austrian National Library

The Austrian National Library (ONB) is establishing a digital repository in line with its legal mandate and responsibility as a national library. Targeted data suppliers are public sector institutions and commercial publishers. Digital material from private persons is accepted to a limited extent. The ONB aspects thousands of objects during next years.

The preservation of digital master files is part of a range of activities at the ONB concerning the long term preservation of offline and online media.

The digital preservation department is part of the department for collection development and processing, where three staff members focus on digital preservation on a strategic and operational level. Additional staff members of the department for collection development and processing are involved in the cataloguing process. Technical infrastructure and support is

[148] Simple Object Access Protocol, http://www.w3.org/TR/soap/.
[149] The Mellon FEDORA (Flexible Extensible Digital Object and Repository Architecture) Project, University of Virginia and Cornell University, http://www.fedora.info/.
[150] ARES - Artiklite Elektrooniline Süsteem / Electronic System of Articles. ARES links in online catalogue http://helios.nlib.ee.

part of the institutional infrastructure supervised by the department for information technology services.

ONB's digital repository – a commercial software – hosts diverse digital archives and connects to other activities like digitisation projects of the ONB as well as to the library system. A completely new and enhanced version of the software will be implemented during summer 2005 and will go online by the end of 2005.

### 3.1.4 Humboldt University Berlin repository edoc

edoc[151] is a thematic institutional repository at Humboldt University Berlin (UBER), which incorporates scientific publications of the Humboldt University and of cooperating partners. The support of the repository is carried out by the joint Electronic Publishing Group of the Computer- and Media Services and the University Library, with permanent and project staff in order to ensure sustainability of the service as well as continuous further development.

With this organisational regulation edoc ensures a comprehensive collection and a high quality service. There are already some 2.000 items in the repository, catalogued and prepared for long-term preservation. Publications, stored within the edoc server use XML as preservation document format. Digital signatures and time stamps ensure the integrity of the materials over time, and open access is granted via an online gateway. Moreover UBER has already set up a professional print-on-demand service for distributing bound copies of the electronic documents.

The edoc server has been set up in September 1997 with a first focus on digital dissertations. It aims to become the digital publication server (and trusted digital repository) where scientific publication of Humboldt-University and associated and cooperating institutions are published in future. There are tendencies to integrate edoc as part of the digital University Press, that is about to be established by the University as a cooperation between the University's Research department, the Vice President for Research, the Computer- and Media Services and the University Library. A deposit of documents is voluntary for authors, doctoral candidates can choose an electronic publication at the edoc-server in order to fulfil their publication duty for receiving the doctor's degree.

The edoc repository will be extended towards publications from the public sector with a thematic restriction to reports, yearbooks or studies from non-profit organisations in the area of science, culture and history. It will also function as hosting server for other universities, universities of applied sciences and institutions. First cooperation has been made with DINI[152], NESTOR[153] and the HTWK Leipzig[154] to host publications.

## 3.2 Stakeholders

The identification of the relevant stakeholders is a key action, and it needs to be performed as one of the first steps in a requirements analysis process, following software engineering

[151] Document and Publication Server of Humboldt University Berlin, http://edoc.hu-berlin.de.
[152] DINI – Deutsche Initiative für Netzwerkinformation [German Initiative for Networked Information], http://www.dini.de/.
[153] Kompetenznetzwerk Langzeitarchivierung NESTOR – Network of Expertise in Long-Term Storage of Digital Resources, http://www.langzeitarchivierung.de/.
[154] HTWK – Hochschule für Technik, Wirtschaft und Kultur Leipzig (FH), http://www.htwk-leipzig.de/.

techniques such as the Use Case methodology.[155] Relevant stakeholders are primary actors and users, but also actors that are not necessarily directly involved with the system. For example library management has a stake in the adequacy of the reUSE service with regard to the mandate and accountabilities of the library, yet they do not interact with the reUSE system as part of their daily tasks.

In reality different roles may, of course, overlap or a single person may combine various responsibilities. For example, a library manager may be author at the same time and her publications may be ingested in a reUSE service. Therefore, the following identification of relevant stakeholders displays prototypical roles and their vested interests and goals with the reUSE system. This synopsis of the stakeholder identifications done by each of the reUSE partners starts with data providers and users, both external to the archive organisation, and it then highlights internal roles at the organisation.

### 3.2.1 Data Providers

All reUSE partners focus on public digital master files, which are acquired through active collection or voluntary deposit. In their selection of digital objects to be ingested into the reUSE service they focus on openly accessible publications by organisations including public sector institutions, non-profit organisations, and higher education. The National Libraries of Austria and Estonia both start from their background as national legal deposit libraries for traditional publications and build on the existing relations with data providers from this context. However, national deposit law in both Austria and Estonia fails to adequately address digital master-files and generally the deposit of online digital publications, and thus the reUSE services are dependent on the voluntary cooperation of the data providers. The situation is similar at the University Library Innsbruck, which is a deposit library for the Austrian region of Tyrol.

SURVEY DATA

(O1.1/3; cf. chapter 3.3) National legislation in the countries of reUSE partners are largely inadequate regarding today's digital challenges. reUSE partners consider adequate **national policies** essential for fulfilling their mandates, and they actively contribute to shaping their legislative environments.

In Austria the National Library together with the Federal Chancellery are working towards an amendment of legal deposit regulations. The National Library of Estonia urged an amendment to legal deposit regulations to be passed, which is due in 2005. With the filed changes the NLE will be obliged to collect online digital publications for free access and preservation.

The Humboldt University Library Berlin operates a thematic repository that is focused on the university. All scientific documents published by the members of Humboldt University are eligible for inclusion in the edoc repository, and submission of publications is actively promoted though not obligatory.[156] This more focused group of data providers enables a more rigid quality control regime that includes deposit guidelines such as adherence to document templates and specific file formats, as well as detailed annotation of the resource with metadata.

[155] Cf. Alistair Cockburn: Writing Effective Use Cases. Boston, London: Addison-Wesley 2001 (=Crystal Series for Software Devolopment).
[156] See the Policy document of the edoc server, http://edoc.hu-berlin.de/e_info_en/policy.php.

While the influence on data providers differs slightly between the reUSE demonstrator organisations, all have a working relationship with the depositors in the traditional print domain. This relationship is an enormous asset when it comes to promoting the reUSE service. After all, the reUSE partners still face a considerable challenge in raising awareness about digital preservation in the public, and in encouraging data producers to contribute to the digital repository and to assume extra effort that may be involved. Besides these promotional activities, reUSE partners are designing their systems and procedures such that the workflow for depositing resources demands minimal effort from the data producer. Added-value services to be designed, such as a print-on-demand service, represent another benefit for the data producer that is expected to encourage participation.

As an add-on to the reUSE service the Austrian partners consider opening up their repositories to publishers with a commercial stake in their products. This, of course, demands the protection of copyrights and other issues with regard to access rights. The necessary rights management schemes need to be adopted in discussions with the respective publishers. Such discussions are ongoing, and generally all reUSE partners are investigating the interests and needs of potential data providers in their area of influence. Template submission agreements have been designed by each reUSE partner and compared amongst each other.[157]

REUSE – CLOSE UP

**National Library of Austria – Data Producers**
**The Federal Law Gazette at the Austrian Federal Chancellery**

The Austrian National Library and the Austrian Federal Chancellery entered a special cooperation encompassing the transfer, preservation and access provision of the Federal Law Gazette. The official publication is thereby transferred to the National Library in the original XML source format, which is particularly conducive to long-term preservation. The Humboldt University Library has long-lasting experience in the management of XML-based formats, and the two reUSE partners mutually benefit from each others experiences. The Austrian National Library is convinced of the benefits of this approach regarding workflow efficiency, (meta)data quality, and suitability for an adequate preservation approach, and plans to initiate similar collaborations wherever possible.

[157] See the following sub-chapter for more on submission agreements.

REUSE – CLOSE UP

**University Library Innsbruck – Data Producers**
RiS-Kommunal

At the time of writing the University Library Innsbruck has already contacted a large number of potential data providers clustered to focus groups such as schools and adult education, political parties, church organisations, or press. In this ongoing effort they probed a variety of strategies for winning data providers over. A promising approach is their contact with the web hosting service RiS (www.ris.at) and specifically their department RiS-Kommunal (www.riskommunal.at). RiS-Kommunal is the web hosting service of choice of about 900 municipalities in Austria and Southern Tyrol, which use it to present an array of issues regarding political and social life in the very community. Among the material featured on such a municipality website are also official publications such as the municipality news, announcements, or election results. Many of the publications will be available only in electronic form in the near future, and they are in imminent danger of being lost.
In their communication with RiS, the University Library Innsbruck aims to establish a direct linkage between the RiS content management system and the ALO digital repository, such that municipalities can transfer their publications for preservation by the mere click of a button and by using the software environment they are accustomed to. This simple feature enables the promotion of the ALO repository at a great number of data providers, and minimises transfer efforts on both sides, the data provider as well as ALO, by plugging into current software environments and practices for maximum automation. While this initiative is still in an early stage, the University Library is confident about its success and that this approach involving an intermediary between the data provider and the repository could be a valuable model for the future.

### 3.2.2 Users

reUSE has a strong commitment to improving the accessibility of public sector information and to satisfy user needs, and therefore all reUSE demonstrators offer open access to the general public.[158] The profile of their dedicated communities overlaps with those of their host institutions and is mainly defined through a geographic scope, i.e. the NLE obviously focuses on the Estonian society and the ONB analogously focuses on Austria. While the ALO consortium is based in universities, the University Library of Graz and Innsbruck also have mandates as local legal deposit libraries and they hence go beyond university borders. Their group of designated users includes the depositors themselves as well as other public sector bodies, libraries, or individuals like scholars and students with an interest in the resources. edoc on the other hand is dedicated to the academic community in the vicinity of Humboldt University Berlin, although it largely offers free public access to the general public

SURVEY DATA

(T3.7) All reUSE demonstrators allow open access and downloading of their resources. ALO and edoc additionally provide print-on-demand and CD delivery as added-value services. These **services** will be further extended in the scope of the reUSE project.

[158] Note that reUSE partners may have other digital repositories aside their reUSE specific collections that may have access restrictions, due to copyright limitations or other issues. These various services could – or even should be tightly integrated. Also the add-on for private sector publications mentioned in the previous section may entail access restrictions.

as well. Therefore it is the needs and expectations of university members who are considered primarily in the design of edoc services.

Whoever the users at the specific organisation are and whatever the access policies and the business model of the organisation, it is essential according to the OAIS reference model that the designated community is clearly identified. This is so important since the resources have to be preserved such that they are understandable for the designated community without needing assistance of dedicated experts or even the experts who produced the information. Moreover, the design of access facilities and other conceivable services are based on the characterisation of the designated community.

All users, of course, have similar basic needs, which include a one-stop gateway for accessing a rich collection free of costs, searching across institutional borders, and an easy-to-use interface. Among other things this demands the integration of the reUSE specific collection with other digital services at the respective library, and preferably also interoperability with repositories at other libraries. Beyond this, the reUSE project also explores potential added-value services. Knowing the designated community is essential for both, basic and added-value services. Indeed, some added-value services may be targeted at sub-groups of the huge general user community that is hard to pin down and analyse.

reUSE – CLOSE UP

**National Library of Estonia – External Stakeholders**
**A Cross-Section of Data Producers and the Designated Community**

DATA PRODUCERS

1. Publishers of public sector institutions and state agencies:
   - State Chancellery of the Republic of Estonia (and its subordinate agencies)
     Ministries (11)
   - County Governments (15)
   - Town Governments
   - Administrations and Boards (24)
   - Inspections (10)
   - Other Governmental Agencies
   - State agencies
   - Foundations
   - State Hold companies
   - Public Institutions (libraries, museums, archives)
   - Universities and research institutions
2. Non-profit organisations and associations
3. Private institutions acting in the field of research, education, culture
4. Private publishers / emphasis on journal publishers (Non-profit organisations, private institutions and private publishers are seen as content providers only if they permit free access to the materials archived in reuse archive.)

DESIGNATED COMMUNITY

1. Estonian society using NLE as
   - a national library (e.g. scientists, researchers, students, teachers)
   - a parliamentary library (providing information services to the Parliament, to the Government, to the governmental institutions and to the Office of the President)
   - a research library for the Humanities and Social Sciences providing information for research activities and offering a wide range of information services
2. ELNET Konsortsium (Consortium of Estonian Libraries Network, http://www.elnet.ee/)

NEEDS OF USERS FROM PUBLIC SECTOR UNITS

- Need for electronic document storage and long-term preservation service free of charge.
- Public sector publishers need feedback coming from the distribution of their production. reUSE will increase the society's attention to their production.
- A large range of users in universities and research institutions (like scientists, researchers, teachers, lecturers, students and specialists) expects additional possibilities for storing their works and expanded opportunities for disseminating their materials.
- Need for print-on-demand service on textbooks, schoolbooks and other materials published by educational institutions.

NEEDS OF NON-PROFIT ORGANISATIONS AND ASSOCIATIONS

- Need a wider dissemination of their materials. The exemplarity of their materials is usually limited because of the lack of publishing finances in the organisation.

NEEDS OF PRIVATE ORGANISATIONS, including private publishers (mainly journal publishers), and private institutions acting in the field of research, education, culture

- Are sometimes editing their books and journals with public funding. Keeping such electronic master files available in electronic format doesn't harm their commercial interests but increases society's attention to their materials and activity.

Further needs:
(1) Every single citizen needs the information published by public sector units.
(2) Network of ELNET Konsortsium (Consortium network of research and scientific libraries of Estonia) needs better access to public sector materials, because their distribution is very often limited and there are not always paper-copies available in all Konsortsium libraries.
(1+2) Users of National Library of Estonia need much better and quicker access to the materials with limited exemplarity.

### 3.2.3 Internal Roles

As an analysis of organisational charts has shown, reUSE demonstrator staff are well integrated in their host organisation. However, this was achieved by inherently different approaches to embedding the repositories in the overall organisation. The NLE is the only one with a stepwise organisational implementation plan, in which it accounts for a transient project phase of demonstrator development and testing. Thereby, the project is split between two teams. One team focuses on the design and implementation of the repository system and consists of four full-time equivalent (FTE) positions with an equal share of technical and library staff. The second team consisting of 2 FTE's is responsible for contacting potential data providers. Both teams are managed by a single project manager and supervised by a council of half a dozen personnel from the senior management of the National Library. This senior management support will be perpetuated once the repository moves from its initial phase into an ongoing service. Staff will then be lowered and split between the collection department and the department for information systems.

At the ONB, staff working on digital preservation on a strategic level and those staff contacting potential data providers are a team, which is a sub-unit of the department for collection development and processing. Technical infrastructure and support is part of the institutional infrastructure and the department for information technology services. In that their organisation is quite similar to the one of NLE in both structure and number of staff, though the ONB has no initial project phase with an increased number of staff. However, ONB's digital preservation strategy ties together their internal efforts to implement a digital repository with international collaborations and their research aspirations. This is exemplified by the position of the ONB coordinator for digital library research and development who is also responsible for internal project management of reUSE and a member of the ONB preservation strategy team.

The ALO repository has a similar split between technology on one side – whereby the technological infrastructure is provided by the University of Innsbruck –, and strategic and operational issues on the other. The latter team is largely an add-on to the library organisation. UBI consists of library entities for the distinct university faculties, and the main university library which is the umbrella for these decentralised entities. ALO is part of the main library, has about half a dozen FTE's, and combines digitisation and digital preservation. This is also in contrast to ONB, where the digital preservation unit is part of the department for collection development and processing whereas the digitisation unit is part of the department for reference and information services.

UBER assumes an entirely different organisational structure. It installed a virtual edoc team that is composed of staff from various departments. Additionally, the technological infrastructure is provided by the university and respective staff is not part of the core edoc team. The core team includes staff from the software and multimedia departments of the Humboldt University Computer and Media Services, and service and operation departments of the University Library. Combined they contribute a workforce of about 6 FTE's. While the team members are assigned to these various departments, the edoc team is an acknowledged entity within the organisation and the team leader directly reports to the heads of both, the Computer and Media Services and the University Library. Therefore, the team leader is equipped with the necessary authority and has well established communication channels to the various departments and to senior management.

Notably, all reUSE demonstrators have for their operational phase about the same number of overall FTE's, namely half a dozen. Staff training is a core requirement in digital preservation

related tasks, for the lack of trained personnel on the market, and to keep up-to-date in the quickly changing digital environment. reUSE partners follow a variety of paths for staff development including attendance at international workshops and support in internal and private training.

In the following the prototypal roles at digital repositories are outlined as they are found at each of the reUSE demonstrator organisations. Only UBER with its focus on quality and control, has a specialised group of staff responsible for guiding publication projects, which is not listed in the subsequent roles.

**Digital Preservation Strategy** – This role largely combines the tasks and functions of two OAIS entities, namely *Preservation Planning* and *Administration*. Staff are responsible for the development of digital preservation strategies suitable for the respective organisational context, design system functionalities and workflows, and monitor international developments regarding digital libraries and digital preservation. They also – which can be mapped to the OAIS entity *Administration* – design the business model, develop relations with data producers, negotiate submission agreements, manage the overall system configuration, and other related tasks. With regard to ongoing demonstrator development, digital preservation strategy staff probably combine the most delicate tasks. Their core interests regarding the demonstrator systems include flexibility for system adaptation to changing requirements and compliance to standards. It is their responsibility to synthesize system requirements to fulfil the functionalities and satisfy the interests of other stakeholders.

**System Administration** – Staff responsible for information technology expects a repository system that can be integrated smoothly in the existing technology infrastructure, is easy to configure and adapt, and is well documented or comes with the necessary external support. IT experts contribute to ongoing technology watch for preservation purposes, and their evaluation of software and tools is important to strategic digital preservation staff for system development and refinement.

**Cataloguers** will be the pivotal players in ensuring the quality of the resources once the repository is operational. They clearly have an interest in well-designed workflows, maximum automation, and easy-to-use cataloguing interfaces.

**Library Management** – While not directly involved in system development and operation, the library management has to ensure that the digital repository contributes to fulfilling the library mandate, and they need to audit and control its adequacy and accountability. Moreover, library management needs regular briefings on the state of the repository for staff planning, financing, and possible organisational support.

## 3.3 Voluntary Deposit Submission Agreements[159]

### 3.3.1 Introduction

When publications are stored, preserved, and made accessible in a digital repository, the archiving institution performs legally relevant activities, concerning especially the copyright of the publication. In order to perform those activities lawfully, the archiving institution has to

[159] Contribution by Christian Recht, Austrian National Library. Please be aware that the following statements do not constitute legal advice in a specific situation.

acquire the consent of the rights holder to these actions in the form of a submission agreement.

In a narrow sense, the term "submission agreement" refers to the depositor's declaration or the mutual contract between the archiving institution and the rights holder transferring those rights to the archiving institution. In a broader sense, the term "submission agreement" applies to all mutual understandings and conditions governing the submission of a publication, whether fixed in a formal license or stated in "terms and conditions" or other information issued by the archiving institution. In this understanding, a formal declaration or contract which only transfers the relevant rights from the rights holder to the archiving institution is but a small part of the overall submission agreement. It is in this broader sense that the term "submission agreement" is used here.

REALITY CHECK

(cf. chapter 3.5.1, OAIS mandatory responsibilities) reUSE demonstrators have designed template submission agreements. These templates differ in several core issues, and their analysis and cross-comparison provides valuable insights. Already an overview analysis by the Austrian National Library has sparked lively discussion among the partners.

### 3.3.2 Framework Submission Agreements

National and international initiatives for voluntary deposit of electronic publications have been mostly driven by the national libraries, the responsible institutions for legal deposit. Common to most framework agreements is their motivation to create a voluntary scheme of deposit for electronic materials ahead of eventual legislation, which is perceived as the ultimate goal.[160]

Most prominent examples for framework agreements include the CENL/FEP statement[161] (2001) at the international level, and the UK Code of Practice[162] (2000), the KB/DPA[163] (1999) and the DDB/BV[164] (2002) at the national level.

### 3.3.3 reUSE relevance of framework agreements

Although the reUSE project is targeted specifically at the digital master files of conventional print publications and not at electronic publications, there is no doubt that the same principles apply as for agreements concerning digital publications because the copyright of the rights owner is – at least in this matter – irrespective of the technical means used for publication. More generally speaking, in order to enjoy copyright protection, a work has to be fixed in some form, but the protection is granted to the immaterial individual and artistic achievement, not the specific material used for giving this achievement the required form. Therefore the

[160] With the exception of the agreements between the Koninklijke Bibliotheek and the Dutch Publisher's Association, considering that Legal Deposit does not exist in the Netherlands.
[161] Conference of European National Librarians / Federation of European Publishers: Statement on the development and establishment of codes of practice for the voluntary deposit of electronic publications, http://www.bl.uk/gabriel/fep.
[162] Code of practice for the voluntary deposit of non-print publications (United Kingdom publications), http://www.bl.uk/about/policies/codeprac.html.
[163] Koninklijke Bibliotheek / Dutch Publishers Association: Arrangement for depositing electronic publications at the deposit of Netherlands publications in the Koninklijke Bibliotheek, http://www.kb.nl/dnp/overeenkomst-nuv-kb-en.pdf.
[164] Die Deutsche Bibliothek / Börsenverein des Deutschen Buchhandels e.V.: Rahmenvereinbarung zur freiwilligen Ablieferung von Netzpublikationen zum Zwecke der Verzeichnung und Archivierung, http://deposit.ddb.de/netzpub/web_rahmenvereinbarung.htm.

digital master files also have to be considered “published” even if the output material used for distribution is different to the material it was fixed on first or during the production process. Therefore the existing framework agreements for electronic publications are in principle also applicable to the reUSE cooperation with publishers.

It has to be said though that most of those framework agreements are in fact much more political statements than binding contracts between two parties. They are unspecific, incomplete[165], and out-dated: still, they provide a good reference point for individual submission agreements as they represent a standard agreed between the concerned interest groups – the libraries on the one hand and the publishers on the other.

This is especially important in relation to the access and the restriction of access to the documents, where individual submission agreement negotiations will benefit from the minimum access defined in the previous documents (e.g. the single concurrent user access at the premises of the institution according to the CENL/FEP statement).

In other respects, the cited framework agreements will hardly be of assistance. Any added value services provided by reUSE project members have no precedent in these agreements, nor do these agreements contain a clear archiving policy. The definition of the objects to be submitted as well as the submission procedures need to be more focused. Basic questions as to which legal procedures to apply need to be defined individually, especially concerning the licensing process (single object or cumulative or a combination of both), the licensing form (mutual agreement or unilateral form or both), and licensing documentation (analogue or digital).

### 3.3.4 Main legal issues for individual submission agreements

When entering an agreement with third party contributors for the submission of digital files, the archiving institution is best advised to draft the agreement to fit best its own archiving and access policy, the target material, the contributors, and the procedures adopted for the submission of the material. In this respect, main legal issues will include definitions of the material and the deposit procedures, provisions concerning the access to the material, the copyright exemptions and liabilities, preservation issues, and miscellaneous points of a more general nature.

**Definitions:** On a general level, sustained document formats, document types, other technical specifications and submission procedures including metadata transmission should be defined whereas the submitted data and the transaction should be documented on an individual level. A clear partner identification is necessary on both levels and on both sides, e.g. the general commitment of an institution to provide/archive master files has to be signed by the legal representatives of both institutions (definitely not the project leaders of some branch or department), whereas the individual submitted document should be traceable to an identified member of the institution entrusted with the transaction of submission/archiving.

[165] Cf. the DDB/BV agreement, which does not contain any provisions concerning the circumvention of technological measures applied to submitted materials. A separate agreement between the DDB, the BV and the Bundesverband der phonographischen Wirtschaft [Federation of the phonographic industry], recently concluded (January 2005), entitles the library to circumvention of technological measures for materials deposited under the legal deposit obligation.

**Access:** If the master files are stored on a server accessible to members of the submitting institution, special access rights for employees of the archiving institution could be considered necessary.

Crucial is the right to grant access to the submitted material to third party users. Different levels of access (special – registered user – anonymous user/on site use – remote use/single access – multiple access) will apply. The possibility to enact restrictions on the basis of the individual document use terms could be considered.

Archiving duties might also include statistics, access monitoring and the obligation to control access. Although the right to circumvent technological measures to enable access will depend on the definitions of the data to be submitted, it should be included at least as a safety measure.

**Copyright:** Of utmost importance is the provenance of the materials and the transfer of rights from the entitled rights holder. The importance of the full rights ownership of the depositor should be stressed by an indemnity clause which holds the depositor liable for any breach of copyright and damages of the archiving institution incurred through the performing of the concluded arrangement according to its clauses.

It should be stated clearly that archiving activities typically involve copying and transferring submitted materials to other formats, data carriers and environments, thereby changing certain aspects and possibly modifying the content as well.

Other copyright relevant uses made by third party users could be subjected to end user licenses, possibly integrating existing models such as Creative Commons[166] licenses, allowing for an individual setting of use terms on a minimal basis (such as author acknowledgment). At any rate, use made according to the fair dealing clause should not be inhibited.

**Preservation:** As outlined earlier, the existing framework agreements are provisional emergency measures to compensate for the so far (in most countries) inexistent legal deposit of electronic online materials. As for legal deposit, a specific archiving commitment is therefore clearly not considered an issue in those statements, as the archiving commitment for materials archived under legal deposit is determined by the institutions' foundation law, which defines the institutions basic duties and responsibilities.

The situation is different however for the voluntary submission of electronic files, which are not within the scope of a legally defined archiving commitment. In individual agreements, which are bound to be more specific than mere political statements, a more substantial commitment might be necessary, which will be defined according to the archiving institutions' policies. It should be stressed however that such a commitment entails responsibilities which definitely outlast archiving on a project basis (if not stated otherwise).

Issues to be taken care of also include the levels of reliability and authenticity for material preservation, which should be rated in relation to data and system requirements; and the archiving institution should be permitted to exclude materials from archiving or access if implicated by technical or legal constraints.

**Miscellaneous:** Standard contract clauses such as applicable law, place of jurisdiction and a saving clause will be of lesser importance than duration and termination clauses and

[166] Creative Commons is a global non profit organization founded in 2001 and based at the Standford Law School. It's mission is to "build a layer of reasonable, flexible copyright", adopted to a variety of national legislations, http://creativecommons.org/.

establishing a mechanism for separate software licensing if needed to make use of the material. According to the procedures for submissions used and the type of submitted materials, consent to the use and transmission of personal data will be required as well.

These contractual issues listed above do not take into account any potential added-value services provided by the archiving institution. In order to include those services in a submission agreement, they would have to be defined first. At this stage of the reUSE project it seems too early[167] to include this possibility in a submission agreement.

### 3.3.5 Examples for individual submission agreements

Two comprehensive examples of the different ways and possibilities to deal with submission agreements are the DSpace License[168] on the one hand, and the AHDS[169] General Licence[170] on the other hand. While the DSpace License is an unilateral, online click-through license form, the AHDS General licence is a mutually signed, paper-form license. While the DSpace License focuses on a minimum of rights granted by the licensor (the right to copy, to reproduce, to change the format, to distribute) the AHDS General licence is exhaustive in its listing of the rights concerned. The same applies to the warranties made by the depositor: a simple claim of rights ownership and of having identified any third party material by the depositor on the one hand (DSpace), a full plethora of warranties and non-warranties by the depositor and the archiving Data Service on the other hand (AHDS). In short, while the DSpace license form focuses on a strict minimum of legal statements needed from the depositor, the AHDS licence is driven by the effort to include all rights and liabilities between the contracting parties in a single document.

### 3.3.6 Submission procedures

Whichever approach is taken, the individual submission agreement is influenced by the overall procedure adopted for the deposit of the material. Generally speaking, at some point before the actual deposit the conditions for deposit have to be agreed on, preferably in a way that is on the one hand minimizing potential risks of misuse, on the other hand minimizing efforts on both parts, embedding this process in a smooth workflow.

While a click-through online license might be a viable option for standardized contributions from individual depositors with a small number of contributions, the deposit of e.g. a big amount of publications in a batch process will necessitate differentiated procedures and therefore individualized agreements. While in the first scenario the individual supply of metadata will be the standard, in the second scenario a more pragmatic approach will be necessary.

---

[167] An agreement on added-value-service which includes commercial exploitation, besides raising questions concerning the repartitions of income, taxes, the legal status of the archiving institution and sublicensing options, might also necessitate a business plan to come to terms with a potential contracting partner.

[168] http://www.dspace.org/implement/policy-issues.html#licenses.

[169] Arts and Humanities Data Service (http://www.ahds.ac.uk/) is, according to its own description, “a UK national service aiding the discovery, creation and preservation of digital resources in and for research, teaching and learning in the arts and humanities”. It is funded by the Arts and Humanities Research Board (AHRB, established 1998), and the Joint Information Systems Committee (JISC).

[170] http://www.ahds.ac.uk/documents/ahds-general-licence-form.doc.

In both scenarios, the general agreement on deposit on the one hand (whether in the form of a mutual, all-encompassing contract or e.g. a rights holder declaration combined with a global reference to terms and conditions of deposit available at the website of the archiving institution) should be separated from the individual filing / registering of the publications concerned.[171]

The point is, however, that submission agreements have to take into account the target contributors and the target material: to some extent, they have to be “made to measure”, otherwise acceptance and relevance will be low. As submission agreements have also to fit the archiving and access policy of the individual archiving institution, they should be drafted on an individual basis.

## 3.4 Workflows

As mentioned previously,[172] careful workflow design and a drive towards automation are core aspects of the operational responsibility of a digital repository, and fundamental building blocks for all attributes of a trusted repository. reUSE demonstrator institutions put a lot of effort in their workflow design. Limited staff resources compel them to optimise workflows for maximum efficiency, and at the same time they aim to enhance their control over the tasks of internal and external actors for the sake of quality assurance.

At the time of writing, demonstrators continue to be refined and various policy and technology issues remain to be resolved. The still variable workflow design of the four demonstrators encumbers comparison at this point of time. On a high level, all demonstrators are in line with the eleven deposit library workflows as identified by the DSEP process model.[173] These eleven workflows are:

1. selection for collection building
2. acquisition
3. delivery/capture/harvest
4. registration
5. verification
6. description
7. storage-handling
8. preservation
9. delivery
10. access
11. monitoring

All these workflows are being designed and implemented by each of the reUSE demonstrators separately, tailored to their respective organisational environment, as well as policy and system qualifications as far as they exist. Most of reUSE partners' attention is currently focused on the phases of – sticking to OAIS terminology – pre-ingest and ingest,

[171] A good example for a one-time-application for deposit and an in-time-registration of publications is the deposit procedure for online publications of the German Nation Library (DDB), cf. at http://deposit.ddb.de/netzpub/np_stepbystep.htm [in German]. Compared to the DDB procedure, the submission procedure adopted by the AHDS [cf. http://www.ahds.ac.uk/depositing/how-to-deposit.htm] seems to be slightly more complicated and troublesome.
[172] Cf. chapter 2.4.
[173] Titia van de Werf: The Deposit System for Electronic Publications. A Process Model. Amsterdam 2000 (= NEDLIB Report 6), http://www.kb.nl/coop/nedlib/results/DSEPprocessmodel.pdf.

which include the workflows 2 to 6 of the DSEP process model. Contacting and negotiation with data producers, and the resource transfer and registration in the digital repository depend on a number of guidelines, agreements, system interfaces, and manual tasks. In addition to their own experiences, reUSE partners review external references such as the OAIS "Producer-Archive Interface Methodology Abstract Standard"[174] document, which focuses on exactly these workflow components. In fact UBI tailored the OAIS Producer-Archive Interface to their own needs and created respective internal checklists.

While reUSE demonstrators put a lot of thought in their workflows, the actual design process is rather informal and implicit at this point of time. For the purposes of this paper the partners described the state of their workflow development in the form of use cases.[175] They thereby determined an overall workflow sequence in the form of a high level use case[176], which branches into several atomic tasks[177]. Three examples of these tasks in the form of use case diagrams are highlighted in the following.[178]

| | | |
|---|---|---|
| INSTITUTION | National Library of Estonia, NLE | |
| USE CASE NUMBER AND NAME | 2. PR coordinator manages agreements | |
| USE CASE SUMMARY | PR coordinator manages publisher profiles and agreements, and keeps track of transactions. Failure would be the inability to save information to the necessary details. | |
| SCOPE | ingest process | |
| PRIMARY ACTOR AND GOAL | PRIMARY ACTOR | GOAL |
| | PR coordinator | keep track of depositor activities |
| STAKEHOLDERS AND INTERESTS | STAKEHOLDER | INTEREST |
| | depositors | Make sure that content provided by depositors is handled according to agreements. |
| PRE-CONDITIONS | New information about a publisher or a new agreement | |
| POST-CONDITIONS | All details about publishers, agreements and transactions are saved. | |
| TRIGGER | PR coordinator discovers a new publisher. | |
| DESCRIPTION OF MAIN SCENARIO | STEP | ACTION |
| | 1 | Inserting information about publisher |

[174] Producer-Archive Interface Methodology Abstract Standard (Fn. 137).
[175] Cf. Cockburn, Effective Use Cases (Fn. 155).
[176] Following Alistair Cockburn's granularity illustration: "way up in the clouds".
[177] Following Alistair Cockburn's granularity illustration: a user-goal at sea level.
[178] The user scenarios and use cases of all reUSE demonstrators have been shared an discussed among the partners. Three exemplary use cases are presented here, whereas the complete documents remain reUSE internal.

| | | |
|---|---|---|
| | 2 | Inserting contract |
| | 3 | Inserting terms of depositing individual publications |
| | 4 | Inserting transactions (times contacted, negotiations) |
| DESCRIPTION OF EXTENSIONS | STEP | BRANCHING ACTION |
| | 1a | New information or update on existing information |
| | 1a1 | Creating a new record |
| | 1a2 | Updating an existing record |

This use case by the National Library of Estonia describes a supportive software tool for a largely manual process. With this tool the NLE aims to improve communication between the data producer and the archive and to enhance transparency and control over the negotiation process. The described task is part of the first DSEP workflow "selection for collection building", and it is featured in the OAIS functional model as the function "Negotiate Submission Agreement", which is part of the Administration[179] entity.

NLE's approach to this task is special in that it is supported by dedicated software functionality. Already during their ARES pilot project[180] the NLE found such a supportive tool essential for managing the multiplicity of potential data producers. Following the NLE example, the ALO consortium has installed a commercial Customer Relationship Management software to facilitate communication with data providers.

| | | |
|---|---|---|
| INSTITUTION | Humboldt University Berlin, UBER | |
| USE CASE NUMBER AND NAME | 4**.** Checking staff performs Check Offline Resource | |
| USE CASE SUMMARY | Checking staff checks the digital resources. The control contains the file formats, the style sheets or templates, and the file set (abstracts (de, en)) and keywords (de, en). | |
| PRIMARY ACTOR AND GOAL | PRIMARY ACTOR | GOAL |
| | Checking staff | Check the files for formats and usage of style sheets or templates and file set |
| SUPPORTING ACTORS AND GOALS | SUPPORTING ACTOR | GOAL |
| | Converting staff | Communicate about style sheets and templates |
| PRE-CONDITIONS | Pre-ingest activities are concluded. Temporary directory for copy of the offline resource has been created in ingest storage. | |

[179] Cf. the respective OAIS functional entity in the upcoming chapter 3.5.3.4.
[180] ARES – Artiklite Elektrooniline Süsteem/Electronic System of Articles. See also the reUSE demonstrator profile of the Estonian National Library in 3.1.2.

| | | |
|---|---|---|
| | Librarian has access to offline carrier. Actual version of template or style sheets is installed on ingest workstation. Librarian has access to file or set of files that have to be checked. | |
| POST-CONDITIONS | File and file set are checked for template or style sheets, virus, completeness, readability. The files are valid. | |
| POST-CONDITIONS FAILURE | Actions taken are documented, Administrator-Staff is informed. | |
| TRIGGER | Successfully uploaded document(s) via the online form. | |
| DESCRIPTION OF MAIN SCENARIO | STEP | ACTION |
| | 1 | Checking staff control the file set (abstract and keyword) |
| | 2 | Checking staff control the style sheets and templates |
| | 3 | Virus scan software scan the files automatically |
| | 4 | The files are valid and the author gets the publication acknowledgement |

The UBER workflow for checking submitted publications comprehends a number of manual and automatic subtasks. Checks include automatic virus scan, conformance to document templates[181], and quality control. The submitted publications may be rejected if they fail any of the checks. All reUSE demonstrators have similar audit procedures, though of course the guidelines, templates and software tools vary. Moreover, UBER's quality check, which also includes the structure of the publication content, is the most comprehensive among reUSE partners. ONB's quality criteria, while remaining on a file format level, include guidelines and recommendations for format types and features.

The UBER use case above matches DSEP process 5 "verification", and it stretches over multiple OAIS functions and entities including "Ingest – Quality Assurance" (cf. OAIS chapter 3.5.3.1), "Administration – Audit Submission" (cf. OAIS chapter 3.5.3.4), and it is influenced by "Preservation Planning – Develop Preservation Strategies and Standards" (cf. OAIS chapter 3.5.3.5).

| | |
|---|---|
| INSTITUTION | Austrian National Library, ONB |
| USE CASE NUMBER AND NAME | 3.1 Import bibliographic metadata from Library System to Digital Repository and export newly created URN from Digital Repository to Library System |
| USE CASE SUMMARY | Cataloguer searches bibliographic record in Library System, transfers parts of it to Digital Repository, and returns the newly created URN from Digital Repository to Library System. |

[181] Humboldt University Institutional Repository: Author Guidelines (in German), http://edoc.hu-berlin.de/e_autoren/.

| | | |
|---|---|---|
| PRIMARY ACTOR AND GOAL | PRIMARY ACTOR | GOAL |
| | Cataloguer | Export bibliographic metadata record from Library System to Digital Repository and newly created URN for it from Digital Repository to Library System. |

| | | |
|---|---|---|
| DESCRIPTION OF MAIN SCENARIO | STEP | ACTION |
| | 1 | Cataloguer searches in Library System for appropriate record. |
| | 2 | Record is found: Cataloguer pushes button. |
| | 3 *** | Digital Repository: MAB2[182]-Fields are converted into Dublin Core or the appropriate metadata fields in Repository. Additional MAB2 fields, that cannot be mapped, remain just in Library System and are not stored within the Digital Repository database. |
| | 4 | Digital Repository opens with metadata editing tool, showing the imported record. |
| | 5 | Cataloguer checks and refines metadata. |
| | 6 | Cataloguer creates URN for record. |
| | 7 | Cataloguer pushes button. |
| | 8 *** | URN is put into appropriate MAB field in Union Catalogue. […] |
| DESCRIPTION OF EXTENSIONS | STEP | BRANCHING ACTION |
| | 1a | Record is not found. |
| | 1a1 | Cataloguer skips action and begins with ingest in Digital Repository. |

The metadata exchange between the Digital Repository and the Library System described in this use case is specific to the Austrian National Library. Since the ONB collects both, print publications by public sector institutions as part of its responsibilities as a deposit library as well as the respective digital master files in the scope of the reUSE project, the descriptive metadata for a specific publication is eventually registered in two separate systems, the Library System and the Digital Repository. Other than manually entering the metadata twice, the metadata is exchanged between the two systems. Since the underlying metadata schemas of the respective systems differ, this process is overseen by a human cataloguer. While at this point of time all steps are manual, step 3 and 8 could partly or fully be automated in the future. Other reUSE demonstrators including the NLE and ALO face similar

[182] MAB2 – Maschinelles Austauschformat für Bibliotheken [Automated Library Exchange Format], http://www.ddb.de/professionell/mab_e.htm.

situations and may adopt a similar workflow or may entirely integrate the two systems with a single database for descriptive metadata. This situation is particular to the very organisational context at legal deposit libraries and can hardly be mapped to the OAIS model directly, though it does not contradict the OAIS model. The DSEP on the other hand is geared towards deposit libraries and accounts for this task in its 6th process "description".

These samples show that a single task may span multiple system functionalities. They underline that careful workflow design leverages quality and efficiency, and in that the reUSE demonstrators may benefit from the perspectives and experience of their partners. However, this exercise also demonstrated the difficulty of workflow comparison. Compliance to a framework such as the DSEP process model and a specification of the description granularity can mitigate this difficulty only to some degree. In the end the inherent dependence of workflow design on the organisational and technological environment encumbers workflow comparison between different institutions, even if they share common starting points and goals as part of the reUSE project.

The preservation community agrees that a promising approach for repository comparison builds on the OAIS reference model. The terms and concepts of the OAIS are considered suitable for comparison, albeit there is as yet no precedence on comparing different archives which claim OAIS compliance.[183] The following section therefore provides a rather informal comparison of the reUSE demonstrators based on the OAIS reference model.

## 3.5 OAIS Mapping

All demonstrators stem from a similar organisational background and pursue similar goals in establishing their reUSE services. This chapter is an effort to contrast the demonstrators and to take stock of what has been put in place already. Development at the reUSE demonstrators is ongoing in a rapid process, including technical adjustments and organisational issues such as negotiations with data producers. As reUSE builds upon already existing systems, these are growing incrementally, rather than following a strict engineering plan with the successive and clearly contained stages of design, implementation, and testing. The incremental development process allows maximum flexibility to accommodate the project constraints, the local requirements at the site of the demonstrator, and the diverse stakeholder interests.

Since system design, policies, responsibilities and workflows are evolving throughout the reUSE project this comparison is but a snapshot of the status quo at the demonstrators in early 2005. Each demonstrator created a comprehensive OAIS mapping of its system. This White Paper already embedded a variety of information extracts from these mappings in the previous chapters, and this section presents a synthesis of the OAIS mappings and mixes description of technical features and human tasks. (The comprehensive OAIS mappings remain reUSE internal.)

[183] Indeed the preservation community still lacks a clear method to attest OAIS compliance, though efforts towards this are ongoing, for example, as part of initiatives on trusted digital repository certification (cf. Chapter 2.1.2, "trusted repositories").

### 3.5.1 OAIS mandatory responsibilities

The OAIS standard defines a minimal set of responsibilities for an archive to be called an OAIS (Open Archival Information System) and be distinguished from other uses of the term 'archive'.[184] It also acknowledges that there are various mechanisms for discharging these minimal responsibilities. Thus, although the standard illustrates the minimal responsibilities with possible approaches and examples, the chosen mechanisms may vary between different OAIS's.

*Negotiate for and accept appropriate information from Information Producers.*

Contacting potential information producers and subsequently negotiating submission procedures are key tasks performed by all demonstrators in the scope of the reUSE project. Thereby, selection criteria are largely given through the nature of the reUSE project and thus, as described above, demonstrators focus on freely accessible publications from public sector institutions or generally in the public domain. Demonstrators only vary slightly in how rigidly these selection criteria are defined. ALO on the one hand has very loose selection criteria, whereas the NLE demands that publications undergo some sort of publishing process where the publications are approved by an editor in some way. UBER, of course, focuses on a specific group of data producers and their documents (e.g. dissertations) have mostly been supervised and controlled by professors or other scholars.

The demonstrators also vary with regard to their demands of metadata provision at the submission of resources. Whereas ALO creates all metadata themselves, UBER has a slim, yet mandatory set of descriptive metadata that the data producer has to provide. ONB negotiates the transfer of metadata on a case-to-case basis, depending on the data producer and what metadata is available in the originating systems, the material to be transferred, transmission channels, and so forth.

*Obtain sufficient control of the information provided to the level needed to ensure Long-Term Preservation.*

Ensuring the necessary rights for performing unrestricted preservation actions on the digital resources is, of course, most relevant in the context of reUSE since all data providers are external to the reUSE organisations. The importance for a passage addressing these issues in the respective submission agreements will even increase with the inclusion of private publishers or other publications that are not in the public domain. For an in depth discussion of Submission Agreements please refer to the respective chapter above (chapter 3.3).

reUSE demonstrators are all working on generic submission agreements, which are then adapted to the requirements of the individual data provider and the specific situation. One interesting feature of the Innsbruck University Library submission agreement is their separation of technological and workflow issues from the actual agreement, since those are rather ephemeral aspects of a cooperation, which is meant to persist in the long-term.

Interestingly enough, the demonstrators differ in some points drastically in their initial approach. For example the NLE takes responsibility to make all efforts to preserve digital information over the long-term that is available in an adequate format. UBER on the other hand denies responsibility beyond a time-span of more than 50 years, and only if the

[184] Consultative Committee for Space Data Systems: CCSDS 650.0-B-1: Reference Model for an Open Archival Information System (OAIS). Blue Book, Issue 1. January 2002, http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf, p 3-1–3-5.

supplied information is available in an XML format and in-line with their respective guidelines. ONB only commits to a best effort attempt to adequate preservation, for the hitherto lacking ultimate preservation solution.

The partners are now in the process of sharing and discussing their draft submission agreements, which may lead to a convergence towards somewhat more consistent approaches. In any case this process will provide valuable insights for all reUSE partners.

> *Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and, therefore, should be able to understand the information provided.*

Since reUSE demonstrators are hosted at institutions with a long standing, the reUSE services have the same designated communities as the host institutions. These have been discussed in the Stakeholders chapter above (chapter 3.2.2). Additionally, some demonstrators plan to introduce digital feedback mechanisms via their online gateway to allow some interactivity. UBER even instituted regular discussion meetings with some selected representatives from their designated community. As the demonstrators further evolve all partners conceive possible business models for the services, and further ahead in the reUSE project in the scope of Work Package 4 the partners will address possible added-value services. Specialised added-value services may alter the designated community slightly, or may focus on subsets of the whole community. In a nutshell, the demonstrators are working with a clear picture of their designated community following their host institutions, however this picture will be further refined in the ongoing demonstrator development, and the designated community will be continuously monitored as the respective repositories move into their operational phase.

> *Ensure that the information to be preserved is independently understandable to the Designated Community. In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information.*

Since all publications collected in the scope of reUSE are publications with a wide audience, it is largely guaranteed that they are independently understandable. In most cases editors will as part of the usual publishing process also ensure this. None of the reUSE demonstrators conducts an analysis of the content of the publications, as this would be infeasible and also unnecessary. Regarding a documentation of the context of a publication, reUSE demonstrators collect limited metadata as described above.

> *Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, and which enable the information to be disseminated as authenticated copies of the original, or as traceable to the original.*

reUSE partners are working towards a framework of policies and procedures. Yet, at the current stage of system development, workflow design, and negotiations with potential depositors the partners are not yet in the position of issuing comprehensive formal policies and procedures, at least none that are sufficiently stable. All partners have strategic objectives that remain largely unwritten at this point of time, and flexible procedures that are adopted as needed. UBER – disposing of a repository that is operational for several years – is currently in the process of consolidating its policies and procedures, and extending its written documentation. ONB already has formal selection criteria, a generic submission workflow, author guidelines for document formats, and similarly targeted documents, which

are tied together by their digital preservation strategy that is endorsed by ONB senior management. All these activities aim towards a complete and stable strategic environment for establishing a trusted digital repository. Also the other partners are working on a strategy framework, which is suitable for their respective technological and organisational environment and which balances the requirements of the specific stakeholders involved.

*Make the preserved information available to the Designated Community.*

All reUSE demonstrators attach great importance to the accessibility of their collections and the usability of their access gateways. Access channels and related issues are discussed in the upcoming section on the OAIS functional entity Access (chapter 3.5.3.6).

### 3.5.2 OAIS information model

In order for an OAIS to fulfil its responsibilities, it needs to store significantly more information than only the actual content information. The OAIS information model[185] describes the information requirements on a conceptual level, without implying any specific implementations. From a workflow perspective, the OAIS distinguishes between the information that is submitted to the archive (Submission Information Package, SIP), the information stored in the archive (Archival Information Package, AIP), and the information provided to the user at access (Dissemination Information package, DIP). When drilling down into the Archival Information Package, the OAIS identifies a variety of information types including Content Information, Preservation Description Information, Packaging Information, and Descriptive Information. These information packages as they are implemented at the reUSE demonstrators are subject of the following analysis.

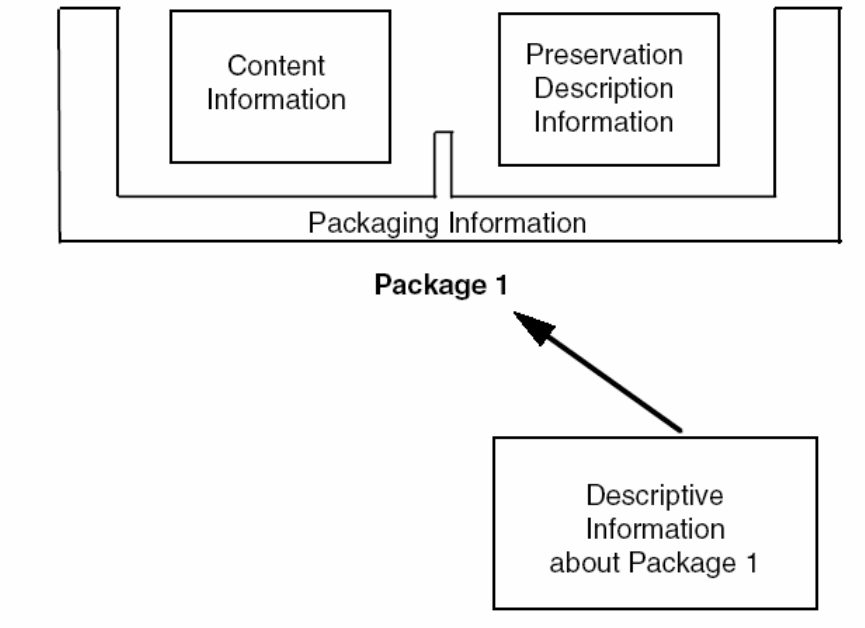



**Figure 2: OAIS Fig. 2-3: Information Package Concepts and Relationships**

[185] OAIS (Fn. 184), p. 4-18–4-47.

While all demonstrators underline that the OAIS-typical distinction between Submission Information Package (SIP), Archival Information Package (AIP), and Dissemination Information package (DIP) has shaped their thinking about repository functions, none of them took the in-depth OAIS information model as a key input to their system design. Moreover, the information models at all demonstrators are subject to regular revisions to come up for gained experiences, the dynamic organisational environment, service development, ongoing demonstrator refinement, and also changing standards in the still young and dynamic international field of digital repositories. Despite this fluctuating state of the demonstrator's information models, the upcoming paragraphs highlight and compare some of their traits along the information package concepts as illustrated in the graphic of the OAIS model above.

#### *3.5.2.1 Content Information*

The content information collected by the reUSE demonstrators are largely static documents based on text and images such as books, journals, reports, newspapers, working papers, and studies. This clear restriction of the content type still yields to numerous different file formats including XML, HTML, Microsoft Word, Postscript, and PDF. Each demonstrator has a slightly different approach to this myriad of possible ingest formats (SIPs), which also defines their preservation format (AIPs) and eventually the possible access format (DIPs).

ALO takes publications in any format at this point of time and will review this approach with gained experience. The format of the submitted object, of the archival object, and of the access object are essentially the same, though ALO expects to assume a preservation approach based on migration as necessity rises. NLE also accepts largely any format provided by the data producer though it has a set of preferred formats. For formats other than the preferred formats the NLE does not assume responsibility for their long-term preservation apart from preserving the bit-stream.[186] The ONB on the other hand actively converts some of the submitted publications into formats suitable for preservation where this is viable (while archiving the original bitstream as well).

UBER has the approach that is most conducive to long-term preservation. UBER actively promotes the usage of document templates[187] when creating LaTeX, Microsoft Word, or OpenOffice documents. These templates enable a largely automatic conversion to an XML based format. UBER will preserve the original source file, a PDF version, as well as the XML. For access, the XML representation enables the automatic generation of a variety of other data formats, and UBER already provides on-the-fly generation of HTML. However, this comprehensive approach is only possible as UBER has some level of influence on the data providers and comes at the cost of a high effort for guiding authors, auditing submitted publications and manually correcting mistakes.

As part of their long-term preservation strategy, ONB is creating guidelines for the creation of documents. Their first guidelines address PDF documents.[188] While these are barely recommendations at this point of time, the library may point out to the data provider that non-

[186] This approach is similar to the three levels of preservation by DSpace, which distinguish between supported formats, known formats, and unsupported formats. Cf. the FAQ on the DSpace website "How does DSpace preserve digital material?", http://www.dspace.org/faqs/#preserve.
[187] Humboldt University Institutional Repository: Author Guidelines (in German), http://edoc.hu-berlin.de/e_autoren/.
[188] The PDF guidelines are available via the website of the ONB digital preservation department, http://www.onb.ac.at/about/lza/.

conformant resources are unsuitable for long-term preservation, or they may even reject such resources. The PDF guidelines are shared and discussed with all reUSE partners, and other demonstrators have expressed their interest in drafting similar recommendations. Furthermore, if ONB converts documents at ingest to a format that is more suitable for preservation, they preserve both the original source file as well as the converted copy, and all migration copies are handled equally with regard to future preservation action.

In the context of Content Information the OAIS model puts a great emphasis on the collection and preservation of 'representation information', which refers to documentation and specifications of specific data formats.[189] As mentioned in a previous section[190], the preservation community is working towards the creation of international data format registries. Though it may still take some time till these are operational, the reUSE partners wait for the developments in this respect. ONB for instance allows for the possibility of linking each format to a future format registry. Only NLE assumes the considerable effort of creating format documentation and aims to create an internal format registry themselves.

| | ALO | NLE | ONB | UBER |
|---|---|---|---|---|
| **ingest format restriction** | × take everything at the moment and will review with gained experience | × ASCII, SGML, XML, HTML, TIFF, GIF, JPEG, PNG, PDF, PS, DOC; accepts also "unknown formats" but without preservation guarantee | ✓ AIFF, ASCII, GIF, JPEG, JPEG2000, PDF, TIFF, UTF8, WAVE, XML – files may be rejected; and in some cases copies of the submitted files are converted to one of these formats by ONB staff | ✓ original source files (DOC, TEX, SXW; PNG, JPG, GIF, TIFF); original print master-file (PDF); XML conversion from source file if possible |
| **object formatting guidelines** | × | × | ✓ guidelines for PDF files are available, for other submission formats they are planned | ✓ templates for each of the formats listed above |
| **object storage format** | METS database | Fedora XML objects (FOXML) | XML data containers (planned for production system) | original files; XML, PDF, and LaTeX according to stringent guidelines |
| **Compression / encryption of** | × | × | × | × |

[189] Actually, the OAIS is more general and defines Representation Information as: "The information that maps a Data Object into more meaningful concepts. An example is the ASCII definition that describes how a sequence of bits (i.e., a Data Object) is mapped into a symbol. " – OAIS (Fn. 184), p. 1-13.
[190] Cf. chapter 2.4.2.

| | | | | |
|---|---|---|---|---|
| **content files** | | | | |
| **data format documentation** | waiting for an international format registry to be created | creating internal format registry | linkage to future format registry | no; metadata and documentation standards are being reviewed |

#### *3.5.2.2 Preservation Description Information*

Preservation Description Information as defined by the OAIS is information that facilitates the understanding of content information over an indefinite period of time. Currently, metadata at the reUSE demonstrators is mostly of a bibliographic nature and lacks a preservation component. The ONB has adopted the NLNZ preservation metadata schema as a model to their metadata. They also feature MIX technical metadata for images[191], and are capable of flexibly accommodating in their production system any metadata schema that has an XML representation. ALO provides for DIG35 metadata for images[192], which includes technical metadata that is also relevant for preservation. Overall, the demonstrators are in the process of refining their preservation requirements and authenticity criteria, and subsequently their preservation metadata needs.

Drilling down in the Preservation Description Information, the OAIS distinguishes between four conceptual modules, being Reference Information, Context Information, Provenance Information, and Fixity Information. It also features an example of Preservation Description Information at Digital Library Collections. This example (table 4-1 of the OAIS) is the basis for the subsequent descriptions and the added table.

Regarding Reference Information, all demonstrators exchange bibliographic metadata with the traditional library catalogues of their host institutions. Whenever specific metadata is missing, this metadata is created manually by repository staff. The creation of metadata by the producer or the transfer of existing metadata from source systems to the repository is at this point of time limited, on a voluntary or case-to-case basis. Bibliographic metadata is perceived as sufficient for documenting the provenance of master files of print publications. However, publications are increasingly also published online or they are only available in electronic form. Some demonstrators still lack a reference to online digital versions on the author's website, or they do not plan to add such a reference for the inherent volatility of these online resources. Since all publications included into the repositories are final, there is no need for managing multiple versions. UBER and ONB additionally preserve all digital surrogates created during migration steps, which yields multiple formats for the same version of a publication. There is no particular preservation of context information since all publications are expected to be self-contained and self-explanatory.

[191] NISO Metadata for Images in XML Schema (MIX), http://www.loc.gov/standards/mix/.
[192] International Imaging Industry Association: DIG35 Initiative Group, http://www.i3a.org/i_dig35.html.

| | ALO | NLE | ONB | UBER |
|---|---|---|---|---|
| **bibliographic description** | link to MAB2 record in library catalogue | transferred from union catalogue (MARC21) | exchanged with library catalogue (MAB2) | exchanged with library catalogue (MAB2) |
| **persistent ID** | × URN (urn:nbn:at) planned | × internal ID, no international persistent ID | ✓ URN (urn:nbn:at) | ✓ URN (urn:nbn:de via DDB) |
| **audit trail** | × | ✓ automatic Fedora feature | ✓ (planned for production system) | × at the moment no systematic documentation |
| **integrity checks** | × | ✓ MD5 checksums | ✓ MD5 checksums, disc-scrubbing | ✓ RSA-Keys |
| **authenticity indicator** | × | × | × | ✓ (in development) digital signatures and timestamps |
| **metadata standard compliance (AIP)** | METS, Dublin Core, DIG35 | Dublin Core | METS, Qualified Dublin Core, NLNZ (with adaptations), MIX (for images); PREMIS, TextMD, AMD, VMD planned for production system | Qualified Dublin Core (with extensions) |

#### *3.5.2.3 Packaging Information*

At all demonstrators there is a clear move towards the usage of files with attached metadata for archival storage, as this is considered a more robust and scalable solution. For instance, the NLE's Fedora repository embeds content data together with metadata in XML files, and indexes these archival objects in a database for increased accessibility. The next version of ONB's software will have a similar architecture.

Since content data, metadata including preservation description information, and descriptive data are potentially separate components, Packaging Information is needed to bind the different components together into an identifiable entity. All demonstrators, of course, have an unambiguous linkage between any separate parts of the archival information package, though in some cases the system of internal identifiers and other packaging information remains to be documented.

| | ALO | NLE | ONB | UBER |
|---|---|---|---|---|
| **storage metadata - content data** | separate: metadata in database, content files in file system | content files are separate from metadata in XML files; selected metadata are indexed in MySQL database | Metadata in database, objects in file system;<br>production system: XML container for metadata in database and in file system, objects in file system | metadata in database, objects in file system |

#### *3.5.2.4 Descriptive Information*

As mentioned above, the demonstrators collect various bibliographic metadata, which is linked to the traditional library catalogues. Additionally these metadata are indexed and available for search in the access interfaces of the repositories. These interfaces are described in more detail in the following mapping of the OAIS Functional Model.[193]

### 3.5.3 OAIS functional model

The OAIS functional model[194] presents a number of system functionalities and human responsibilities clustered according to the six functional entities: Ingest, Archival Storage, Data Management, Administration, Preservation Planning, and Access. The following mapping highlights a number of technological aspects rather than dissecting all aspects of the OAIS along the dimensions policies / standards and people / systems. At the current stage of demonstrator implementation it is too early to comprehensively contrast these various dimensions that are largely still in flux, although such a broad perspective would better capture the nature of the functional model. The following mapping is clustered along the six functional entities and cites the short entity description by the OAIS document, summarises some general reUSE related observations, and lists a variety of atomic system features of the demonstrators for comparison.

#### *3.5.3.1 Ingest*

> This entity provides the services and functions to accept Submission Information Packages (SIPs) from Producers (or from internal elements under Administration control) and prepare the contents for storage and management within the archive. Ingest functions include receiving SIPs, performing quality assurance on SIPs, generating an Archival Information Package (AIP) which complies with the archive's data formatting and documentation standards, extracting Descriptive Information from the AIPs for inclusion in the archive database, and coordinating updates to Archival Storage and Data Management.
>
> OAIS, p. 4-1

Ingest is arguably one of the most important processes for reUSE demonstrators, and also perhaps the most work intensive task now and in the operational phase of the repository. The

[193] See the mapping of the functional entity access in section 3.5.3.6.
[194] OAIS (Fn. 184), p. 4-1–4-18.

transfer of publications from the data suppliers to the reUSE service is a delicate task from an organizational as well as technical perspective. After all, this step is the basis for a comprehensive archive with high-quality collections. Adequate digital objects and rich metadata are the result of a tight cooperation between the data supplier and the reUSE service, and ultimately a well-designed workflow of the whole transfer and ingest process.

In an optimal situation the producer would transfer his digital objects with annotated metadata in the required form to the library's digital repository for direct ingest. However, if such an optimal workflow cannot be attained and if, for example, the producer fails to supply the requisite metadata, the repository has to create the metadata manually in what is a time-consuming process. Other typical ingest tasks include a virus check of the resource after some time of quarantine, and quality checks regarding content, structure and format. The submission of an unsupported format or of a format violating institutional guidelines may call for conversion action to be taken. Flaws in the submitted resource or a lack of information may necessitate the correction and resubmission by the data supplier.

As an impeding factor in the creation of an optimal workflow, the digital repository may be faced with a variety of heterogeneous systems at the producers. The systems, data and metadata formats may vary considerably between the various data suppliers and the ingest process may need to be tailored to the specific situation. Also, reUSE demonstrators have made the experience that data providers often are cooperative only to such a degree where they do not incur any extra effort. Minimal effort, of course, can only be achieved with a workflow that is as smooth and as automatic as possible. This calls for the careful design and iterative improvement of workflows, and the deployment of (semi-)automatic tools, for example for metadata extraction.

To come up for all these requirements and dependencies between the data producer and the repository, the cooperation between the two parties needs to be defined in an adequate submission agreement. The agreement defines not only the transfer route – such as FTP or offline CDs – but also the amount of data to be transferred, metadata, data format guidelines, and the delivery schedule.[195] The negotiation of submission agreements together with the selection and contacting of potential data suppliers, and the promotion of the service are some of the tasks that need to be conducted before ingest can start. These pre-ingest activities are actually part of another OAIS functional entity, namely 'Administration'.

For a detailed step-by-step description of the phases of the cooperation between data producer and the repository, the OAIS “Producer-Archive Interface Methodology Abstract Standard”[196] and also the ERPANET Ingest Guidance[197] are reference documents with all-encompassing descriptions and useful checklists. In fact, UBI builds strongly on the OAIS Producer-Archive Interface Methodology and undertook to tailor the document to their specific situation. English and German versions of this downsized document have been shared and discussed with all reUSE partners.

[195] Refer to chapter 3.3 on Submission Agreements for more details.
[196] Producer-Archive Interface Methodology Abstract Standard (Fn. 137).
[197] ERPANET Guidance: Ingest Strategy (September 2004), http://www.erpanet.org/guidance/docs/ERPANETIngestTool.pdf.

| | ALO | NLE | ONB | UBER |
|---|---|---|---|---|
| **submission agreements** (negotiated individually) | ✓ | ✓ | ✓ | ✓ |
| **receive: bulk transfer** | ✓<br>FTP | ✓<br>FTP | ✓<br>FTP | × |
| **receive: online submission form** | × | ✓ | ✓ | ✓ |
| **receive: web harvesting** | × | ✓<br>HTTrack | ✓<br>HTTrack | × |
| **receive: offline** | any offline media | CD, DVD | any offline media, preferably CD, DVD | CD, DVD |
| **receive: email** | ✓ | ✓ | ✓ | ✓ |
| **virus check** | ✓ | ✓ | ✓ | ✓ |
| **quality check** | ✓<br>according to submission agreement | ✓ | ✓<br>according to submission agreement | ✓ |
| **conversion at ingest** | × | × | ✓<br>on a case-to-case basis, original file is also preserved | ✓<br>original file is also retained |
| **metadata and documentation by producer** | × | ✓<br>optionally, if supplied by the producer | ✓<br>on a case-to-case basis | ✓<br>a mandatory set of metadata is demanded at submission via the web interface |
| **manual / automatic metadata creation** | manual;<br>some technical metadata is extracted automatically | manual;<br>technical metadata is extracted automatically | automatic creation of technical metadata where possible, other metadata are created manually | manual<br>(limited automatic creation is planned) |

### *3.5.3.2 Archival Storage*

> This entity provides the services and functions for the storage, maintenance and retrieval of AIPs. Archival Storage functions include receiving AIPs from Ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the media on which archive holdings are stored, performing routine and special error checking, providing disaster recovery capabilities, and providing AIPs to Access to fulfil orders.
>
> OAIS, p 4-1f

The low-level storage functionalities in this OAIS entity are handled by both the storage facilities and the repository software. Other than suggested by the OAIS reference model, only one of the four demonstrators, ALO, builds on a hierarchical storage management system. The others operate with online disk arrays and tape backup facilities. For all demonstrators the storage facilities are managed as part of the institutional technology infrastructure. Accordingly, refreshing, error checking, disaster recovery plans and the like are established by those responsible for the institutional technology infrastructure. Close collaboration between those responsible for the corporate preservation strategy and the respective institutional technology departments ensure the organisational viability and technological suitability of this arrangement.

| | **ALO** | **NLE** | **ONB** | **UBER** |
|---|---|---|---|---|
| **hardware storage** | hierarchical storage management (IBM Tivoli) | external disk array (Fujitsu-Siemens PRIMERGY SX30) | disk array | disk array |
| **backup** | tapes, geographically separate | tapes, geographically separate (planned) | replicated disk array and tapes, geographically separate | tapes, geographically separate |
| **disaster recovery** | according to university standards | replicated disk array | tapes in governmental high-security silo in more than 300 km distance | according to university standards |
| **refreshing** | according to university standards | incremental backup, and a complete backup every 7 days | incremental backup, and a complete backup every 7 days | according to university standards |

### *3.5.3.3 Data Management*

> This entity provides the services and functions for populating, maintaining, and accessing both Descriptive Information which identifies and documents archive holdings and administrative data used to manage the archive. Data Management functions include administering the archive database functions (maintaining schema and view definitions, and referential integrity), performing database updates (loading

> new descriptive information or archive administrative data), performing queries on the data management data to generate result sets, and producing reports from these result sets.
>
> OAIS, p 4-2

Data management functions in this OAIS entity are mostly inherent repository system features. The two demonstrators with a university background, ALO and UBER both have a self-made system, whereas the NLE installs Fedora and the ONB a commercial system. All systems aim to be flexible and extensible to support maintenance over time. Their compliance with XML and METS for metadata management is part of this strategy.

Another preservation-related functionality is the audit trail. An audit trail records any changes to a digital object, which includes the application of preservation methods such as successive migration to current formats. It thus is an essential tool in establishing the provenance and consequently the authenticity of the digital object. reUSE demonstrators are in the process of defining their preservation strategies, and therefore audit trails may become a high-priority feature of all reUSE repositories in the future. In the context of the formulation of a comprehensive preservation strategy, reUSE partners may also address whether the integrity of archival files can be checked by use of checksums or similar codes, or whether it is ensured at a low system level that archival objects cannot be manipulated without a record of that change or even a new version of the object being created. Whether or not such or similar system features are needed, of course, demands weighing the risks and requirements regarding the specific service.

Another high priority data management goal is the integration of the reUSE specific system with other institutional services, as it potentially raises quality and accessibility of the services while reducing costs. In the previous section on workflows (chapter 3.4) the ONB use case highlighted the possibility for metadata exchange between the demonstrator and the existing library system. Also references in the library catalogue to reUSE objects are part of the drive for better integration. Comprehensive integration, of course, can only be achieved in tandem with organisational measures and on a strategic level.

| | **ALO** | **NLE** | **ONB** | **UBER** |
|---|---|---|---|---|
| **embedding** | linkage to ALO from university catalogue and Austrian union catalogue[198] | linked with OPAC and other bibliographical databases | metadata exchange with library catalogue (planned);<br>linked with Austrian union catalogue | metadata exchange and linked with regional union catalogue[199] |
| **audit trail** | × | ✓<br>automatic Fedora feature | ✓ (planned for production system) | ×<br>at the moment no systematic documentation |

[198] Austrian library union catalogue based on ALEPH system, http://www.bibvb.ac.at/aleph.htm. – ExLibris: ALEPH integrated library system, http://www.exlibrisgroup.com/aleph.htm.
[199] The ALEPH system of the regional library union: Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV), http://www.kobv.de.

| | | | | |
|---|---|---|---|---|
| **integrity checks** | × | ✓<br>MD5 checksums | ✓<br>MD5 checksums, disc-scrubbing | ✓<br>RSA-Keys |
| **archival files can/not be changed** | manual changes possible | manual changes possible with automatic audit trail entry | strategy decision for invariant files but technically not enforced | invariant;<br>but deletion is possible with the requisite rights |
| **usage logs** | ✓ | ✓ | ✓ | ✓ |
| **generate reports** | × | monitoring system status information | monitoring system status information and generating regular system reports for digital preservation department | automatic reports |

#### *3.5.3.4 Administration*

> This entity provides the services and functions for the overall operation of the archive system. Administration functions include soliciting and negotiating submission agreements with Producers, auditing submissions to ensure that they meet archive standards, and maintaining configuration management of system hardware and software. It also provides system engineering functions to monitor and improve archive operations, and to inventory, report on, and migrate/update the contents of the archive. It is also responsible for establishing and maintaining archive standards and policies, providing customer support, and activating stored requests.
>
> OAIS, p 4-2

Policies and standards remain to be fully developed as the demonstrator implementation proceeds. Once the demonstrators are in full production with the necessary policies in place, there will be means for audit and measurement, as demanded by the "Trusted Digital Repositories" report.[200] Currently the most active component of this functional entity at all demonstrators is the refinement of template submission agreements and contacting potential data providers. To date UBI, for instance, collected about 2.000 addresses of potential data providers, clustered them into focus groups, and is preparing the installation of a Customer Relationship Management tool together with other ALO consortium partners to be able to handle this enormous amount of contacts.

The Administration entity is at all reUSE demonstrators tightly integrated in the organisational environment. Those tasks of the functional entity that involve infrastructural issues, technological or other, are handled by the host institutions and along their policies.

[200] RLG Working Group on Digital Archive Attributes: Trusted Digital Repositories: Attributes and Responsibilities. RLG, OCLC Report. Mountain View CA 2002, http://www.rlg.org/longterm/repositories.pdf.

| | ALO | NLE | ONB | UBER |
|---|---|---|---|---|
| **submission agreements** (negotiated individually) | ✓ | ✓ | ✓ | ✓ |
| **ongoing system maintenance** | ✓<br>hardware by Innsbruck University information technology, software by ALO consortium | ✓<br>by NLE Information Systems Department | ✓<br>cooperatively by the information technology and the digital preservation departments | ✓<br>by UBER information technology department |
| **physical access control** | according to institutional standards | according to institutional standards | according to institutional standards | according to university standards |
| **audit and update of standards and policies** | internally on an ad-hoc basis | internally on an ad-hoc basis | internally on a regular basis, every 6 months | internally on an regular basis, every 12 months |

#### *3.5.3.5 Preservation Planning*

> This entity provides the services and functions for monitoring the environment of the OAIS and providing recommendations to ensure that the information stored in the OAIS remains accessible to the Designated User Community over the long term, even if the original computing environment becomes obsolete. Preservation Planning functions include evaluating the contents of the archive and periodically recommending archival information updates to migrate current archive holdings, developing recommendations for archive standards and policies, and monitoring changes in the technology environment and in the Designated Community's service requirements and Knowledge Base. Preservation Planning also designs IP templates and provides design assistance and review to specialize these templates into SIPs and AIPs for specific submissions. Preservation Planning also develops detailed Migration plans, software prototypes and test plans to enable implementation of Administration migration goals.
>
> OAIS, p 4-2

Preservation Planning is the OAIS entity which is least developed among reUSE demonstrators at this point of time. However, preservation is mainly a strategic issue, and the partners are confident to catch up with ongoing implementation of the demonstrators.

UBER has relatively stable goals and procedures, though they remain unwritten to date. They are now in the process of consolidating these strategies and practices, supplementing them where necessary and writing them up in a formal policy framework. UBER employs XML as preservation format from which a variety of access formats can be created dynamically and automatically. Additionally, the original source file and a PDF version are preserved, along with all intermediate versions created through successive migration steps.

As part of their participation in the OpenOffice-Developer-Group, UBER instituted an ample technology watch.

The National Libraries of Austria and Estonia are converging towards a preservation strategy similar to the approach by DSpace[201], which applies preservation methods on some selected file formats and informs the depositor at ingest whether a format is supported, known but unsupported, or unknown and unsupported.

| | ALO | NLE | ONB | UBER |
|---|---|---|---|---|
| **monitor designated community** | × | ✓ | ✓ | ✓ interaction with designated community by way of discussion meetings |
| **documentation standards** (systems and resources) | × | × | × | × |
| **technology watch** | × on an informal basis | ✓ internal technology watch | × (planned) | ✓ partner of the OpenOffice-Developer-Group |
| **format restriction** (preservation format) | × | ✓ | ✓ | ✓ |
| **data format documentation** | wait for an international format registry to be created | create internal format registry | linkage to future format registry | no systematic documentation; metadata and documentation standards are being reviewed |
| **definition of authenticity, significant properties** | × | × | × | × |

[201] Julie Walker: DSpace, digital preservation, and business models. Presentation at the ERPANET Seminar "Business Models related to Digital Preservation". Amsterdam (September 20–22, 2004), http://www.erpanet.org/events/2004/amsterdam/presentations/erpaTraining-Amsterdam_Walker.pdf.

#### *3.5.3.6 Access*

> This entity provides the services and functions that support Consumers in determining the existence, description, location and availability of information stored in the OAIS, and allowing Consumers to request and receive information products. Access functions include communicating with Consumers to receive requests, applying controls to limit access to specially protected information, coordinating the execution of requests to successful completion, generating responses (Dissemination Information Packages, result sets, reports) and delivering the responses to Consumers.
>
> OAIS, p 4-2

All reUSE demonstrators endeavour to establish popular services that fit the needs of their designated communities. The user interface of choice is, of course, the web interface with various means for search and retrieval. But also other channels including FTP access, CDs and print on demand are being installed. The partners will explore other added-value services in the scope of the reUSE project[202], and design suitable business models as the demonstrator implementation proceeds and as potential added-value services are identified.

Generally, reUSE material is openly accessible, however most demonstrators account for access restrictions for supplementary material that is out of the scope of the reUSE project. Access restrictions may thereby range from metadata access only to no restrictions, and rights are applied on an object level for an individual user. Most reUSE demonstrators already implement an OAI interface, which is considered a key feature in the digital repository community. In fact the DINI certificate[203] and other initiatives from the open access movement[204] call for the installation of OAI.

[202] Added-value-services will be addressed as part of reUSE Work Package 4.
[203] Arbeitsgruppe Elektronisches Publizieren: DINI Zertifikat Dokumenten- und Publikationsserver [DINI Certificate for Open Access Document Servers], http://www.dini.de/documents/Zertifikat.pdf.
[204] A OAI interface is, for example, a key selection criteria for featuring repository software in the OSI Guide: Open Society Institute (OSI): A Guide to Institutional Repository Software. 3rd Edition, August 2004, http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf.

| | ALO | NLE | ONB | UBER |
|---|---|---|---|---|
| **rights management** | open access to all objects | user authentication for object rights | user authentication for object rights | user authentication for object rights |
| **external system interfaces** | SOAP, OAI (planned) | OAI | Z39.50, OAI, SOAP | OAI, ProPrint |
| **web interface** | ✓ | ✓ | ✓ | ✓ |
| **other access interfaces** | FTP, CD on request | FTP (for depositors for getting back their materials on request) | × | Proprint print-on-demand |
| **search interface** | boolean | boolean metadata search | boolean, metadata, and full-text search | metadata search |
| **access formats** | original format is supplied for access | the formats accepted at ingest | PDF, DOC, ZIP, HTML, ISO (for disc images), RTF, TIFF, JPEG, PNG … (may be enlarged) | author submitted PDF, and HTML generated on-the-fly from XML if available |

## 3.6 Outlook

reUSE partner institutions embarked on an ambitious project. The preservation of digital master files of print publication has not previously been conducted in practice on such a large scale with so many different actors involved. While the challenges regarding system deployment and digital preservation planning are considerable, a salient challenge is of an organisational nature, namely the relationship with data producers and the future business model of the repository. Tackling these challenges, the priority activities at the time of writing of this paper include establishing contact and collaboration with potential data providers and workflow design, in tandem with system deployment and the integration of the reUSE service in the technological and organisational context. As the project unfolds, and new information and experiences are gathered in a step-by-step process, system design and organisational planning is improved iteratively. The cooperation of the reUSE partners is crucial in ensuring that this iterative process is swift and robust. reUSE offers the possibility for exchanging experiences on collaboration with data providers and organisational approaches, and comparing financial numbers and the success of business models. Such practical experiences and numbers are scarce in the young field,[205] yet indispensable for business planning and establishing a sustainable and trustworthy service. It is thereby particularly valuable that reUSE partners are from a range of institutions with different organisational settings and, hence, different approaches to establishing a digital repository service. The comparative analyses yield optimized planning, new perspectives, and may eventually lead to a series of independent yet cooperating, trusted digital repositories.

This White Paper, which was composed as part of reUSE Work Package 1, had the dual purpose of providing an environmental scan of international developments in the field of digital repositories, and to reflect the current status of reUSE demonstrator development, a snapshot of a rapidly advancing process. The authors also had the goals to raise awareness among the reUSE partners on the importance of repository design and strategic issues, and to share the notions and requirements of individual demonstrators among all reUSE partners in order to facilitate discussion and the improvement of demonstrators.

Another milestone on the path towards trustworthiness of the individual digital repositories will be the demonstrator evaluation, which is set to be conducted in the framework of reUSE Work Package 3. The existence of methodologies for system evaluation is, in fact, a core aspect of trustworthiness.[206] None of the partners fulfils this requirement at this point of time, yet the results and experiences of the reUSE evaluation may feed into ongoing evaluation at the reUSE partner institutions. The National and University Library of Slovenia, Die Deutsche Bibliothek and the University of Ljubljana Faculty of Civil and Geodetic Engineering are in the process of designing the evaluation framework accordingly.

Furthermore, reUSE will examine ways towards rising the accessibility of reUSE material, establishing sustainable business models, and boosting the trust of all stakeholders in the reUSE repositories. Work Package 4 therefore investigates enhanced and advanced means for installing added-value services, such as print-on-demand or advanced information retrieval. These services will raise the standing of the individual repository, and it will also contribute to building a market place for digital content.

[205] Cf. ERPANET Seminar: Business Models related to Digital Preservation. Final Report; Amsterdam, (September 20–22, 2004), http://www.erpanet.org/events/2004/amsterdam/Amsterdam_Report.pdf.
[206] Trusted Digital Repositories (Fn. 200).

# 4 Annex

## 4.1 reUSE Partner Organisations

### 4.1.1 The University Library of Innsbruck

The foundation of the University Library of Innsbruck[207] is marked by the year 1745. The University of Innsbruck[208] is about 100 years older, and had been installed as a so called "full" university (covering all scientific disciplines).The Foundation Charta of the library quotes its status as "an institution for scientific and public use", and therefore belongs to the first public libraries in Europe in the modern times.

Today the library is ranking 4th in Austria concerning the dimension of stored media: ca. 2,8 million volumes (including ca. 1.100 manuscripts and 2.000 incunabula), 6.500 scientific journals and periodicals, a wide range of online-databases, electronic journals and a growing number of full-text electronic resources which are made available campus-wide.

The Library is a member of the Austrian Library Consortium[209], a network more than 60 scientific libraries in Austria (among them the Austrian National Library and all university libraries). In January 1999 the first libraries of this consortium (among them the University Library of Innsbruck) switched to the Ex Libris[210] library system ALEPH-500. Aleph is used both locally both as a network solution in Austria. Innsbruck is playing an important and leading part in the development and testing of the new programme features.

Apart from supporting the University's research and teaching across a full range of subjects, the Library functions as the regional centre library for the Province of Tyrol. Therefore the library's special collections are of regional and historical importance. The substantial part forms a mostly unique and almost complete collection of historical Tyrolean printings (from its beginnings in the 16th c. to 1850). Being intensely used both by members of the university both by many visitors from elsewhere this collections functions as test-site for the METAe[211] Project.

### 4.1.2 The National Library of Estonia

**History**

The history of the National Library of Estonia[212] begins on 21 December 1918, when the Government of the Estonian Republic passed a law creating the State Library with the aim of collecting printed matter essential for governing the new state. In 1989, the National Library of Estonia regained its status as a parliamentary library responsible for serving the information needs of Parliament and Government. Today, the National Library of Estonia is a legal entity in public law which operates pursuant to the National Library of Estonia Act, and the Articles of Association approved by the Supervisory Board of National Library.

[207] University Library of Innsbruck, http://www.uibk.ac.at/ub/.
[208] University of Innsbruck, http://www.uibk.ac.at/.
[209] Austrian Library Consortium / Österreichischer Bibliothekenverbund, http://www.bibvb.ac.at/.
[210] Ex Libris, http://www.exlibrisgroup.com/.
[211] METAe – Meta Data Engine, http://meta-e.uibk.ac.at/.
[212] National Library of Estonia, http://www.nlib.ee/.

**Functions**

The National Library of Estonia is:

- a national library with the aim of collecting, preserving and making publicly accessible documents published in Estonia, in Estonian, or about Estonia, regardless of their place of publication
- a national centre of statistics on book publishing and on Estonian libraries
- the national ISBN and ISSN agency
- a parliamentary library with the aim of providing information services to the Riigikogu (Estonian Parliament), to the Government, and to governmental institutions
- a research library with the aim of providing information for research activities in the field of humanities and social sciences
- an information centre whose services are based on the use of local databases and other reference tools, as well as on-line search possibilities
- a centre for research on book preservation and conservation
- a professional development centre that provides LIS information to interested persons and further education in the LIS field in Estonia
- the location of the Estonian Librarians Association and the publishing group of the journal Raamatukogu (Library), the only Estonian scholarly publication for librarians
- a book research centre pursuing research in librarianship and information sciences and responsible for related development activities
- a member of IFLA and other leading international organisations in the field of librarianship and information sciences, such as LIBER, CENL, IALL, IAML, etc.
- a cultural centre where various book and art exhibitions are held along with concerts, conferences and other cultural activities

### 4.1.3 The Slovenian National and University Library (NUK)

The National and University Library (NUK)[213] incorporates two functions. It is the Slovenian national library and at the same time it acts as the main library of the University of Ljubljana. Within the frame of the national functions it also provides the national bibliographic service and national centre for library network development (advising and monitoring included), National Information and Referral Centre. There are also the Centre for Research and Development for Librarianship, and the National centre for Preservation of Slovenian Written Heritage. The library is among the fundamental institutions of Slovenian infrastructure, closely connected to cultural, educational and scientific life.

NUK is a public institution. Its founder is the Republic of Slovenia. As a public research library it was first founded in 1774, in 1921 it acquired the function of a university library, in 1945 it became the national library of Slovenia.

### 4.1.4 Institute "integrated study" – i3s3

The i3s3 department[214] was established in October 1991 as a model project which tries to support blind and visually handicapped students in their studies. The main part of the support

[213] The National and University Library (NUK), http://www.nuk.si/.

[214] i3s3 – Interuniversitäres Institut für Informationssysteme zur Unterstützung Sehgeschädigter Studierender / Austrian Institute for Information Systems and Support Services for Blind and Visually Handicapped University Students, http://www.integriert-studieren.jku.at/.

activities is the digital preparation of all studying materials such as books, lecture notes, overhead sheets, exercises, contents of the blackboard and so forth for blind and visually handicapped students all over Austria. Research and teaching is also mainly directed towards this field. In 1995 the model project was established as the Department Computer Science for the Blind. In 2000 an Austrian wide institute was established.

**Research activities:**

- research for principles of the non-visual man-machine-interaction for an adapted interface
- electronic libraries and electronic publishing for people with disabilities
- access to graphical user interfaces for the visually impaired users
- design for all of hard- and software
- adapting graphic-software for the support of the production of tactile pictures for the visually impaired (e.g. automatic lettering of pictures in Braille)
- research in pedagogical, psychological and social aspects of the use of the computers for the support of the integration of visually impaired
- access to notations of mathematics, chemistry and music for blind people
- practical teaching and supporting of blind students in all courses of studies (alternative concepts of representations of the methods of scientific fields)
- the social consequences of applying technology for the disabled
- inclusive teaching
- speech recognition applications for people with disabilities

The department is involved in the Austrian wide production of teaching and learning materials for blind and visually handicapped students in primary and secondary schools which is organised by the ALS (Arbeitsgemeinschaft zur Lehr- und Lernmittelerstellung für Sehgeschädigte). i3s3 is also involved in teachers education, especially concerning IT usage in integrated education.

Staff of the department has been working for several years as evaluator and reviewer for the European Commission.

### 4.1.5 University Library Graz

The University Library Graz[215] is the library for the Karl-Franzens-University Graz[216] (founded in 1585 as a Jesuit University, actually about 28.000 students) and open also for the general public. Legal deposit for Styria since 1781. The library consists of one main library, one faculty library (law and social sciences), nine branch libraries and about 100 institute libraries (staff: 120; librarians: 90; number of members: 65.000; number of visits: 600.000). Library Holdings: books 2.600.000, periodicals: 20.000, current periodicals: 10.000 i.e. volumes: 1.000.000. Core of library holdings: humanities, theology, social sciences, law, medicine, sciences, chemistry. UBG was the first Austrian Academic library to introduce a workstation for blind users and to train and employ a blind librarian.

[215] University Library Graz, http://www.kfunigraz.ac.at/ub/.
[216] Karl-Franzens-University Graz, http://www.kfunigraz.ac.at/.

University Library Graz is digitising some 2.000 of its medieval manuscripts for the UNESCO's Memory of the World programme. Excellent dept for conservation and restoration. UBG offers at present some 50.000 pages of 19th century scientific periodicals via the Web.

### 4.1.6 Austrian National Library

As the main scientific library of the Republic of Austria, the Austrian National Library (ONB)[217] can look back on a history rich in tradition dating to the 14th century. It is a living bridge between the rich heritage of the past and the claims of a modern informed society oriented toward the future.

The ONB regards itself as

- a centre of information and research set up to provide services;
- the country's excellent bank of memories;
- a multilevel centre of education and culture.

In accordance with the legislative act on research organisation, the following are defined as the Austrian National Library's main commitments:

- storing and recording all literature and other information carriers published in Austria
- collecting and recording all literature published abroad on Austria or Austrian men and women; further, in collaboration with other academic libraries, collecting and recording literature on the arts
- recording and complementing valuable holdings in special collections
- performing central tasks in Austrian libraries (e.g. publishing the Austrian bibliography, Austrian periodicals' database, librarianship training etc.)
- providing information service
- providing publications, exhibitions and other special events to publicise the library's holdings

As an information centre for providing services, the ONB offers its visitors access to and professionally competent advice on its own holdings (over seven million objects), and links to international databases as well. In addition it accepts research commissions and consults documentation centres and services, e.g., on literature devoted specifically to women. Since the beginning of the digital age a constantly growing portion of the service is carried out through the homepages of the ONB.

Because of a requirement of the Austrian Media Law the ONB is the only library in the country that receives a copy of every publication appearing in Austria, including university theses and products of the electronic media. Those obligatory items are simultaneously the basis on which the Austrian Bibliography is published. In addition to that the Library systematically chooses and collects literature from foreign countries that specifically refer to Austria, and literature on the humanities, which is of particular relevance for the collections.

In its ten collections the ANL, as the heir to the Habsburg Imperial Court Library, conserves an important part of the world's written cultural heritage and feels a duty to guard that permanently. Of special international significance are the holdings of manuscripts from Antiquity, the Middle Ages and the modern era, originating in the most varied cultural areas; add to that the music texts and the incunabula and old prints, historical maps, portraits and

[217] Austrian National Library (ONB), http://www.onb.ac.at/.

other pictorial documents, posters, ex-libris and pamphlets. No less important are the literary estates of Austrian authors. Three museums that are a part of the Library present the public with objects of the world's greatest papyrus collection, of a unique holding of historic globes of the earth, and an impressive collection on the international language Esperanto.

All of the collections are places of lively scientific research carried out in close contact with other scientific research institutes (partly on the basis of projects supported by the EU). The essential basis of that work is consistent formal gathering of the collection items that are constantly being supplemented. A particular challenge is that of gradually presenting the holdings digitally.

### 4.1.7 Computer and Media Services (CMS) of Humboldt University Berlin

The Computer and Media Service of Humboldt-University Berlin[218] is a central institution, the primary objective of which is to back the university's education, research and administration by providing the required computing technology – focussing on support and services, such as:

- maintaining and extending the university's computer network and providing connectivity to remote networks
- assisting faculties, institutes and central institutions with decisions on computing technology and supporting them in conceiving, extending and maintaining their local networks
- setting up and providing information services and supporting clients in setting up their individual information systems
- providing hardware services for decentralized computing technology and local networks
- providing software services and support in the application and purchase of software products
- supporting and training clients in using the services and systems, centrally provided, as well as the workstations and software
- conceiving, implementing and adapting data processing applications in the university's administrative sector
- providing media services such as maintaining the central and local WWW servers, providing services with visualization and animation, photography, teleconferencing, assisting with multimedia applications and projects for the university
- together with the university library supporting services for electronic publishing

The Computer and Media Service is actively participating in building new technologies for university systems infrastructure. In this context it has taken and takes part in several projects, including

- Electronic Publication of Dissertations / Digitale Dissertationen, http://edoc.hu-berlin.de/
- The Networked Digital Library of Theses and Dissertations (NDLTD), http://www.ndltd.org/
- UNESCO Guide on electronic theses and dissertations, http://www.etdguide.org/
- DINI – Deutsche Initiative für Netzwerkinformation / The German Initiative for Networked Information, http://www.dini.de

[218] Computer and Media Service of Humboldt-University Berlin, http://www.cms.hu-berlin.de/.

- Proprint – Printing on Demand, http://edoc.hu-berlin.de/proprint/

Further information about the projects undertaken by the Computing Center may be found at:
http://edoc.hu-berlin.de/e_projekte/

### 4.1.8 Die Deutsche Bibliothek

Die Deutsche Bibliothek[219] is the national library and national bibliographic information center for the Federal Republic of Germany. It is responsible for the collection, processing and bibliographic indexing of all German and German-language publications issued since 1913. Die Deutsche Bibliothek cooperates closely with all national and international library institutions and organisations. Within this context, it has a leadership role in the development and application of common rules and standards for Germany.

The cataloguing records created for the German national bibliography are kept in a data pool, from which manifold bibliographic services are generated in either printed or electronic form. These services are constantly being upgraded, like for example the service "DDB online" facilitating the transfer of bibliographic records in a structured form into the customer's system, and the document delivery service enabling to order excerpts from publications available in Die Deutsche Bibliothek. Furthermore, Die Deutsche Bibliothek hosts and maintains a number of national reference tools and authority files (union data catalogue of serials, name authority file, subject heading authority file) that are constantly being enhanced both in content and technically. As a national library Die Deutsche Bibliothek is striving for solutions concerning archiving, access, and long-term preservation of electronic publications. Since 1998 online dissertations and theses have been collected, archived, and made available on a document server. Since 2000 electronic periodicals have been collected, and since 2001 Die Deutsche Bibliothek has been operating a submission interface for online publications.

### 4.1.9 University of Ljubljana

The University of Ljubljana (LJU)[220] is the largest university in Slovenia and has over 56.000 students and over 3.700 staff. The participating group is the Chair of Construction Informatics from the Faculty of Civil and Geodetic Engineering. This group was a pioneer in the introduction of computers into the Slovenian construction industry. Computer integrated construction (CIC) was systematically researched at LJU since the late 1980s. LJU has been actively participating in the work of the CIB's working group W78 on CIC, the corresponding group WG6 of IABSE, in eCAADe and ECCPM. They have also been studying the computerisation of building regulations, the electronic technical document management, integrated information systems, and object-orientated data-bases.

In 1993, LJU was among the first in the field to start publishing on the Internet and the WWW (among the first 1.000 web servers world wide) and studied the role of communication media, such as the Internet, on the construction industry's practice. A number of frequently visited services were prepared, aimed at the international community of researchers and scholars. LJU also hosts the first Web based academic journals ITcon, an internationally renowned earthquake engineering database EASY, bibliographic database on CAAD CUMINCAD,

[219] Die Deutsche Bibliothek, http://www.ddb.de/.
[220] University of Ljubljana, http://www.uni-lj.si/.

software.orAEC.com and many others. LJU has been taking part in several EU TEMPUS and ERASMUS projects, in Esprit ToCEE, Esprit SCENIC, ETTN-CONNET and in IST-ISTforCE, IST-ICCI, IST ProdAEC and has been the coordinator of the IST project SciX.

## 4.2 reUSE Demonstrator Self-Assessment

### Survey on Trusted Digital Repositories for the project reUSE
### Conducted in October 2004

**Introduction**

reUSE will set up trusted digital repositories maintained by national and university libraries in order to collect, to preserve and to make available electronic documents which are currently published by public sector units in printed form only. Public sector units, as well as universities, NGOs, and publishers will be encouraged to use this service and to incorporate their digital content into the repository.

reUSE will secondly try to reach a critical mass of digital content which will allow to create new added value services from these digital data collections. reUSE will therefore not only improve accessibility of public sector information but make a contribution to build up a market place for digital content.

reUSE will finally be organized in a way that the implementation of its objectives is systematically evaluated on a European level and proved by critical observers interested in the effectiveness, impact, and results of the project. The organizational structure of the project will reflect this approach towards transparency, provide serious data and allow decision makers to build upon the projects outcomes.

**Purpose of this survey**

The following survey will aid to systematically evaluate organizational and technical aspects of the three project demonstrators of digital repositories in Austria, Estonia, and Germany as well as of other already established systems in order to describe their characteristics and point out some of the basic features that make a digital repository trustworthy.

**Resources**

The "Reference Model for an Open Archival Information System (OAIS)"[221] of the Consultive Committee for Space Data Systems represents the theoretical frame for this survey.

The questions were formulated mainly with reference to the following documents:

- dpc: Institutional Repositories in the context of Digital Preservation[222]
- RLG-OCLC: Trusted Digital Repositories: Attributes and Responsibilities[223]
- RLG: Digital Preservation Needs and Requirements in RLG Member Institutions[224]
- OCLC: PREMIS Implementation Survey[225]
- DINI: Fragebogen an Anbieter digitaler Dokumente[226]
- Open Archives Forum: OAI in Europe: Technical Validation Questionnaire[227]
- AWIICS: Certification (Best Practices) Checklist[228]

**Terminology**

The following terms are introduced within the OAIS reference model. Please refer to the definitions offered here when responding to the questions of the survey.

| | |
|---|---|
| *context information* | The information that documents the relationships of the Content Information to its environment. This includes why the Content Information was created and how it relates to other Content Information objects. |
| *designated community* | An identified group of potential consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. |
| *fixity information* | The information which documents the authentication mechanisms and provides authentication keys to ensure that the Content Information |

[221] http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf.
[222] http://www.dpconline.org/docs/DPCTWf4word.pdf.
[223] http://www.rlg.org/longterm/repositories.pdf.
[224] http://www.rlg.org/preserv/digpres.pdf.
[225] http://www.oclc.org/research/projects/pmwg/survey.doc.
[226] http://www.dini.de/fragebogen/tvqval.php4?my_initglobal=digdoc/digdocinitglobal.inc.
[227] http://www.oaforum.org/resources/tvquest.php.
[228] http://ssdoo.gsfc.nasa.gov/nost/isoas/awiics/CertifBase.ppt.

| | |
|---|---|
| | object has not been altered in an undocumented manner. An example is a Cyclical Redundancy Check (CRC) code for a file. |
| *provenance information* | The information that documents the history of the Content Information. This information tells the origin or source of the Content Information, any changes that may have taken place since it was originated, and who has had custody of it since it was originated. Examples of Provenance Information are the principal investigator who recorded the data, and the information concerning its storage, handling, and migration. |
| *reference information* | The information that identifies, and if necessary describes, one or more mechanisms used to provide assigned identifiers for the Content Information. It also provides identifiers that allow outside systems to refer, unambiguously, to a particular Content Information. An example of Reference Information is an ISBN. |
| *representation information* | The information that maps a Data Object into more meaningful concepts. An example is the ASCII definition that describes how a sequence of bits (i.e., a Data Object) is mapped into a symbol. |

**Questions**

***Repository***

**R3.3 What is the *operational status* of your repository?**

| | | | | | |
|---|---|---|---|---|---|
| o | in planning | o | in development | o | in production |

**R3.4 Which *types of documents and manifestations* do you store in your repository?**

| | | | | | |
|---|---|---|---|---|---|
| o | digital master files of printed materials | o | E journals/articles | o | preprints |
| o | monographs | o | reports | o | annuals |
| o | theses/dissertations | o | master/diploma theses | o | seminar papers |
| o | lecture scripts | o | conference proceedings | o | speeches |
| o | multimedial content | o | retrodigitized materials | o | web pages |
| | other: _____ | | | | |

**R3.5 Which of the following *mandatory responsibilities* for OAIS archives are, spontaneously answered, fulfilled by your repository at the moment?**

| | |
|---|---|
| Negotiate for and accept appropriate information from information Producers | o |
| Obtain sufficient control of the information provided to the level needed to ensure Long-Term Preservation | o |
| Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and, therefore, should be able to understand the information provided | o |
| Ensure that the information to be preserved is independently understandable to the Designated Community. In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information | o |
| Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, and which enable the information to be disseminated as authenticated copies of the original, or as traceable to the original | o |
| Make the preserved information available to the Designated Community | o |

**R3.6 Do you have any *comments* concerning these mandatory responsibilities?**

I: We are approaching these criteria but do not fulfill them.
E: All these mandatory responsibilities will be answered and fulfilled in the end of reUSE project.

**R3.7 Which of the following *expectations* according to trusted digital repositories mentioned in the RLG-OCLC-report are, spontaneously answered, fulfilled by your repository at the moment?**

| | |
|---|---|
| accept responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users | o |
| have an organizational system that supports not only long-term viability of the repository, but also the digital information for which it has responsibility | o |
| demonstrate fiscal responsibility and sustainability | o |
| design its system(s) in accordance with commonly accepted conventions and standards to ensure the ongoing management, access, and security of materials deposited within it | o |
| establish methodologies for system evaluation that meet community expectations of trustworthiness | |
| be depended upon to carry out its long-term responsibilities to depositors and users openly and explicitly | o |
| have policies, practices, and performance that can be audited and measured | o |

**R3.8 Do you have any *comments* concerning these expectations?**

**R3.9 Do you plan to *certify* your repository as a trusted digital repository as soon as this is possible?**

o yes o no

**R3.10 If yes, what is the *name* of the certificate you are going to acquire and by which *institution* is it provided?**

**R3.11 Do you have any *comments* concerning the certification of your repository?**

***Organizational Aspects***

O1 *policies*

**O1.1 Are there *national policies* on digital preservation in your country?**

o yes o no

**URL (if available):**

**O1.2 Is there a *deposit law* for digital publications in your country?**

o yes, for offline publications only (floppy disks, CD-ROMs, DVDs etc.)
o yes, for both
yes, for online publications only
no

**URL (if available):**

**O1.3 Do you have any *comments* to the *current status* of legal policies concerning digital preservation in your country?**

**O1.4 Does your institution have a *mission statement*?**

o yes, web accessible o o no

**URL (if available):**

**O1.5 Does your repository have a *mission statement*?**
o yes, web accessible o yes, not web accessible o no
**URL (if available):** _____

**O1.6 Do you provide a written statement about your repository's *designated community*?**
o yes, web accessible o yes, not web accessible o no
**URL (if available):** _____

**O1.7 Do you provide a written policy for the *ingest* process to your repository?**
o yes, web accessible o yes, not web accessible o no
**URL (if available):** _____

**O1.8 Do you provide a written policy for the *storage* process within your repository, including strategies for regular backup and disaster preparedness?**
o yes, web accessible o yes, not web accessible o no
**URL (if available):** _____

**O1.9 Do you provide a written policy for the *long-term preservation* process within your repository, including strategies as migration or emulation?**
o yes, web accessible o yes, not web accessible o no
**URL (if available):** _____

**O1.10 Do you provide a written policy concerning *access* to your repository, including access rights statements for one or more user groups?**
o yes, web accessible o yes, not web accessible o no
**URL (if available):** _____

O2 *rights*
**O2.1 Do you provide written *submission agreements* for content producers?**
o yes o no
**URL (if available):** _____

**O2.2 If yes, which kinds of submission agreements do you use (check all that apply)?**
o individual agreement for every object of a certain producer
o collective agreement for all objects of a certain producer
o collective agreement for all objects of a certain group of producers

O3 *preservation planning*
**O3.1 Are your activities based on regular workplans?**
o yes, for every _____ month(s) o no

**O3.2 Do you regularly review your *written policies* to include current best practices?**
o yes, appr. every _____ month(s) o no

**O3.3 Do you regularly audit your *processes* to assure their quality?**
o yes, appr. every _____ month(s) o no

O4 *staff*
**O4.1 Please, give a brief description of how your digital preservation team is *embedded* into the structure of your institution:**

| |
|---|
| |

**O4.2 How many *staff members* does your digital preservation team consist of?**
for organizational purposes _____ for technical purposes _____ for other purposes _____

**O4.3 Did you *outsource* any tasks necessary for the running of your repository?**
o yes o no

**O4.4 If yes, which tasks did you outsource?**

O5 *knowledge*

**O5.1 What is the *highest level* of knowledge concerning digital preservation available within your institution?**

o expert
o intermediate
o novice
o none

**O5.2 Which methods does your institution use or plan to use to *increase* the level of staff expertise concerning digital preservation?**

o regular technology watch
o training by commercial institutions
o training by software vendors
o international workshops
o internal studies by institution members
o external studies by research institutions
o hire consultants
o hire staff with expertise
other: I: international projects; cooperation.

**O5.3 Does your institution *cooperate* with other institutions in the field of digital preservation, for example by carrying out common projects?**

o yes
o no

**O5.4 If yes, with which institutions and in which digital preservation projects does your institution work at the moment?**

**O5.5 Does your institution provide trainings for *data producers* about the *ingest* process of your repository?**

o yes
o no

**O5.6 Does your institution provide trainings for *users* about the *access* process of your repository?**

o yes
o no

O6 *funding*

**O6.1 How is your repository *funded* (check all that apply)?**

o part of institutions operational budget
o fee for services
o other:
o grant funded internal to institution
o grant funded external to institution

**O6.2 Is your repository provided with a *special budget* for longtime preservation?**

o yes
o no

**O6.3 Do you estimate that the *financial resources* of your repository will be sufficient to guarantee the preservation of digital objects over longer periods of time?**

o yes
o no

O7 *information*

**O7.1 Which kinds of *representation information* do you store about the data objects in your repository?**

o hardware descriptions
o software descriptions
o other: o
o operating system descriptions
o descriptions of standards, data types or mapping rules

**O7.2 Which kinds of *reference information* do you store in your repository to uniquely identify the content information?**

o system-generated internal identifier
o other:
o persistent identifier

**O7.3 If you use or plan to use *persistent identifiers*: Which systems do you or will you apply?**

| | | | |
|---|---|---|---|
| o | URN based on NBN (national bibliography number) | o | DOI |
| o | URN | o | PURL |
| o | other: | | |

**O7.4 If you use or plan to use *persistent identifiers*: In which way are the persistent identifiers assigned?**

| | | | |
|---|---|---|---|
| o | by institution itself | o | centralized, by national or international organization |
| o | other: | | |

**O7.5 Which kinds of *context information* do you store in your repository to describe the relationships between content information objects?**

| | | | |
|---|---|---|---|
| o | info about subject-based collections | o | info about versions of the same content in alternative formats |
| o | info about parent-child relations | o | pointers to related content |
| o | other: | | |

**O7.6 Which kinds of *provenance information* do you store in your repository to document the history of the content information?**

| | | | |
|---|---|---|---|
| o | creation/change history | o | refreshment/migration history |
| o | other: | | |

**O7.7 Which kinds of *fixity information* do you store in your repository to validate the authenticity and integrity of the content information?**

| | | | |
|---|---|---|---|
| o | checksums | o | digital signatures/watermarks |
| o | other: | | |

**O7.8 If you are using or plan to use elements from one or more of the following published *metadata schemes*, which schemes are you using?**

| | | | |
|---|---|---|---|
| o | AUDIOMD | o | MPEG7 |
| o | CEDARS | o | MPEG21 (B: planned.) |
| o | Creative Commons Metadata | o | NEDLIB |
| o | Dublin Core 1.1 | o | National Library of Australia |
| o | Dublin Core qualified | o | National Library of New Zealand |
| o | MAB | o | OCLC Digital Archive Metadata |
| o | MARC21 | o | TEXTMD |
| o | METS | o | UniMARC |
| o | METSRights.xsd | o | VERS |
| o | MIX or Z39.87 | o | VIDEOMD |
| o | other: | | |

***Technical Aspects***

T1 *ingest*

**T1.1 Could you briefly describe the *most important workflow* within the ingest process of your repository?**

**T1.2 Which possibilities do you offer for the *receipt* of content?**

| | | | |
|---|---|---|---|
| o | e-mail with attachment | o | ftp |
| o | web upload | o | OAI harvesting |
| o | other: | | |

**T1.3 Which *submission formats* does your repository support?**

| | | | |
|---|---|---|---|
| o | ASCII (flat text) | o | PDF |
| o | SGML | o | Postscript |
| o | XML | o | LaTeX |
| o | HTML | o | MS Word |
| o | TIFF images | o | WAV audio files |

| | | | |
|---|---|---|---|
| o | GIF images | o | MP3 audio files |
| o | JPEG images | o | Real Audio files |
| o | MPEG movies | o | Real Video files |
| o | AVI movies | o | Quicktime movies |
| o | ZIP archives | o | TAR archives |
| o | ISO disk images | | |
| o | other: _____ | | |

**T1.4 Which *integrity check mechanisms* do you use after the receipt of content?**

| | | | |
|---|---|---|---|
| o | checksums | o | virus check |
| o | data format check (JHOVE etc.) | | |
| o | other: _____ | | |

**T1.5 Which *authenticity check mechanisms* do you use after the receipt of content?**

| | | | |
|---|---|---|---|
| o | digital signatures | o | digital watermarks |
| o | other: _____ | | |

**T1.6 Do you *keep log files* of the data flows that emerge from the receipt of content?**

| | | | |
|---|---|---|---|
| o | yes | o | no |

T2 *archival storage*

**T2.1 Could you briefly describe the *most important workflow* within the storage process of your repository?**

| |
|---|
| |

**T2.2 On which *physical location* do you store your data?**

| | | | |
|---|---|---|---|
| o | in house | o | with a third-party provider |
| o | other: _____ | | |

**T2.3 Is your data additionally stored on another physical location for reasons of *disaster preparedness*?**

| | | | |
|---|---|---|---|
| o | yes | o | no |

**T2.4 Which *medium* do you use to physically store your data?**

| | | | |
|---|---|---|---|
| o | hard disk | o | open reel magnetic tape |
| o | optical disk (CD-ROM, DVD etc.) | o | cassette or cartridge |
| o | other: _____ | | |

**T2.5 Which *operating system* do you use on your repository server?**

| | | | |
|---|---|---|---|
| o | Linux (open source) | o | MS Windows NT |
| o | Unix (commercial) | o | MS Windows 2000 |
| o | MacOS | o | MS Windows XP |
| o | other: _____ | | |

**T2.6 Which *data base management system* do you use on your repository server?**

| | | | |
|---|---|---|---|
| o | ORACLE | o | Informix |
| o | mySQL | o | IBM DB2 |
| o | MS SQLServer | o | XML-based system |
| o | other: _____ | | |

**T2.7 Which kind of *repository software* do you use on your server?**

| | | | |
|---|---|---|---|
| o | commercial | o | open source |
| o | self developed | | |

**T2.8 What is the *name and version* of the repository software that you use?**

| |
|---|
| |

**T2.9 Which *storage formats* do you allow in your repository?**

| | | | |
|---|---|---|---|
| o | ASCII (flat text) | o | PDF |
| o | SGML | o | Postscript |
| o | XML | o | LaTeX |
| o | HTML | o | MS Word |
| o | TIFF images | o | WAV audio files |
| o | GIF images | o | MP3 audio files |
| o | JPEG images | o | Real Audio files |
| o | MPEG movies | o | Real Video files |
| o | AVI movies | o | Quicktime movies |
| o | ZIP archives | o | TAR archives |
| o | ISO disk images | | |
| o | other: _____ | | |

**T2.10 How many *database records* does your repository appr. contain now?** _____

**T2.11 Please estimate, how many database records your repository will cover in 5 years:** _____

**T2.12 How much *storage* does your repository appr. cover now (in MB, GB or TB)?** _____

**T2.13 Please estimate, how much storage your repository will cover in 5 years (in MB, GB or TB):**

**T2.14 What is the size of the *biggest single file* within your repository?** _____

**T2.15 How often do you backup the data stored in your repository?** every _____ days

**T2.16 How often do you perform a data integrity check in your repository?** every _____ days

**T2.17 Which *functions* of the OAIS functional model does your *repository software*, spontaneously answered, cover?**

| | | | |
|---|---|---|---|
| o | ingest | o | administration |
| o | archival storage | o | data management |
| O | access | o | preservation planning |

**T2.18 Does your repository software support automated mechanisms for *preservation planning* purposes (like dates of expiration for object versions)?**

| | | | |
|---|---|---|---|
| o | yes | o | no |

**T2.19 If yes, could you briefly *describe* these mechanisms?**

| |
|---|
| |

**T2.20 Do you store *statistical data* concerning the data flows in your repository?**

| | | | |
|---|---|---|---|
| o | yes | o | no |

T3 *access*

**T3.1 Could you briefly describe the *most important workflow* within the access process of your repository?**

| |
|---|
| |

**T3.2 Which *user interfaces* do you offer for the access to your repository?**

| | | | |
|---|---|---|---|
| o | static web pages | o | ftp |
| o | dynamic web pages | o | OAI harvesting |
| o | other: | | |

**URL (if available):** _____

**T3.3 Which software tools were used to create these user interfaces?**

| | | | |
|---|---|---|---|
| o | Perl | o | Java |
| o | PHP | o | JavaScript |
| o | other: _____ | | |

**T3.4 Do you *restrict access* to your repository to special groups of users?**

o restricted to institution members
o restricted to registered users
o other: _____
o no restrictions

**T3.5 Do you distinguish between several user groups with *different access rights* in your repository?**

o yes
o no

**T3.6 Which *search mechanisms* do you provide to the users your repository?**

o metadata search
o full text search
o other: _____
o use of boolean operators
o use of advanced search criteria

**T3.7 Which *services* do you offer (S) to the users of your repository and which of them are *with costs (C)*?**

o preview data reduced manifestations of objects (thumbnails of images etc.)
o view objects
o print on demand
o other: _____
o download data reduced manifestations of objects (thumbnails of images etc.)
o download objects
o burning of CD-ROMs

**T3.8 Are the services you offer part of an *existing business model* for the access to your repository?**

o yes
o no

**T3.9 Which *dissemination formats* do you provide to the users of your repository?**

o ASCII (flat text)
o SGML
o XML
o HTML
o TIFF images
o GIF images
o JPEG images
o MPEG movies
o AVI movies
o ZIP archives
o ISO disk images
other: _____
o PDF
o Postscript
o LaTeX
o MS Word
o WAV audio files
o MP3 audio files
Real Audio files
Real Video files
Quicktime movies
TAR archives

**T3.10 Do you store *access statistics* for your repository?**

o yes
o no