Astrid Recker

The Preservation of Digital Objects in German Repositories

Three Case Studies

Master's Thesis

Studiengang: Master of Library and Information Science (MALIS)

Fakultät für Informations- und Kommunikationswissenschaften

Fachhochschule Köln vorgelegt am 04.11.2009

Supervisor: Prof. Dr. Achim Oßwald

Revised version: April 3, 2010

URN: urn:nbn:de:bsz:14-qucosa-33364

Table of Contents

Abstract

Taking its cue from the increasing amount of digital content deposited into institutional and subject repositories as well as the open question of repositories' role in long-term preservation, this study presents case studies of three German institutional and subject repositories all of which are in a different stage of establishing a (cooperative) framework for the long-term preservation of their digital collections. Drawing on different sets of criteria for trustworthy repositories, it is investigated which strategies the selected repositories pursue to preserve the digital assets in their collections, and how these strategies are implemented with the help of both human repository staff and the repository software used.

The following repositories are considered: pedocs (Deutsches Institut für Internationale Pädagogische Forschung), JUWEL (Forschungszentrum Jülich), and Qucosa (SLUB Dresden). In that the latter can be regarded as examples for common types of (German) repositories, the results of this study might on the one hand serve as a guideline for repositories that intend, similar to the ones described here, to explore questions of long-term preservation in the near future, or are even taking their first concrete steps in this field. On the other hand, it is hoped that this work can at least give some hints as to the stage and status of long-term preservation in the German repository landscape.

Acknowledgements

The author gratefully acknowledges the help and support she received from those responsible for developing, managing and maintaining the repositories considered in the present work. In particular, I would like to thank Dr. Julia Kreusch and Thomas Oerder (pedocs), Dr. Alexander Wagner (JUWEL), and Dr. Andreas Kluge, all of whom took supported me during this project by answering all my questions and by discussing aspects of long-term preservation relevant to institutional and subject repositories with me.

I would like to thank Steve Hitchcock, Tim Brody, Jessie M.N. Hey, and Leslie Carr for giving their permission to reproduce two illustrations from their article “Digital Preservation Service Provider Models for Institutional Repositories: Towards Distributed Services” (2007). The copyright of these illustrations belongs to Hitchcock et al.

List of Abbreviations

AIP	Archival Information Package
BMBF	Bundesministerium für Bildung und Forschung / Federal Ministry of Education and Research
CPA	European Commission on Preservation and Access
CRL	Center for Research Libraries
DCC	Digital Curation Centre
DCMI	Dublin Core Metadata Initiative
DDC	Dewey Decimal Classification
DFG	Deutsche Forschungsgemeinschaft / German Research Foundation
DIAS	Digital Information Archiving System
DIN	Deutsches Institut für Normung / German Standardization Organization
DINI	Deutsche Initiative für Netzwerkinformation
DIP	Dissemination Information Package
DIPF	Deutsches Institut für Internationale Pädagogische Forschung
DNB	Deutsche Nationalbibliothek / German National Library
DPE	DigitalPreservationEurope
DRAMBORA	Digital Repository Audit Method Based on Risk Assessment
DRIVER	Digital Repository Infrastructure Vision for European Research
DRM	Digital Rights Management
ISO	International Organization for Standardization
JHOVE	JSTOR/Harvard Object Validation Environment
JISC	Joint Information Systems Committee
JUWEL	JUelicher Wissenschaftliche Elektronische Literatur
KB	Koninklijke Bibliotheek / National Library of the Netherlands
koLibRI	kopal Library for Retrieval and Ingest
kopal	Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen / Co-operative Development of a Long-Term Digital Information Archive
LMER	Long-term preservation Metadata for Electronic Resources
METS	Metadata Encoding and Transmission Standard
NARA	National Archives and Records Administration
nestor	Network of Expertise in Long-Term Storage and Availability of Digital Resources
OAIS	Open Archival Information System
OCLC	Online Computer Library Center
OCR	Optical Character Recognition
PLANETS	Preservation and Long-term Access through Networked Services
PLATTER	Planning Tool for Trusted Electronic Repositories
Qucosa	Quality Content of Saxony
RLG	Research Libraries Group
ROAR	Registry of Open Access Repositories
RVK	Regensburger Verbund-Klassifikation
SIP	Submission Information Package
SLUB Dresden	Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden
SSOAR	Social Science Open Access Repository
TRAC	Trustworthy Repositories Audit & Certification: Criteria and Checklist
UOF	Universal Object Format
URN	Uniform Resource Name

1. Introduction

> "[The institutional repository] is like a roach motel. Data goes in, but it doesn't come out." (Salo 2007)
>
> "Library silos aren't much better than publisher silos" (Geoffrey Bilder)[1]

While above quotations are meant to take issue with the lacking acceptance and use of repositories – institutional ones in particular –, one might also read them differently: namely, as a comment on the threat posed to the digital assets stored in these repositories by the deterioration of storage media, file format obsolescence, or loss of interpretability and understandability due to insufficient metadata. Picturing repositories as closed-up, inaccessible, and even deadly spaces[2], Salo and Bilder (unintentionally) draw an image of repositories as data cemeteries contributing to the imminent danger of what is frequently referred to as a "digital dark age." This looming threat of "whole portions of the scholarly and cultural record [...] on the brink of disappearing" (Lavoie and Dempsey 2004, no pag.) is primarily attributable to the vulnerability of digital materials, which

> generally do not afford the luxury of procrastination. The fragility of digital storage media, combined with a high degree of technology dependence, considerably shortens the 'grace period' during which preservation decisions can be deferred. Issues of long-term persistence can arise as soon as the time digital materials are created: for example, in choosing between a widely-used, stable digital format, and one that is obscure or on the verge of obsolescence. (ibid.)

The awareness of this danger to our digital heritage has, over the past decade, led to the "realization that perpetuating digital materials over the long term involves the observance of careful digital asset management practices diffused throughout the information lifecycle" (ibid.). [3]

As will be discussed in more detail below, such practices are primarily concerned with the "accurate rendering of authenticated content over time" (ALCTS 2007, 2) or, as stated in the OAIS reference model, "[t]he act of maintaining information, in a correct and Independently Understandable form, over the Long Term."[4]

---

1 See http://sspnet.org/News/Gems_from_the_SSP_30th_Annual_Me/news.aspx – 29.10.2009.

2 Thus the "roach motel" is an insect trap for cockroaches which used to be advertised with the slogan "Roaches check in, but they don't check out." Ironically, according to the Wikipedia disambiguation page, the term "roach motel" is also "used to refer to a proprietary file standard – 'you can check your data in, but you can't check it out'" (http://en.wikipedia.org/wiki/Roach_motel – 30.10.2009).

3 As this quotation already suggests, digital preservation is not conceived of as a single act or event but rather as "a set of activities required to make sure digital objects can be located, rendered, used and understood in the future" (http://www.digitalpreservationeurope.eu/what-is-digital-preservation/ – 30.10.2009). This understanding of digital preservation as "an ongoing activity" which becomes increasingly "difficult to distinguish […] from the routine, day-to-day management of digital materials" (Lavoie and Dempsey 2004, no pag.) has led to the coining of a second, related term: digital curation. According to the DCC definition, the latter term is used to describe "the actions needed to maintain digital research data and other digital materials over their entire life-cycle and over time for current and future generations of users. Implicit in this definition are the processes of digital archiving and preservation but it also includes all the processes needed for good data creation and management, and the capacity to add value to data to generate new sources of information and knowledge" (http://www.dcc.ac.uk/about/what/ – 30.10.2009). The concept is visualized in the DCC Curation Lifecycle Model (http://www.dcc.ac.uk/docs/publications/DCCLifecycle.pdf – 30.10.2009). While the focus of digital curation is therefore possibly somewhat wider than of digital preservation, the distinction between the two concepts is not clear-cut.

4 Consultative Committee for Space Data Systems 2002, 1-11 (hereafter cited as OAIS 2002). It is a truism that no general consensus exists as to how long "long term" actually is where digital materials are concerned. This is aptly illustrated by Jeff Rothenberg's ironic stance that "digital information lasts forever

Both institutional and disciplinary or subject repositories[5] are playing an increasingly important part in scholarly communication worldwide. Thus, while they used to (and partly continue to) be faced with lacking interest and use[6], these problems have abated as a result of the growing importance and acceptance of the Open Access movement, and with it the Green Road to Open Access in particular.[7] This acceptance brings institutional and subject repositories increasingly to the attention of scholars across disciplines as a veritable expansion of and alternative to the traditional publishing system. In consequence, an important and ever-increasing part of our digital cultural heritage is stored in institutional and subject-based repositories. Therefore the question of how these assets will be preserved for the future and what the role and responsibility of repositories is in this context, becomes more and more pressing, and it seems inevitable that repository managers look into the complex issue of digital preservation and draw conclusions as to how to position themselves and their repositories with regard to it.

However, while awareness of these problems has grown over the past years, in particular due to initiatives and networks such as nestor (Network of Expertise in Long-Term Storage and Availability of Digital Resources) in Germany or Digital Preservation Europe (DPE)[8] in the European context, it seems that for institutional and subject repositories (as much as for many other cultural heritage institutions), the consideration of long-term preservation often comes only as an afterthought and is hence not regarded as a genuine, "traditional" task of repositories. This is for example suggested by Barbara Siermann in her contribution to the study A DRIVER's Guide to European Repositories (Weenink, Waaijers, and van Godtsenhoven 2008), in which she points out that

> consumers of the repository trust they will be able to have access to the repository over the years. This demand of the public requires that the repositories start to think about the measures to be taken to keep these repositories accessible for a long time. Nowadays the main focus of

– or five years, whichever comes first" (1999, 2). The OAIS definition of "long term" as a time span "long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community," possibly "extend[ing] indefinitely" (OAIS 2002, 1-1) is widely accepted, but hardly more concrete. Nonetheless at least some consensus seems to exist that "long term" begins with approximately 30 years and more (see also Rusbridge 2009a).

5 According to Clifford Lynch's comprehensive and by now seminal definition, "a university-based institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these materials, including long-term preservation where appropriate, as well as organization and access or distribution […]. At any given point in time, an institutional repository will be supported by a set of information technologies, but a key part of the services that comprise an institutional repository is the management of technological changes, and the migration of digital content from one set of technologies to the next as part of the organizational commitment to providing repository services. An institutional repository is not simply a fixed set of software and hardware" (2003; no pag.). In essence, this definition also applies to disciplinary or subject(-based) repositories, except that these provide services to a community of scholars not defined by institutional membership/boundaries but by their working in the same discipline. In addition, as Kingsley points out, an important distinction between the two types of repositories is that in institutional repositories "the policies on the selection and retention of material, as well as the general scope and organization of the repository, is determined by the institution. This stands in contrast to the discipline- or subject-based repository where depositing policies are determined by the research communities" (2008, 3).

6 It seems that institutional and subject repositories differ in their acceptance in their respective target communities (see, for example, Herb, Kersting, and Leidinger 2008; see Kingsley 2008 for similar observations for the international (or rather anglophone) repository landscape; Salo 2007, while not addressing subject repositories, gives a detailed account of the various reasons why researchers are so hesitant in depositing content in institutional repositories.

7 See Harnad et al. 2004 for the distinction between the "Green" and "Golden Road" to Open Access.

8 http://www.langzeitarchivierung.de, http://www.digitalpreservationeurope.eu – 03.11.2009.

> the institutional repositories is often on collecting material, storing it into the repositories, and making it accessible for a wide community of interested people. Digital preservation itself is often not yet part of the daily workflow. Nor is it clear why and how to perform digital preservation and who should take care of this process. (Siermann 2008, 155)

This is also suggested by Julie Allinson, who observes that "[m]any repositories [...] would not cite preservation as their primary function and might not immediately see the relevance of OAIS to them […]," giving "the relative infancy of digital curation, the 'unknown' aspect of attempting to identify the threats and losses that may occur into the future and the perceived dislocation between preservation activity and the more-pressing need to populate repositories with content" as possible reasons for this lacking interest in long-term preservation (2006, 4-5; see also Knight and Hedges 2007, 66). Moreover, as Hichcock et al. remark,

> [t]aking the age profile of most repositories into account, the need for preservation is perhaps less critical than for older digital content sources, but other factors such as growth and diversity point towards a more urgent need to plan for preservation by the more content-rich repositories. (2007; no pag.)

While from the perspective of long-term preservation these latter factors clearly need to be taken into account, this might not be so apparent to repository managers who are not experts in questions of long-term preservation.

The impression that repositories do not generally see themselves as responsible for the long-term preservation of digital materials also seems to be confirmed by Chris Rusbridge of the Digital Curation Centre (DCC), who in two blog posts recounts an Ideascale[9] discussion and three small polls among UK repository managers concerning the question of whether repositories are responsible for the long or at least medium term preservation of their assets. In this discussion, the statement "The repository should be a full OAIS preservation system"[10] was voted down with -13 votes (net result). In contrast, the revised idea "Repository should aspire to make contents accessible and usable over the medium term"[11] turned out to be more successful and gained +12 votes (net result) (see Rusbridge 2009a). While, as Rusbridge points out himself, these results are far from representative (see 2009b), they do indicate that not all repository managers see the long-term preservation of their assets as their (primary) responsibility.[12] That this might be the case is also suggested by my own (hardly more representative) experience in attempting to find a German repository willing to become the subject of this thesis. Thus, often repository managers cited long-term preservation as a potential future perspective but had

9 http://www.ideascale.com/ – 01.11.2009.
10 http://jiscrepository.ideascale.com/akira/dtd/2276-784 – 29.10.2009.
11 http://jiscrepository.ideascale.com/akira/dtd/2643-784 – 29.10.2009. The explanatory text to this Idea defines "medium term" as a time span of about 10 years, and "long term" as 30 years and more.
12 Although perhaps the comment left by Chris Keene goes to the heart of the problem when he states that the question whether long-term preservation is important is "one of those 'would you like more money?' questions" with only one answer to them (see Rusbridge 2009a). This implies that if repository managers had ample (financial and human) resources, they would embrace long-term preservation more wholeheartedly.

not considered this perspective in any detail yet as they were faced with other, more urgent problems in establishing, advocating, and maintaining the repository.[13]

However, as will become apparent throughout this study, many of the minimum requirements that have to be met in order to stand a chance of preserving digital material for the long term, such as persistent and unambiguous identification or protective measures to maintain the authenticity and integrity of objects stored in an archive, are also minimum requirements that have to be met if the publications stored in a repository are to be usable (i.e. citable) for scholarly purposes. Only if scholars can be certain that the repository documents or data sets they cite will remain unaltered and accessible, and only if they can be certain about the source of this material, open access publications will continue to gain currency and will become a trusted and accepted form of scholarly communication. It seems that with regard to repositories, one task of both long-term preservation initiatives and projects and repository network initiatives, some of which are introduced below, is to make manifest and visible the overlap that exists between what might be called "good repository practice" for scholarly purposes and efforts to keep digital materials accessible over the medium and long term.

The hesitance of repository managers to tackle the question of long-term preservation might, as is also suggested by Allinson, be at least partly a result of the fact that the preservation of digital materials is posing a challenge to which so far no standard solution has been found. Moreover, given the continuous changes in the overall digital landscape, including technologies and infrastructures, an ever-growing number of digital objects and new types of digital objects among other things, it is also highly unlikely that there ever will be such a standard, "one size fits all" solution. Thus although a considerable number of initiatives and projects exists which contribute to the solution of this complex of problems, we are still far away from routine and large-scale solutions adequate to the sheer mass of digital content produced and disseminated every day.

Nonetheless, an increasingly networked and standardized German and European repository infrastructure is being built in projects and initiatives such as, for example, DRIVER (Digital Repository Infrastructure Vision for European Research), the Deutsche Initiative für Netzwerkinformation (DINI) and its Open Access Network[14], many of which are also aiming to establish and maintain interfaces to national and international digital preservation initiatives and communities such as nestor, DPE, Preservation and Long-Term Access Through Networked Services (PLANETS), the British Digital Curation Centre (DCC) as well as JISC-funded projects in the field of digital preservation and curation.[15] Thus, both nationally and internationally, possibilities exist for repositories to participate in

13 Note, however, the DRIVER study on the European Repository Landscape for which repositories were asked among others: "Is the long-term availability of the materials in the repositories secured?" (van der Graaf and van Eijndhoven 2008, 41). In answer, nearly 73% (n=114) of the repositories claimed that this was the case, either by "internal procedures" or due to a cooperation with a national library (see ibid., 42).

14 http://www.driver-community.eu/, http://www.dini.de, http://www.dini.de/projekte/oa-netzwerk/ – 03.11.2009. See Müller et al. (2009) for an introduction to the Open Access Network.

15 http://www.planets-project.eu/, http://www.dcc.ac.uk/, http://www.jisc.ac.uk/whatwedo/topics/digitalpreservation.aspx – 03.11.2009.

and benefit from the development and implementation of preservation strategies and activities.

It is against the backdrop of the foregoing observations – the increasing amount of digital content deposited into repositories and the open question of repositories' role in long-term preservation in particular – that this work will present case studies of three German institutional and subject repositories with the objective of showing how these position themselves with regard to long-term preservation. It will be considered in detail which strategies these repositories pursue to preserve the digital assets in their collections, and how these strategies are implemented with the help of both human repository staff and the repository software used. In that the selected repositories can be considered as typical examples for common types or classes of (German) repositories, the results of this study might on the one hand serve as a guideline for repositories that are, similar to the ones described here, considering to explore questions of long-term preservation in the near future, or are even taking their first concrete steps in this field. On the other hand, it is hoped that this study can at least give some hints as to the stage and status of long-term preservation in the German repository landscape. Thus, this study investigates a subject-based repository (pedocs, Deutsches Institut für Internationale Pädagogische Forschung, DIPF), an institutional repository managed at the central library of a research institution with a multi-disciplinary focus in the STM field (JUWEL, Forschungszentrum Jülich), and a university-based institutional repository which, however, also offers services for other institutions (HE and other) in the federal state of Saxony (Qucosa, SLUB Dresden).[16]

While initially this study set out to look at how German repositories *do* long-term preservation, it turned out quickly that only very few – if any – repositories have already implemented concrete long-term preservation strategies and measures. In consequence, the original focus of this study was slightly shifted. It now centers on three repositories in various stages of establishing a framework for long-term preservation of their digital collections, and particularly looks at tasks and functions relevant to long-term preservation that these repositories are already fulfilling or will have to fulfill in the future in order to preserve the digital objects published through them. All of these repositories have opted (or are planning to opt) for a cooperative solution to long-term preservation involving the repository and a preservation service provider (see below). They thus do not aim at shouldering the preservation of their assets all by themselves but in cooperation with a long-term archive which will be responsible for a considerable part of the preservation activity. Thus, pedocs is already in the process of working out concrete steps for a cooperative preservation workflow with the German National Library (DNB), while Qucosa

16 The attempt was also made to focus on repositories using different software. However, due to the above-mentioned difficulties in finding repositories willing to become the subject of this study, only two different kinds of software – DSpace and OPUS – are now represented. This, however, makes for an interesting match as OPUS is the software most often used by German repositories while DSpace is the most widely distributed software internationally.

is at the beginning of a DFG-funded project during which a disaggregated preservation model (cf. Knight and Hedges 2007) will be explored and implemented. JUWEL is currently planning to establish a cooperative solution in the future without having taken concrete steps in establishing such a cooperation as of yet; however, it is actively working to ensure that the materials it collects will be in a suitable state for submission to a long-term archive. However, regardless of the fact that the repositories discussed in this study do not aim at becoming long-term archives themselves, it will become apparent in the following that a considerable number of important tasks and responsibilities lie with them, and that the fulfillment of these tasks is a crucial factor in the success of the cooperative preservation solution.

1.1 Criteria Catalogs for Trustworthy Digital Repositories[17]

The framework for the description, analysis, and evaluation of the activities of the three selected repositories relevant to the preservation of their digital assets presented in this study is provided by three criteria catalogs, which will be briefly introduced in the following. Although archives, libraries, museums, and other cultural heritage organizations have acquired and built expertise in preserving objects from the analog world successfully over centuries, and can thus be said to have earned our trust in this domain, such long-standing expertise is still missing where digital objects are concerned. Thus, as the authors of the RLG-OCLC Report on Trusted Digital Repositories claim, on one hand

> [l]ibraries, archives, and museums are entrusted with the materials and objects that document our cultural heritage. They are trusted to store these valuable materials. They are trusted to provide access to them in order to document and reveal history […]. They are trusted to preserve these items to the best of their ability for future generation. (RLG-OCLC 2002, 8)

Yet this trustworthiness cannot automatically extend to the digital domain for the reasons outlined above, and hence a need exists for cultural heritage institutions to prove that they can be trusted to preserve and provide access to digital objects over the long term as much as with analog ones. This need was beginning to be perceived and addressed as early as the mid-nineties, when the question of sustainable digital archives was first raised with some force.

Beginning with the work of the Commission on Preservation & Access (CPA) and the RLG Task Force on Archiving of Digital Information between 1994 and 1996, a number of attempts were made to "articulat[e] the nature of a sustainable digital archives [sic],"[18] quickly moving in the direction of repository certification. Among the results of these distributed but often networked efforts were the Reference Model for an Open Archive Information System (OAIS 2002), the RLG-OCLC Report on Trusted Digital Repositories:

17 Please note that in this context, the term "repository" is ambiguous. Thus, repositories as considered by the nestor and the TRAC catalog of criteria, for example, are digital long-term archives. A sub-group of such digital archives can be subject or institutional repositories; as outlined above, however, often these are not capable of or willing to offer long-term preservation services and thus represent digital short- or medium-term archives, where "archive" is taken to mean "a collection of information."

18 http://www.oclc.org/research/activities/past/rlg/trustedrep/default.htm – 03.11.2009.

Attributes and Responsibilities (2002), and the Audit Checklist for the Certification of Trusted Digital Repositories developed by RLG-NARA Digital Repository Certification Task Force (founded in 2003). After some revisions the latter resulted in Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC) published by the Center for Research Libraries (CRL) and OCLC in 2007. In the European context, similar initiatives and task forces were established by Digital Preservation Europe, nestor, and the Digital Curation Centre, resulting in tools and documents such as DCC's Data Audit Framework and the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA co-developed with DPE, 2007), nestor's Catalog of Criteria for Trusted Digital Repositories (version 1 was published in 2006, version 2 – currently available only in German – in 2008[19]) as well as DPE's Planning Tool for Trusted Electronic Repositories (PLATTER).[20]

As a result of these efforts, a number of definitions of what it means for a repository or archive to be trustworthy exist, often emphasizing different aspects, ranging from the IT-concepts of integrity and authenticity, over administrative and procedural accountability, to financial and organizational sustainability (cf. RLG-OCLC 2002, 13). While the 1996 CPA/RLG report highlights, for example, that "[f]or assuring the longevity of information, perhaps the most important role in the operation of a digital archive is managing the identity, integrity and quality of the archives itself [...]" (23; also qtd. in RLG-OCLC 2002, 8), the nestor catalog mentions IT-security as one core element of trustworthiness when defining the latter as follows:

> Vertrauenswürdigkeit […] wird als Eigenschaft eines Systems angesehen, gemäß seinen Zielen und Spezifikationen zu operieren (d.h. es tut genau das, was es zu tun vorgibt). Aus Sicht der IT-Sicherheit stellen Integrität, Authentizität, Vertraulichkeit und Verfügbarkeit Grundwerte dar. IT-Sicherheit ist somit ein wichtiger Baustein für vertrauenswürdige digitale Langzeitarchive. (2008, 5)

Suggesting a somewhat more comprehensive – and more detailed – definition of trustworthiness, the authors of TRAC remind us that

> [i]n determining trustworthiness, one must look at the entire system in which the digital information is managed, including the organization running the repository: its governance; organizational structure and staffing; policies and procedures; financial fitness and sustainability; the contracts, licenses, and liabilities under which it must operate; and trusted inheritors of data, as applicable. Additionally, the digital object management practices, technological infrastructure, and data security in place must be reasonable and adequate to fulfill the mission and commitments of the repository.[21]

But although these definitions – each taken from the introductory section of the cited documents – highlight different aspects, a closer look at the actual requirements on and responsibilities of trustworthy digital repositories outlined by the respective authors shows

19 Please note that hereafter the German catalog will be used and quoted as it is the most current version.

20 According to the DPE website, PLATTER "provides a basis for a digital repository to plan the development of its goals, objectives and performance targets over the course of its lifetime in a manner which will contribute to the repository establishing trusted status amongst its stakeholders. PLATTER is not in itself an audit or certification tool but is rather designed to complement existing audit and certification tools by providing a framework which will allow new repositories to incorporate the goal of achieving trust into their planning from an early stage" (http://www.digitalpreservationeurope.eu/platter/ – 30.10.2009).

21 CRL and OCLC 2007, 3. Hereafter cited as TRAC 2007 and page or criterion number.

that high-level consensus exists about important elements without which trustworthiness cannot be achieved. The latter fact is particularly mirrored in the existence of a set of Core Requirements for Digital Archives, devised by representatives of DCC, DPE, nestor, and the CRL in 2007, and intended "to guide further international efforts on auditing and certifying repositories"[22]:

> 1. The repository commits to continuing maintenance of digital objects for identified community/communities.
> 2. Demonstrates organizational fitness (including financial, staffing structure, and processes) to fulfill its commitment.
> 3. Acquires and maintains requisite contractual and legal rights and fulfills responsibilities.
> 4. Has an effective and efficient policy framework.
> 5. Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.
> 6. Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.
> 7. Creates and maintains requisite metadata about actions taken on digital objects during preservation as well as about the relevant production, access support, and usage process contexts before preservation.
> 8. Fulfills requisite dissemination requirements.
> 9. Has a strategic program for preservation planning and action.
> 10. Has technical infrastructure adequate to continuing maintenance and security of its digital objects. (ibid.)

In addition to the nestor and TRAC catalogs, the catalog of criteria used by the DINI certificate 2007 for open access repositories ("DINI-Zertifikat Dokumenten- und Publicationsservice 2007") will be used in this study, as it plays an important role in building and shaping the German repository landscape. Despite the fact that the nestor, TRAC, and DINI criteria have similar goals to some extent, a number of differences exist between them due to their different target groups (institutional and subject repositories in the case of DINI[23], long-term digital archives in the case of nestor and TRAC). Thus, although long-term preservation and availability of digital objects do play a role in the DINI criteria, this is clearly not the central focus of the certificate, which primarily evaluates aspects such as visibility, author support, and the existence of logs and statistics among other things (cf. DINI 2007). According to Dobratz and Schoger,

> [f]or DINI the primary objective of the guidelines and criteria was to improve interoperability and cooperation between German higher education institutions that run digital repositories and to provide an instrument for the repository operators that could be used to raise the visibility, the recognition and the importance of the digital repository within the university. The certificate shows potential users and authors of digital documents that a certain quality level in operating the repository is guaranteed and that this distinguishes it from common web servers of institutions. In addition, DINI sees its certificate as an instrument to support the Open Access concept. (2005, 1)[24]

It follows that although the DINI certificate does not center on long-term preservation, its criteria, similarly as those comprised in the nestor and TRAC catalogs, aim at establishing

22 http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re – 30.10.2009.
23 Although DINI primarily intends to enable repositories to "position themselves as institutional repositories" (2007, 8; my translation), it is also used to certify subject repositories such as SSOAR, the Social Science Open Access Repository (http://www.ssoar.info/en.html – 03.11.2009) and is hence relevant to all three repositories discussed below. Of these three repositories, only Qucosa is already DINI 2007 certified.
24 See also DINI 2007 (8).

the trustworthiness of the repositories seeking certification (i.e. the quality of their services). Among the markers of quality considered by DINI are many that are also highly relevant in the context of digital preservation – a fact that again suggests the existence of considerable overlap between what might be described as "traditional" repository functions and services and functions that need to be fulfilled if the repository seeks to preserve its collections for the long or medium term. Thus, many of the tasks and functions described below should be fulfilled by *any* repository that aims at providing high-quality and sustainable services in the field of scholarly communication, and one might regard the DINI catalog of criteria as a manifestation of this overlap. At the same time, the "gaps" in the DINI catalog, which omits many of the requirements made by nestor and TRAC, also serve as an indication of the boundary that separates more traditional repository "business" from the "business" of long-term preservation.

Possibly due to its specific target group, the requirements formulated by the DINI certificate tend to be much more specific and concrete than the criteria of nestor and TRAC, both of which offer criteria in a far more general form, thus leaving institutions and organizations considerably more freedom how to implement them.[25] In the light of these circumstances this study will regard the DINI criteria (those which are relevant to long-term preservation concerns, that is) as a specification, a spelling-out of the more general catalogs, nestor and TRAC. Thus it should be kept in mind that DINI is only one possible "manifestation" of the general functions and requirements outlined in nestor and TRAC, and it will be considered as such alongside the two .

All three criteria catalogs primarily used in this study have in common that they aim at establishing a framework for the (self- or external) evaluation of repositories. Thus, the authors of TRAC write, "[r]egardless of size or purpose, all repositories should be encouraged to use this checklist as a tool for objective evaluation whether it is accomplished in-house or by an objective, third-party auditor […]" (2007, 5). Similarly, the nestor catalog of criteria seeks to provide cultural heritage organizations both with a guideline for conceptualizing and building a trusted digital repository and with a means for self-evaluation:

> Der vorliegende Kriterienkatalog richtet sich in erster Linie an Gedächtnisorganisationen (Archive, Bibliotheken, Museen) und dient als Leitfaden, um ein vertrauenswürdiges digitales Langzeitarchiv zu konzipieren, zu planen und umzusetzen. Ferner kann er auf allen Stufen der Entwicklung zur Selbstkontrolle eingesetzt werden […]. (2008, 7)

At the same time all three catalogs go one step further in also aiming at establishing an audit and certification procedure for trusted digital repositories. Thus, more than 30

25 For example, nestor and TRAC demand the existence of a policy; DINI voices very concrete requirements on the content of the policy. Similarly, while nestor and TRAC make the use of metadata a requirement, DINI makes Dublin Core the minimum requirement. This practice is both helpful in that it offers repositories a very pragmatic approach at certification, it is, however, also problematic – for example because it requires the DINI Certificate to be revised on a regular basis in order to keep it compliant with technological and policy-related current trends and developments. Repositories which follow either nestor or TRAC on the other hand, will generally have to invest considerably more time in the attempt to spell out what the generic criteria mean with regard to their particular repository. This allows for more flexibility, but requires more expertise.

German repositories have already been awarded the 2004 or 2007 DINI Certificate. The nestor catalog is in the process of being transformed into a DIN standard[26], and TRAC forms the basis of the efforts of the Birds of a Feather (BOF) working group currently creating a document to be submitted to ISO for consideration as an international standard.[27]

This study uses selected criteria from the above-mentioned catalogs to investigate the approaches to long-term preservation taken by the three repositories on which it will focus. The benefits which can be gained from this, even if the repositories to which the catalogs will be applied do not aim at becoming long-term archives themselves, are described in a recent article by Steinhart, Dietrich, and Green in which they summarize their experiences in applying the TRAC criteria to DataStaR, a data-staging repository currently developed at Cornell University. As they explain, they

> decided to investigate and incorporate best practices related to digital preservation to the fullest extent possible even though DataStaR is not intended to serve as a long-term preservation repository […]. There are good reasons for taking this approach. First, […] policies and best practices for repositories seem to be best developed in the digital preservation community, and digital preservation frameworks have much to offer that bears on responsible management of repositories, regardless of a repository's stated preservation commitment [...]. Digital preservation frameworks also emphasize the importance of establishing trust, and how repositories can demonstrate trustworthiness with certain kinds of evidence. (2009, no pag.)

As this quotation indicates, traditional repository tasks and long-term preservation tasks are not as far removed from each other as one might think: thus, trustworthiness is something that both traditional and preservation-centered repositories should strive to achieve. At the same time, the observations of Steinhart et al. serve to "justify" the approach taken in this study, which uses the criteria catalogs introduced above as guidelines as to which aspects need to be considered in particular if a repository wants to be prepared to take at least a shared responsibility for the long-term preservation of its digital assets, that is, without becoming a long-term archive itself.

In order to keep this study within the set limits (both with regard to time and to length), the number of criteria used has been reduced. Not only were some criteria simply not relevant in the current context, but also the decision was made to focus in particular on the actual management of digital objects (implemented procedures carried out both by human staff and software-aided) as well as the areas of policy and technical infrastructure. Finally, the (organizational and disciplinary) context in which each repository is embedded was also considered briefly. In contrast, despite their importance criteria dealing primarily with questions of user access have been largely disregarded for the reasons just mentioned. Instead I have worked based on the assumption (whether this is entirely justified is certainly to be questioned) that a repository which fulfills the non-

26 A first version of the standard was scheduled to be presented in the DIN Standards Committee meeting on October 13, 2009 (cf. the timetable of the nestor working group "Vertrauenswürdige Archive" on http://www.langzeitarchivierung.de – 30.10.2009).

27 See http://wiki.digitalrepositoryauditandcertification.org/bin/view for further details – 30.10.2009.

access related criteria will also have little difficulty in implementing the access functional entity according to the requirements voiced in the catalogs.

In order to make working with the different catalogs easier, in addition the decision was made to map them onto each other based on the functional entities outlined in the OAIS reference model.[28] Taking the nestor catalog of criteria as a basis, criteria of the other two catalogs were matched with the nestor criteria. In a second step, these matched criteria were then reordered according to the OAIS functional entity or entities for which they seemed most relevant (see Appendix A for the mapping of the criteria). This procedure seemed adequate not only because OAIS is the accepted international standard for the (high-level) description of digital archives, but also because in particular nestor and TRAC clearly work against the backdrop of the OAIS reference model by using its terminology and by referring to its functional entities.

1.2 OAIS and Disaggregated Preservation Models for Repositories

> "OAIS is not an architectural model. It is an ontology, a terminology underlying a shared view and, as such, provides a means of communication […]." (Allinson 2006, 11)

In that OAIS is one of the main conceptual frameworks for the following discussion, it needs to be considered briefly to what extent and in which ways the model's terminology is really applicable to institutional and subject repositories. As the commentaries to the aforementioned blog posts by Chris Rusbridge illustrate, whether OAIS is a suitable model for institutional or subject repositories, especially if they do not see long-term preservation as one of their central tasks, continues to be subject to debate. Nonetheless, the present work will use OAIS terminology to describe the elements and activities of the selected repositories, even if the latter do not consider themselves as and are no long-term archives.[29] That this is possible is argued, among others, in Allinson's article "OAIS as a Reference Model for Repositories" (2006), in which she opens her argument by pointing out that

> most repositories, perhaps without realising it, are offering some level of preservation. They are storing and managing materials on behalf of others, they are committed to gathering metadata and they have agreements and policies to ensure a certain level of service. (5)

Arguing that it is "relatively easy" for repositories "to conform to the OAIS model" (2006, 5), she explains that the requirements for OAIS-compliance are merely a "small set of high-level goals, providing a loose framework for best practice and communication between repositories" embodied in "six responsibilities encompass[ing] many of the tasks that institutional repositories are already fulfilling […]" (2006, 6). A similar point is made by Thibodeau, who argues that

28 The present work will refer to the OAIS reference model as published in 2002 by the CCDS (ISO 14721:2003). A new (preliminary) OAIS version was presented in May 2009 for public examination and comment (hereafter cited as OAIS 2009). This version was scheduled to be submitted to ISO in or after June 2009 but is not an ISO standard as of yet.

29 Please note that while some of the terminology will be explained in the course of this study, a general familiarity with the terminology of OAIS on the side of the reader is assumed.

> [t]he Open Archival Information System (OAIS) reference model, which provides an abstract description of the function of any system used to preserve any type of information for any significant period of time, as well as a detailed delineation of the information management required not only to ensure that the information survives but also that it can be accessed and correctly understood in the future, explicitly offers itself at least as a benchmark for evaluating of digital repositories. (2007, no pag.; cf. also Hitchcock et al. 2007)

Both Allinson and Thibodeau thus suggest (even if implicitly in the latter case) that the OAIS model may well apply to (digital and analog) archives which do not aim at storing their assets for the "long term." Moreover, the application of OAIS and its terminology is particularly useful in cases where a cooperative approach to long-term preservation is taken. Thus, Winkler observes:

> Die Aufgaben eines Repositoriums und eines Langzeitarchivs unterscheiden sich aufgrund der spezifischen Systemanforderungen. Beide Systeme können jedoch mit Hilfe des OAIS-Modells beschrieben und ausgestaltet werden. Eine konzeptionelle Verzahnung in einem kooperativen Modell lässt sich weitaus besser realisieren, wenn beide Akteure nach denselben Prinzipien verfasst sind. (2008, 71)

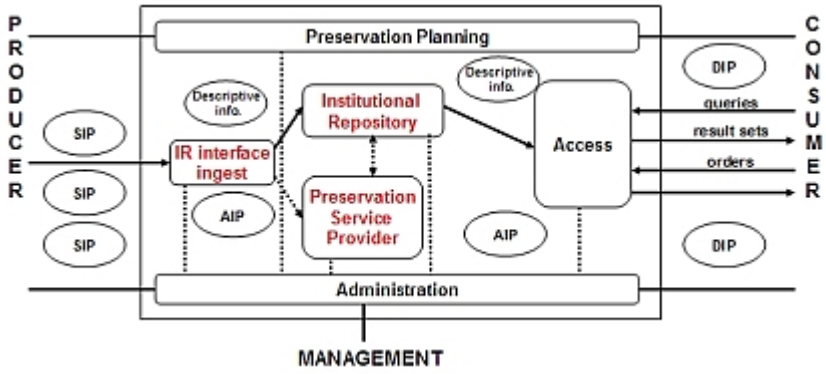
As the repositories selected for this study cooperate with a digital preservation service provider or plan to do so in the future, it seems helpful to describe repository elements and organizational units as well as preservation-related activities and services with the help of the well-defined OAIS terminology, and to model the (planned) cooperation with reference to the OAIS functional entities, as both are the accepted standard.

The possibility of modeling such a cooperative or disaggregated approach to digital preservation according to OAIS was explored as part of the PRESERV project[30], which among others proposed a service provider model (see ill. 1 and 2), in which "[t]he archival storage, or service provider, element in principle covers the full range of preservation services, from bit-level storage to migration and emulation," while other responsibilities either remain with the institutional or subject repository, or are carried out in cooperation (Hitchcock et al. 2007, no pag.). As the illustrations show, repository and preservation service provider therefore have shared but interlocking responsibilities in this model. As observed by Winkler (2008), in such cooperative models, the repository becomes the producer or content provider for the long-term archive (see 65).[31] One consequence of this constellation is that information packages have, as Winkler points out, a "double value" (2008, 69; my translation). Thus, repository Dissemination Information Packages (DIPs) become Submission Information Packages (SIPs) for the preservation service provider, while the DIPs of the latter are re-ingested into the repository as SIPs.[32]

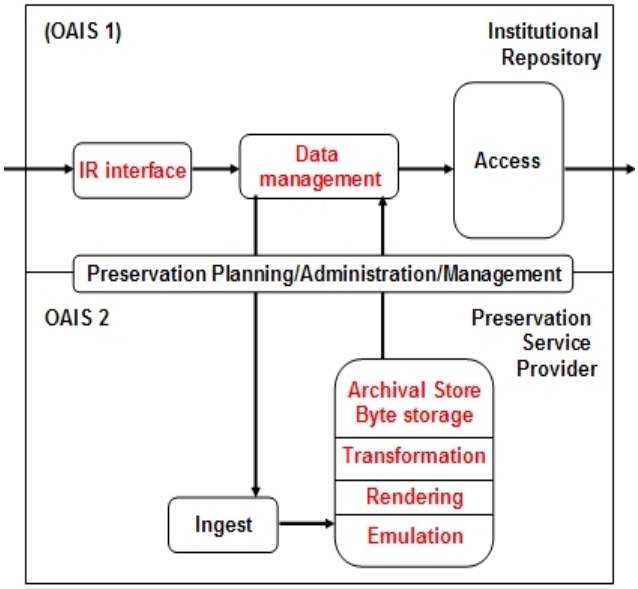
30 http://preserv.eprints.org/ – 03.11.2009.

31 It will not be considered in the following whether the cooperating repositories are partial or full OAIS, as this is not of utmost relevance in the present context. As Hitchcock et al. state, "in the service provider model the IR could be OAIS-compliant, but it need not necessarily be if the service provider delivers that compliance" (2007, no pag.).

32 In the following the OAIS designations for information packages will be used for the packages accepted, managed, stored, and disseminated both by the repositories and the preservation service provider. Thus, similar to the long-term archive, the repository ingests what will be referred to as (pre-)SIPs, stores (pre-)AIPs and disseminates DIPs. This seems possible in particular because, as Allinson remarks, "[f]or institutional repositories, it is conceivable that the SIP, AIP, and DIP are all the same, that a submitted package is ingested, stored and delivered in an unchanged state. There is nothing in OAIS to say that this should not happen, so long as the necessary information is captured at submission and the necessary planning for preservation has been made" (2006, 11).



III. 1 (Hitchcock et al. 2007, no pag.)



III. 2 (Hitchcock et al. 2007, no pag.)

Both illustrations[33] make clear that in the proposed model the OAIS functional entity "Access" remains wholly in the responsibility of the repository. While it is thinkable that access is also provided by the preservation service provider, Winkler points out correctly that the publication/dissemination of documents is one of the central tasks of repositories, which have key competencies in this field:

> Des Weiteren ist die Publikation von Dokumenten gerade die originäre Aufgabe eines digitalen Repositoriums, d.h. es stellt die Infrastruktur für die Recherche und Informationsverteilung in Endnutzersysteme bereits von Haus aus zur Verfügung [...]. Außerdem aggregiert das Publikationssystem Mehrwertdienste, wie vergleichende Nutzungsstatistiken und verknüpft in Zukunft die publizierten Dokumente mit den zitierten Dokumenten und kann somit an ein System zur Durchführung von Zitationsanalysen ankoppeln. (2008, 66)

33 Illustrations reproduced with permission from Steve Hitchcock, Tim Brody, Jessie M.N. Hey, and Leslie Carr. Illustration 1 derives from the OAIS model, which is not reproduced here. © S. Hitchcock, T. Brody, J.M.N. Hey, L. Carr.

Thus it seems highly recommendable to leave the Access functional entity in the responsibility of the repository entirely. We will see in the following, however, that this distribution of responsibilities also means that repository and preservation service provider have to work out a detailed plan governing how communication between them takes place – that is, a set of well-defined rules has to be created which determine, for example, which information packages are exchanged when and how. Among the issues that have to be addressed according to Knight and Hedges are

1. The method of enabling machine-to-machine transfer between two repositories;
2. Maintaining consistent identifiers between the digital repositories;
3. Maintaining authentic records between the digital repositories. (2007, 69)

While the first and second issue can be solved by means of protocols (e.g. OAI-PMH), exchange formats and standards (e.g. UOF or ONIX[34]), and persistent identifiers, the third issue in particular depends on more than merely "technical" considerations. Thus it needs to be agreed upon what happens if, for example, a digital object is migrated to another format by the preservation service provider – if it is to be re-ingested into the repository, at which point and in which form will this be the case? How and to what extent will metadata be updated and expanded as migrated objects are re-ingested into the repository? How can it be assured that the repository and/or repository users have the software needed to display the transformed digital objects at their disposal?[35]

Users accessing digital objects through a repository must be entirely certain that the digital objects they receive are authentic copies of the original objects submitted to the repository – a requirement that any repository will have to meet, regardless of whether it is involved in long-term preservation efforts or not. Thus the repository, just like a preservation service provider, will therefore have the responsibility to protect the integrity and authenticity of the information packages on its servers. In consequence, a repository cooperating with a preservation service provider needs to be able to make sure that on the one hand the SIPs submitted to the long term archive for preservation contain digital objects that are uncorrupted; on the other hand the same needs to be guaranteed for the objects and metadata contained in Dissemination Information Packages accessed by the repository users, as these will not be generated by the long term archive but from the information packages archived by the repository. As all of this makes very clear, even if the repositories considered here are not directly involved in taking concrete preservation action, for example, by converting digital objects to a different format in response to the threat of obsolescence, they nonetheless play an important part in curating these objects over their lifecycle and hence form part of the preservation system and workflow (see the DCC Curation Lifecycle-model for a possible visualization of this lifecycle).

34 See Steinke 2006 for information on the Universal Object Format (UOF) and http://www.editeur.org/8/ONIX/ for information on ONIX – 03.11.2009.
35 On this issue, see, for example, chapter 9.2 of the nestor Handbuch (Neuroth et al. 2009).

2. Approaches to Long-Term Preservation in German Repositories[36]

2.1 Introduction and Overview: pedocs, JUWEL, and Qucosa

2.1.1 pedocs

pedocs[37] is a repository hosted and managed by the DIPF (Deutsches Institut für Internationale Pädagogische Forschung / German Institute for International Educational Research), an institute whose "profile is shaped by two focal areas of activity, i.e. educational information and educational research."[38] Among its tasks the DIPF sees the provision of information services to researchers in educational science with the objective of facilitating an enhanced science, for example, by augmenting "an integrated structure of portals by means of modern information and communication technology" and by "rendering historical stocks from the domain of educational history accessible."[39] Part of these efforts is the Fachportal Pädagogik (German Education Portal[40]), providing access to databases from the field of educational science, among them the bibliographic database FIS Bildung (German Education Index[41]) and pedocs, both managed and maintained by the DIPF.

As an open access repository, pedocs collects and provides access to scholarly publications from the field of educational science and research, and is hence a disciplinary or subject repository. As stated in its policy document ("Leitlinien"/"Guidelines"[42]), pedocs' central concern is both to make scholarly publications from educational science visible and openly accessible and to preserve them for the long term in cooperation with the German National Library (see below). The former objectives – visibility and accessibility (in the sense of "findability") – are primarily achieved by inclusion in (open access) search engines and databases. Thus, publications can be found through the following channels:

- Search options on the German Education Portal and the German Education Server;
- Google indexations, which we actively promote, and other search engines;
- Delivering the data to so-called OAI servers, which maintain nodes for searching scientifically relevant literature (e.g. BASE; OAIster; MeInD);
- Documentation by the German Education Index [...];
- The online catalogue of the German National Library (OPAC).[43]

In the German repository landscape, pedocs is exceptional on the one hand because from the outset it has sought to cooperate with publishers in educational science in order to establish cooperation models which allow the open access publishing of publisher products according to the green road.[44] Second, pedocs has been set up as a service

36 The major work on this thesis was conducted between June and November 2009. All repositories have undergone further development in the meantime, and hence some of the observations made in the following chapters may no longer reflect the present circumstances.
37 http://www.pedocs.de – 03.11.2009
38 http://www.dipf.de/en/institute/organisation – 03.11.2009
39 http://www.dipf.de/en/educational-information/educational-information – 03.11.2009
40 http://www.fachportal-paedagogik.de/start_e.html – 03.11.2009
41 http://www.fachportal-paedagogik.de/fis_bildung/fis_datenbank_e.html – 28.07.2009
42 http://www.pedocs.de/leitlinien_e.html – 03.11.2009. Hereafter cited as pedocs Guidelines.
43 http://www.pedocs.de/publizieren_mit_pedocs_e.html – 01.11.2009.
44 See http://blog.bildungsserver.de/?p=269 for a brief report on the pedocs workshop "Open Access Erziehungswissenschaften" carried out in August 2009 with representatives of publishing houses –

actively pursuing the long-term preservation of its collections. In this respect, it differs from the majority of German repositories, many of which are, as outlined above, set up as platforms focused primarily – and often exclusively – on (present) access. The approach taken by pedocs is different in the sense that from its inception it was planned with an emphasis on the requirements of digital preservation, and thus its processes, policies, and standards have been (and are being) developed to meet the requirements of long-term preservation, which will be carried out in cooperation with the DNB. In consequence, long-term preservation features did not have to be added to pedocs retrospectively, and thus added "on top" of an existing and finished system so to speak, but were implemented right from the beginning.

The types of publications accepted by pedocs include monographs (in particular out-of-print and digitized monographs), conference proceedings, essays, journal articles, pre- and postprints, doctoral or habilitation theses, as well as gray literature. Currently, a great number of the documents archived in pedocs are essays and articles published before 1995 by authors who were able to secure the right to publish these works online before the 2008 change in German copyright law (the so-called "zweiter Korb"). However, the collection of current publications is growing steadily.

pedocs is funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) and is currently preparing to apply for DINI certification.

pedocs Software

The pedocs software is based on a strongly modified version of the German OPUS software (version 3.1)[45] and runs on the common LAMP stack.[46] The decision to use OPUS was made on the one hand because it runs "out of the box." On the other hand – and more importantly – however, it uses the same software components as the Fachportal Pädagogik so that the expertise required to run and maintain OPUS was readily available at the DIPF. In addition, OPUS has a wide user community in Germany, and expectations were at the point when the decision was made that future support and development of the software were likely.[47] Early in the development of pedocs it became apparent, however, that the original OPUS, which was geared towards the storage and management of (doctoral) dissertations and strongly mirrored the institutional structure of German universities (e.g. it contained required fields such as university, institute, etc.), needed to be adapted in order to make it useful for the objectives pursued by pedocs. In

03.11.2009.

45 See http://www.carpet-project.net/en/content/tools-and-services/carpet/opus-1/ for summary information (including an illustration of its architecture) about OPUS. Additional information is available at http://elib.uni-stuttgart.de/opus/doku/dokumentation.php – 03.11.2009. Scholze and Summann 2009 gives a brief introduction to OPUS (note that the publication, although very recent, describes a previous OPUS version). Winkler 2008 also presents information on the conception and architecture of the OPUS software.

46 I.e. Linux, Apache, MySQL, and PHP4.

47 According ROAR, 34 out of 111 German repositories currently use OPUS, while according OpenDOAR among the 137 German repositories listed there, 52 use OPUS. See http://roar.eprints.org/ and http://www.opendoar.org/ – 01.11.2009. See http://www.driver-support.eu/national/germany.html for a short overview of the German repository landscape – 03.11.2009.

consequence, for example, new document types were created, fields were added and renamed. Further changes were and continue to be made in order to adapt the pedocs software to the requirements resulting from long-term preservation concerns.

Expectations are that within two years from now the pedocs software will only be remotely related to the original OPUS base. The downside of the adaptations made to the OPUS software on which pedocs was originally based are that an update to subsequent versions of OPUS will not be possible as any changes made in the past would be lost. In addition, pedocs no longer partakes in the advantages of using a software with a wide (national) user community.

pedocs-DNB cooperation

As mentioned above, pedocs is in the process of implementing a preservation workflow in cooperation with the German National Library (DNB). Currently, an agreement is being drafted which will specify the details of the cooperation, the basis of which is among others formed by the legal deposit regulation (Pflichtablieferungsverordnung, PflAV[48]), the National Library Law[49], as well as the DNB's collection guidelines. Thus, pedocs will be treated similarly to a publisher required to submit publications to the DNB; the DNB itself, on the other hand, is required by law to collect and preserve electronic publications.[50] The cooperation agreement will have to specify, among others, the design of the Submission Information Packages provided by pedocs, how frequently these will be harvested by (or submitted to) the DNB, and at which points and in which form the DNB will return migrated objects back to pedocs for re-ingest.

2.1.2 JUWEL (JUelicher Wissenschaftliche Elektronische Literatur)

JUWEL[51] is a service offered by the Central Library of the Forschungszentrum Jülich (Research Center Jülich), which was established as part of the Open Access strategy proposed in 2003 and since then implemented at the Forschungszentrum (FZ).[52] With key competencies in the field of physics, biophysics, and scientific computing, research at the Forschungszentrum Jülich is also carried out in the areas of health (biotechnology, neuroscience), energy and environment, and information technologies of the future.[53] Thus, the Central Library is a special library providing services to a very heterogeneous user community, which makes both the monitoring of JUWEL's designated communities and the subject indexing of publications submitted to the repository a challenging task.

48 http://bundesrecht.juris.de/pflav/index.html – 01.11.2009.
49 http://bundesrecht.juris.de/dnbg/index.html – 01.11.2009.
50 For the DNB's preservation strategies and principles, see Neuroth et al. 2009 (chapter 18.2) as well as the short introduction to kopal given below.
51 http://juwel.fz-juelich.de:8080/dspace/ – 03.11.2009
52 See http://www.fz-juelich.de/zb/index.php?index=758 (03.11.2009). for information on the FZ's Open Access strategy which is also outlined in a 2003 lectured delivered by the then director of the FZ's central library, Dr. Rafael Ball (see Ball 2003).
53 See http://www.fz-juelich.de/portal/research – 03.11.2009

Like pedocs, the repository is not a “standalone” service but is integrated into a (preexisting) network of services and grown structures, including a database containing bibliographic records for publications by researchers employed at the Forschungszentrum (“Veröffentlichungsdatenbank” [VDB]) and a campus press (electronic and print publications), both hosted at the Central Library.[54]

In line with the Open Access strategy mentioned above, the primary objective in the implementation of JUWEL was to make the results of the research carried out at the Forschungszentrum openly accessible in full text and thus, although one was aware of the importance and relevance of digital preservation for JUWEL and its services, questions and concerns of long-term preservation have not been a priority so far.[55] Thus, although it has always been attempted to also address requirements of long-term preservation in the development of the repository's services, no comprehensive policies regarding digital preservation have been developed as of yet, nor are long-term preservation strategies implemented and applied. Instead it has become clear that most likely a cooperative solution for long-term preservation will be sought. While this means that actual long-term preservation measures such as migration or emulation will not be carried out in JUWEL, the repository and its staff will nonetheless have important responsibilities in Administration, Data Management, and Preservation Planning among others, as it will be their task to make sure that the long term archive receives digital objects and metadata adequate to long-term preservation.

JUWEL currently contains about 3.500 titles and runs on the following system:

- LINUX OS
- DSpace 1.4.2[56]
- SQL Server: PostGreSQL 8.07
- Webserver Tomcat 5.5.17
- Backup: IBM Tivoli Storage Manager (see Hinz 2008)

Although the question of long-term preservation capabilities did play a role in selecting DSpace[57] as a software for JUWEL according to Dr. Alexander Wagner, who is part of the team responsible for managing JUWEL at the FZ's Central Library, this was not the primary concern at the time the selection was made. Instead, the decision to use DSpace was on the one hand governed by the fact that it has comparably well-written code and

54 http://www.fz-juelich.de/zb/index.php?index=251 and http://www.fz-juelich.de/zb/verlag/ – 01.11.2009.

55 JUWEL will, however, serve as a prototype for a (closed access) repository for (historical and partly rare) collections, which will be digitized by the Central Library of the Forschungszentrum for the purpose of long-term preservation. Hence considerations of how long-term preservation can actually be achieved will gain in importance in the near future.

56 The 1.4.2 release of DSpace was thoroughly tested by a National Library of Medicine working group during the implementation of plans to establish a digital repository. For these tests, requirements were formulated on the basis of the OAIS model so that in the documentation of the test results, a matching of DSpace functions with OAIS functional entities is available. Based on these tests the use of Fedora as a basis for the digital repository was recommended. For further information and details, see Marill and Luczak 2009 as well as http://www.nlm.nih.gov/digitalrepository/ which contains links to the project documents (NLM 2009), including the Policies and Functional Requirements Specification (NLM-DRWG 2008) as well as the Recommendations on NLM Digital Repository Software (NLM-DRESWG 2009), which also contains the test results for DSpace in Appendix C.

57 http://www.dspace.org/ – 03.11.2009

uses technology which the responsible IT-staff considered preferable (e.g. Java and JavaService Pages). On the other hand, the fact that DSpace has a wide international user community, and that hence its ongoing development and support at this point seem fairly certain[58], played a role in the decision as the availability and further development of the software is also an important consideration in the context of long-term preservation. There is, however, also a drawback to using DSpace for a German repository, as currently only very few (if any) German developers contribute to the DSpace source code on a regular basis, which means that certain characteristics and aspects relevant only to the German repository landscape, such as the DNB's XMETADISS, are not implemented into DSpace.[59] Contributions to DSpace source code follow very strict coding and documentation standards which are difficult to adhere to if one is not involved in DSpace programming work full time.[60]

Like pedocs, JUWEL is currently preparing to apply for the DINI 2007 certificate.

2.1.3 Qucosa (Quality Content of Saxony)

Launched in July 2009, when it replaced a repository which had been in operation since 1996, Qucosa[61] is a service hosted by the Sächsische Landesbibliothek – Staats- und Universitätsbibliothek (SLUB) Dresden. It is co-funded by the European Regional Development Fund (ERDF). Further partners in this project are the academic and research libraries in Saxony.

According to Dr. Andreas Kluge, head of the information technology department at the SLUB Dresden and responsible for the management of Qucosa, a concern with issues and questions of digital long-term preservation has existed at the SLUB Dresden for quite some time, especially as in addition to Qucosa, the SLUB also operates a digitization center.[62] The latter carries out mass digitization projects, among others, of manuscripts, maps and photographs.

This awareness of long-term preservation and its concerns lead, for example, to a close monitoring of the German kopal project[63] (see below) and to considerations to establish a cooperative long-term preservation system for Qucosa drawing on services developed by kopal. In order to pursue the possibilities offered by kopal further, an

58 According to ROAR, on November 1st, 2009 DSpace had 485 installations worldwide, thus making it the repository software with the biggest user community (see http://roar.eprints.org/).

59 As the FZ does not have the right to award doctorate degrees (“Promotionsrecht”), the XMETADISS standard is not relevant to JUWEL. In addition, German DSpace installations for university-based repositories exist, which must therefore have found a workaround for this problem. XMETADISS is, however, not part of the DSpace standard installation.

60 See, for example, DSpace's Contribution Guidelines at http://wiki.dspace.org/index.php/ContributionGuidelines – 31.10.2009. See also Salo (2007), who comments on available repository software that “[...] the architectures of these software packages are deeply innovation-unfriendly. DSpace and EPrints are heavily overengineered and written in Java and Perl respectively, rather than one of the simpler Web-scripting languages such as PHP or Ruby. Even slight modifications are out of reach of most members of 'the community' due to lack of specialized expertise and steep learning curves” (no pag.).

61 http://www.qucosa.de – 03.11.2009

62 http://digital.slub-dresden.de/digitalisierungszentrum/ – 03.11.2009

63 http://kopal.langzeitarchivierung.de/ – 03.11.2009

application for a project grant was submitted to the German Research Foundation in January 2008. The application was approved in the spring of 2009; it is currently expected that the first step towards long-term preservation services for the digital materials created and collected at the SLUB Dresden will be completed over the coming two years. For this purpose, a preservation workflow in which Qucosa will serve, in OAIS terminology, as a producer for the long-term archive (e.g. DIAS; see below) will be developed and tested.

Like pedocs, Qucosa is based on the OPUS software – however, it was updated to OPUS 4 just recently. It thus uses the newest generation of OPUS, which underwent a complete, cooperative re-development in the course of a project funded by the German Research Foundation between July 2008 and June 2009. Although the development of OPUS 4 has at this point not been completed, the decision was made to update Qucosa to OPUS 4 despite the fact that certain functions are not yet available. Thus, while Qucosa is currently based on an OPUS 4 core, certain components of the working system – including the Web UI – were developed by the SLUB Dresden.

The original decision to use OPUS was made in the context of the ERDF grant application for Qucosa, and very similarly as in the case of pedocs, this decision was strongly motivated by the fact that OPUS has a very broad and strong user community in Germany. Thus, although Fedora and DSpace were also considered, there was a strong concern with being able to actively contribute to the development of the software in the community, and it was feared that to become an active member of international communities strongly dominated by Anglo-American institutions and programmers, would be difficult. The future development both of Qucosa and of OPUS 4 will have to show whether Qucosa can indeed partake in the benefits of a big user community.[64]

In contrast to pedocs and JUWEL, Qucosa has already been awarded the DINI 2007 certificate. In consequence, the DINI criteria will not be addressed explicitly in the following.

kopal and koLibRI

With a strong focus on cooperation, sustainability, re-usability, the kopal project (2004-2007), funded by the BMBF, aimed at “develop[ing] a technological and organizational solution to ensure the long-term availability of electronic publications.”[65] The project, which was strongly concerned to implement national and international standards such as Dublin Core, LMER, OAIS, METS, and URN, was carried out by the German National Library (DNB), Göttingen State and University Library (Niedersaechsische Staats- und Universitätsbibliothek [SUB] Göttingen), the Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG), and IBM Deutschland GmbH.[66] The result of

64 The prerequisite for this is the existence of a critical mass of users for OPUS 4, and that by the time this critical mass exists the Qucosa system is not too far removed from its OPUS 4 base.

65 http://kopal.langzeitarchivierung.de/index_ziel.php.en – 03.11.2009.

66 See http://kopal.langzeitarchivierung.de/index.php.en – 03.11.2009. A summary of the project-related activities is provided by the presentation “'kopal goes live' Nutzungsmodelle und Perspektiven eines Langzeitarchivs digitaler Informationen” (Altenhöner 2007). More recent perspectives on kopal are offered in a presentation by Tobias Steinke (2008).

the project is a system based on a DIAS core (see below) available for reuse by institutions interested in taking the preservation of digital material into their own hands. Alternatively, “an institutional user can have its [sic] own 'locker', i.e., secure storage space with data under its [sic] own administrative control.”[67]

The Digital Information Archiving System (DIAS) developed by IBM provides a storage and preservation solution for digital objects and was developed originally for the KB in 2000.[68] The DIAS core consists of modules corresponding to the OAIS functional entities, including a Preservation Manager performing the following tasks:

> 1. Identifying the digital assets in danger of becoming inaccessible/unusable due to technology changes
> 2. Planning the activities associated with preservation, i.e. implementing migration and/or emulation strategies
> 3. Specifying the software and hardware environments required to render a digital asset.[69]

Among the software tools developed for kopal is koLibRI, the kopal Library for Retrieval and Ingest,

> a framework for the integration of a long-term archiving system, like the IBM Digital Information Archiving System (DIAS), into the infrastructure of an institution. In particular, it organizes the creation and ingest of archival packages into DIAS and provides functionalities to retrieve and manage these packages […]. In short, koLibRI generates a [sic] XML file according to the METS schema out of the metadata, provided with the object to archive or generated by JHOVE, bundles it into an archive file together with the object (.zip or .tar) and delivers this Submission Archiving Package (SIP) to the DIAS system. (Funk et al. 2007, 4)

Providing an interface between the database or repository containing the material to be submitted to the long-term preservation system, koLibRI consists of several internal and external tools and elements, among others a database for storage of metadata (e.g. for statistics or rights management), a module which will retrieve descriptive metadata from PICA-OPACs, as well as a tool for the correction of TIFF metadata and JHOVE for format recognition and validation as well as the extraction of technical metadata (see Ludwig 2007).[70] Qucosa is planning to generate SIPs by means of the koLibRI software tool in the future, and to submit these SIPs to the long term archive (e.g. a “locker” in the DIAS core). Currently, tests with koLibRI are carried out and it is hoped that the tool can be implemented into Qucosa by March 2010.

2.2 Ingest

> This entity provides the services and functions to accept Submission Information Packages (SIPs) from Producers […] and prepare the contents for storage and management within the archive. Ingest functions include receiving SIPs, performing quality assurance on SIPs,

67 http://kopal.langzeitarchivierung.de/index_service.php.en – 03.11.2009.
68 Cf. http://www-935.ibm.com/services/nl/dias/is/implementation_services.html – 24.10.2009
69 http://www-935.ibm.com/services/nl/dias/is/preservation_manager.html – 24.10.2009
70 In addition, koLibRI contains the MigrationManager, a component “which manages and executes migrations” on the basis of “individual migration workflows” which can be configured according to the requirements of the institution (see Funk et al. 2007, 24). However, the MigrationManager is according to the documentation only a prototype at the current stage not suitable for use in complex scenarios.

> 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 2002, 4-1)
>
> Ingest relies upon rules established on the organizational side to determine the metadata that must be present, the formats that are acceptable, the means that may be used for transferring objects, and the quality checks that must be performed […]. The ingest functions must be able to determine that the files and their metadata are complete and correct as sent. Next the metadata must be generated to tie the objects into the structure of the archive by generating the Archival Information Package (AIP). Any text that will be used for searching or for display must also be created and associated with the objects – the Descriptive Information – and sent to Data Management. After all the complex objects are created, they are moved to Archival Storage.[71]

During Ingest the repository will typically receive an Information Package from a producer.[72] It is with Ingest that "the repository takes intellectual control of the [Information Package] and processes it for preservation" *(*Bodleian Library 2007, 55). In the following, the process in which an institutional or subject repository receives a digital object for inclusion in its digital collection will be described as Ingest, despite the fact that the repositories treated here are not the long-term archive. While it would be possible to designate this process as a kind of pre-Ingest, it will become apparent in the following that important data and information is collected at this stage, without which long-term preservation would not be possible later, and that hence the repository Ingest can at least conceptually be regarded as part of the "Ingest proper."

From the point of Ingest onward the repository has the responsibility for the ingested Information Package and the preservation of the content information according to the defined storage or preservation goals. In order to fulfill its tasks, the repository must in particular have physical and technical control over the digital objects it receives, a fact that is reflected in all three criteria catalogs (see Appendix A). In particular, this means that the SIP must be free from any Digital Rights Management software or technical measures limiting, for example, the possibilities to view, save, copy, or print the digital object. This latter aspect, which also is highly important in the context of user access to objects stored in a repository, is explicitly mentioned in the explanatory notes to the nestor criterion 9.3[73] as well as in the DINI criteria (see DINI 2007, 2.8[74]). In contrast, TRAC, while stating that

71 http://www.icpsr.umich.edu/dpm/dpm-eng/foundation/oais/overview.html – 24.10.2009

72 However, the Ingest process may also be carried out for an existing AIP which has been changed or updated in any way. Thus, the OAIS itself might act as producer, for example, when AIPs which underwent a migration are re-submitted to the repository as SIPs. See OAIS 2002, chapter 4.2 for a detailed definition of its information model, including information packages in chapter 4.2.2.

73 Hereafter, the criteria catalogs will be cited as short name (nestor, TRAC, and DINI respectively), year, and number of the criterion rather than page number.

74 Note, however, that DINI is somewhat unclear in its remarks concerning technical measures for digital rights management. Thus, the following minimum standard is described under 2.8 Langzeitverfügbarkeit (Long Term Accessibility): "Die gegebenenfalls *zusätzlich* zu den eingereichten Originaldateien des Autors erstellten Archivkopien sind frei von Schutzmaßnahmen (DRM), die eine Anwendung von Strategien zur Langzeitverfügbarkeit (Migration, Emulation) verhindern" (2007, 2.8; emphasis added). Hence DINI makes a distinction between the submitted files, which are *not* required to be free from DRM, and archival copies, which *can* but do not have to be created, and which would have to be free from DRM. Thus the creation of DRM-free archival copies is currently not a requirement in the DINI criteria, a fact that is, from the perspective of long-term preservation, certainly problematic as a repository that will decide in the future to undertake long-term preservation activities may have a large amount of digital objects unfit for preservation in its collection. In addition, the following is recommended (i.e. not required): "Nutzung von offenen Dateiformaten, die zur Langzeitarchivierung geeignet […] und frei von Schutzmaßnahmen (DRM) sind" (DINI 2007, 2.8). This statement is similarly unclear, in that DRM is something applied to files or

“[t]he repository must obtain complete control of the bits of the digital objects conveyed with each SIP” (2007, B1.5), does not focus on DRM or similar issues so much but points to the fact that digital objects may contain references to other digital objects, and that the repository should attempt to harvest and ingest these objects as well in order to guarantee that the digital objects it preserves are as complete as possible.

Part of the Ingest process is Quality Assurance, during which the repository ascertains the integrity and authenticity of the SIP before archiving it in the repository. Definitions of either concept abound, and Clifford Lynch's observation, dating back to the year 2000, that authenticity and integrity are “elusive properties” which, as we try to define them, “recurse into a wilderness of mirrors, of questions about trust and identity in the networked information world” (2000, no pag.), still rings true.[75] Authenticity can be understood as a measure for the digital object's “trustworthiness” in the sense that an object is authentic to the degree that it is what it seems or purports to be (see nestor 2008, 7).[76] As Factor et al. note,

> [a]uthenticity refers to the reliability of the data in the broad sense [...]. To validate authenticity of a preserved data object provenance is needed, i.e., the documented history of creation, ownership, accesses, and changes that have occurred over time for a given data object. Also a means is needed to guarantee that data is whole and uncorrupted (integrity). (2009, no pag.)

It follows that authenticity strongly depends on the repository's ability to ascertain and guarantee that a digital object was created by the specified author/source at the specified time – information that is also indispensable if the object is to be usable for/in scholarly research – and on a documentation of any transformation the object may have undergone since submission:

> Ein wichtiger Aspekt ist, dass das vorliegende Objekt von der angegebenen Quelle und zur angegebenen Zeit erstellt wurde. Ferner schließt Authentizität den lückenlosen Nachweis aller im Sinne der Erhaltungsmaßnahmen durchgeführten Transformationen an den Objekten mit ein. (nestor 2008, 7)

Such provenance information is crucial in that in the context of long-term preservation we need to take into account that

> in most cases digital objects cannot be preserved without any change in the bit stream, and we have to modify the original object to have the ability to reproduce it in the future. Unfortunately, this runs counter to the assumption that preserving authenticity implies retaining the identity and integrity of a digital object, i.e. free from tampering or corruption. It is a sort of paradox, where preservation entails change, while authenticity needs fixity. (Factor et al. 2009, no pag.)

---

software rather than file *formats*.

75 See the *DigiCULT* thematic issue on “Integrity and Authenticity of Digital Cultural Heritage Objects” (2002) for very similar observations. http://www.digicult.info/downloads/thematic_issue_1_final.pdf – 31.10.2009. “Authenticity” is also the focus of two letters to the editor published in issue 4.2 (2009) of the International Journal of Digital Curation (see Wilson 2009 and Jantz 2009).

76 See also the revised version of the OAIS model (May 2009 version for public comment), which defines authenticity as “the degree to which a person (or system) may regard an object as what it is purported to be” (1-8). The question of authenticity is also considered by Wilson in the InSPECT 2.2 work package (2007). Wilson draws attention to the fact that rather than adhering to a “broad meaning” of authenticity, with “all the connotations of that much overused word truth,” it makes sense in the context of digital long-term preservation to “limit 'authenticity' to its archival meanings to do with what a record purports to be and how it was created. Variations of this definition abound but the central core of the concept is fixed [...]. For the UK National Archives assessing authenticity involves establishing the integrity and identity of the object – integrity here referring to the objects 'wholeness and soundness', and identity referring to attributes such as context and provenance” (Wilson 2007, 4).

Integrity can be defined as “completeness” and “intactness” of the digital objects – in particular, of their significant properties (nestor 2008, 6; my translation), and can be put at risk both by malfunctioning technology and by human action (intentionally or by mistake) (nestor 2008, 6). During Ingest, integrity may, among others, be ascertained and protected by means of virus checking, cyclic redundancy checks or other checksums, as well as by ascertaining the validity of the file format. Thus, in order to determine the digital object's properties as comprehensively as possible, the following actions should be performed:

> - Format *identification* is the process of determining the format to which a digital object conforms; in other words, it answers the question: “I have a digital object; what format is it?”
> - Format *validation* is the process of determining the level of compliance of a digital object to the specification for its purported format, e.g.: “I have an object purportedly of format *F*; is it?” […].[77]

For the purpose of long-term preservation, such identification and validation processes, which can, for example, be carried out with the tool JHOVE (JSTOR/Harvard Object Validation Environment)[78], should be combined with format validation, i.e. “the process of determining the format-specific significant properties of an object of a given format, e.g.: 'I have an object of format F; what are its salient properties?'” (ibid.).[79] Such format identification and validation measures clearly exceed the range of “traditional”, non-preservation-related repository tasks. Nonetheless it would be desirable if available repository software included such features as a standard. Thus, if the software made functions important in the context of long-term preservation easily available, repositories deciding in the future to begin implementing long-term preservation strategies would find that their collections are – at least to some extent – already prepared for such preservation efforts, a circumstance that might make it considerably easier for them to decide taking a step towards long-term preservation.

Both authenticity and integrity of files may be protected by making use of secure connections (e.g. by means of SSL or HTTPS protocols) for file and metadata upload by submitting authors. In the context of institutional and subject repositories, this not only helps to protect the personal information authors give about themselves when submitting a digital object (i.e. information which will not be visible in the public record for the publication), but also prevents third parties from interfering with the submission. In order to identify and authenticate a depositor, many repositories work with user accounts and log

77 http://hul.harvard.edu/jhove/ – 03.11.2009.

78 Another freely available tool for format identification is DROID (Digital Record Object Identification), which was developed by the UK National Archives. It is available for download from http://droid.sourceforge.net/. Together with JHOVE and other tools, DROID forms part of FITS (File Information Tool Set), an open source tool which “combines the abilities of many different open-source file identification, validation, and metadata extraction tools. The File Information Tool Set (FITS) acts as a wrapper around these tools, invoking, normalizing, and combining their output” (Spencer et al. 2009; no pag.). Further format identification and validation tools are listed in the “Preservation-related Tools” section of the IDEALS (Illinois Digital Environment for Access to Learning and Scholarship) Wiki at https://services.ideals.uiuc.edu/wiki/bin/view/IDEALS/Internal/PreservationTools – 03.11.2009.

79 In this context, so-called format registries play an important role. Among the most well-known registries are those built up by the PRONOM or GDFR initiatives, both of which “joined forces” in April 2009 to form the Unified Digital Formats Registry (UDFR) (http://www.gdfr.info/index.html – 03.11.2009).

in procedures or even require authors to identify themselves by means of PGP/GPG keys.[80]

The nestor and TRAC catalogs contain criteria dealing with integrity and authenticity of the files submitted to and archived by the repository. While, as we have seen, the nestor criteria mention the two concepts explicitly (cf. criteria blocks 6 and 7), TRAC only refers to “completeness and correctness” of the submitted objects (2007, B1.4). Important aspects of the concept of authenticity as outlined above are touched upon in TRAC criterion B1.3, focusing on “authenticat[ing] the source of all materials,” and requiring the repository to “ensure the digital objects are obtained from the expected source, that the appropriate provenance has been maintained, and that the objects are the expected objects” (2007, B1.3). DINI's overall approach to authenticity and integrity is primarily from the technical side. Thus 2.5 (Sicherheit, Authentizität und Integrität) among others[81] recommends the use of advanced digital signatures according to SigG 2001 (Gesetz über Rahmenbedingungen für elektronische Signaturen)[82] (DINI 2007, 2.5.2) and requires that (technical) measures are taken to prevent that files are uploaded onto the server which do not meet the criteria outlined in the repository's policy (DINI 2007, 2.5.1). Thus, for example, files containing viruses or files with formats not accepted by the repository should be rejected automatically.[83]

During the Generate AIP functional sub-entity, an Archival Information Package is generated from the Submission Information Package. This AIP must, according to OAIS, “conform to the archive's data formatting and documentation standards” (2002, 4-6; emphasis omitted), that is, it must be built and structured in accordance with packaging designs developed by the Preservation Planning Functional Entity (see below) and adopted by Administration. Like all OAIS Information Packages, the AIP

> is a conceptual container of two types of information called Content Information and Preservation Description Information (PDI). The Content Information and PDI are viewed as being encapsulated and identifiable by the Packaging Information. The resulting package is viewed as being discoverable by virtue of the Descriptive Information. (OAIS 2002, 2-5)

80 TUBdok, the OPUS-based repository of the Technical University Hamburg-Harburg, uses PGP/GPG keys to authenticate the source of materials published on its server. Thus, every record contains a link to a page serving as evidence for the attached document's integrity (“Unversehrtheitsnachweis”), including a reference to digital signatures of the author(s) and the library. See Marahrens 2005 for a description of the project in which these and other changes were implemented.

81 In addition, DINI criterion 2.5 for example suggests access controls to the server (required; see 2007, 2.5.1) and contains the requirement that a document whose content was changed has to be treated like a new document (see 2007, 2.5.2). These and other relevant criteria, requirements, and recommendations from 2.5 will be listed and discussed under the functional entities for which they are most relevant.

82 It does not become entirely clear how the signature is to be used and what is meant to be signed with it. See Winkler (2008) 76-78 for an explanation of possible uses of digital signatures in the context of long-term preservation. Becker explicitly comments on the DINI suggestions and points to possible problems the use digital signatures can pose for long-term preservation efforts (2008, 36-37).

83 Of course, each repository must decide for itself whether it wants to restrict the formats it will accept, and whether these should really be rejected upon upload. As pointed out by Dr. Wagner, such an automated rejection process might mean that files which could, for example, be converted into a different format will not reach the repository at all as users might not want to make a second attempt or contact repository staff after a file was rejected. This scenario shows quite clearly that in contrast to digital long-term archives solely serving the purpose of preservation, for institutional or subject repositories long-term preservation efforts might conflict with usability and the concern to acquire a critical mass of repository content.

It is with this step that the metadata information (with the exception of certain Descriptive Metadata) for the Archival Information Package must be generated or assembled in the correct manner.[84]

The nestor and TRAC criteria catalogs take different – but equally accepted – approaches to the classification of metadata. While nestor refers to descriptive, structural, technical, provenance, and rights management metadata[85], TRAC adheres more closely to OAIS terminology in identifying Descriptive Metadata/Information, Representation Information (RI), and Preservation Description Information (PDI). According to OAIS, RI is "[t]he 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 2002, 1-13). The Paradigm Workbook describes RI as follows:

> Representation Information (RI) provides structural and semantic information that permits the interpretation of Content Data Objects (i.e. the archival material accessioned) so that they may be rendered accessible: translating bits into information that is meaningful to the repository's Designated Community […]. (Bodleian Library 2007, 79)[86]

Dividing into Provenance, Context, Reference, and Fixity information, PDI "is needed to preserve the Content Information, to ensure it is clearly identified, and to understand the environment in which the Content Information was created" (OAIS 2002, 2-6). The four different types of PDI are briefly described as follows:

> - **Reference information** – allows for consistent, enduring and unique identification
> - **Provenance information** – details the origin and custody of an information object, including events in its life once submitted to the repository
> - **Context information** – documents the relationships of an information object with its environment, including why it was created and relationships with other information objects
> - **Fixity information** – documents authenticity using mechanism which evidence that no undocumented alterations have been applied. (Bodleian Library 2007; 80; emphasis in the original; see also OAIS 2002, 2-6)

Despite the differences in the TRAC and nestor approaches to classifying the required metadata the respective catalogs' requirements are largely similar and include all types of metadata crucial for digital long-term preservation.

---

84 Note that some information can only be completed as the AIP enters Archival Storage. Thus, for example, the Coordinate Updates function "incorporates the storage identification information into the Descriptive Information for the AIP [...]" (OAIS 2002, 4-6).

85 Cf., for example, NISO's Understanding Metadata (2004), which distinguishes the following "types of metadata":
- Descriptive metadata describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords.
- Structural metadata indicates how compound objects are put together, for example, how pages are ordered to form chapters.
- Administrative metadata provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it. There are several subsets of administrative data; two that sometimes are listed as separate metadata types are:
- Rights management metadata, which deals with intellectual property rights, and
- Preservation metadata, which contains information needed to archive and preserve a resource. (1)

86 This information can, for example, be generated with the help of JHOVE, which reports a "set of characteristics […] about a digital object […] known as the object's representation information […]. The standard representation information reported by JHOVE includes: file pathname or URI, last modification date, byte size, format, format version, MIME type, format profiles, and optionally, CRC32, MD5, and SHA-1 checksums. Additional media type-specific representation information is consistent with the NISO Z39.87 Data Dictionary for digital still images and the draft AES metadata standard for digital audio" (http://hul.harvard.edu/jhove/ – 03.11.2009).

In contrast to nestor and TRAC, the DINI catalog primarily focuses on descriptive metadata, the use of subject terms and/or classification in particular (see 2.6.1 Sacherschließung), in its requirements. The creation of technical metadata, format identification, and preservation metadata is merely recommended (see recommendations 2.6.2, 2.8), a fact that highlights again that the DINI catalog is first and foremost addressed at "traditional", non-preservation-centered repositories. In addition, DINI requires metadata to be structured according to Dublin Core Simple and recommends the Dublin Core Qualified element set.[87]

All three criteria catalogs require repositories to make use of persistent identifiers (PIs) in order to make it possible to address their digital objects permanently. According to the Paradigm Workbook, in contrast to URLs, PIs "provide a means of connecting and distinguishing between an identifier for an object (which should be permanent) and an object's location (which may change)" (Bodleian Library 2007, 48). As explained by the DNB, persistent identifiers serve to

- permanently address a digital object,
- simultaneously refer to several storage locations,
- unequivocally identify a digital object as information entity worldwide, but also
- reliably identify single parts of a digital object.[88]

All of these functions are important if not indispensable not only in the context of digital long-term preservation, but also in the context of everyday scholarly work, which similarly depends on the ability to permanently identify and address (referenced) digital objects in a manner that is wholly unambiguous.

2.2.1 pedocs Ingest

Ingest: Receive Submission

Digital objects ingested into pedocs can be deposited through two channels: Depositors can upload born-digital or digitized documents through a web interface[89] requiring them to enter a certain amount of (descriptive) metadata. In addition, some documents – especially digitized material – is uploaded into pedocs by the repository staff. According to the pedocs policy Guidelines, only PDF or PDF/A file format are accepted for publication in the repository[90], and hence file formats different from PDF will not be accepted for the purpose of archiving unless an individual agreement has been made between the author and the repository staff. In such exceptional cases, repository staff will convert submitted files to PDF.

Both the repository's policy document and the author agreement, which was developed in cooperation with a law firm and will be publicly available soon, specify that

87 This requirement is related to the fact that Dublin Core metadata can be harvested through OAI-PMH, which is particularly relevant to institutional and subject repositories.
88 http://www.persistent-identifier.de/?link=203&lang=en – 11.10.2009.
89 See http://www.pedocs.de/uni/index.php – 10.10.2009.
90 Technically, however, upload of different file formats is possible, as these are not rejected by the repository software. Note that this might mean that the DINI requirement defined in criterion 2.5.1 is not satisfied.

documents submitted to the repository must be free from any DRM, password protection, or other technical protection measures that might be used, for example, to limit or disable the print option. Thus – in the case that authors follow the guidelines – pedocs obtains full physical/technical control over the digital objects as required by nestor, DINI, and TRAC. As of October 2009, a newly developed tool is used which automatically identifies documents with DRM restrictions during upload. As a result, while pedocs still contains files with such technical limitations, uploaded before the new tool was implemented[91], these can now be identified and further action be taken where necessary.

Ingest: Quality Assurance

Integrity: As outlined above, virus checking and cyclic redundancy checks help to ascertain a submitted file's completeness and correctness (on a technical/bitstream level), and to determine whether any errors occurred during the transmission. Currently, virus checks are not implemented into the pedocs workflow as such, i.e. files uploaded through the web interface are stored in the pedocs file system without having been checked for viruses. However, every uploaded file is opened by repository staff in order to see if it is functional and is thus automatically scanned by the antivirus software installed on each computer.

As explained above, a digital object's integrity also depends on the validity of its file format. Thus, small mistakes in a file's code might make it harder to preserve the object for the long term, for example, because the behavior of corrupted files during format migrations cannot be predicted. A format validation (preceded by a format identification), which ascertains that a file's bitstream is composed exactly according to the specifications, is thus highly important in the context of long-term preservation. pedocs carries out format validation and characterization as part of the ingest process by means of JHOVE, which is used to generate a limited amount of technical metadata to be submitted to the DNB as part of information packages (i.e. file format, format version, supplemented by checksum; also see below). Invalid files will not be submitted to the long-term archive.

The completeness and correctness of the SIPs is additionally ascertained by means of intellectual control, especially with regard to metadata. Thus, for the purpose of quality assurance the metadata submitted by authors is sometimes modified; also, further metadata can be added by repository staff in order to complete the information package stored and published on the pedocs server. This practice is explicitly mentioned in the pedocs policy and thus authors are aware of the fact that their submitted metadata may be subject to modification.

Authenticity: In the context of the Ingest functional entity, the question of authenticity of digital objects is primarily a question of the extent to which the submitting person or

91 Thus, for example, the documents attached to the record 807 are both encrypted and password protected. In addition, the possibility of copying text from the PDF has been disabled. See http://www.pedocs.de/volltexte/2009/807/ – 30.10.2009.

institution can be trusted – e.g. trusted to be the author of a work and/or to have the right to publish it through the repository. Although there are different possible ways of authenticating a user (see above), none of these is currently used at pedocs, mainly because there is a concern not to create too many obstacles/hindrances that might keep authors from submitting their works to the repository. In contrast to other types of repositories, as a subject repository pedocs is not in the convenient situation that its users already have an ID and/or password, e.g. for the use of other services. Thus, for a repository belonging to the services offered by a University Library it can be comparably easy to (re-)use the ID and password its patrons already have (e.g. for the use of the library's on- and offline services) as a means of user authentication with the repository – an approach taken by the TU Hamburg-Harburg, for example. As a disciplinary repository, pedocs does not have such infrastructure ready at hand. This is relatively uncritical where pedocs enters into direct contact with producers, as is the case, for example, with many publishers approached by the repository. However, where contact between repository and submitting person is only based on e-mail, authenticating the source of submitted material remains a problem.

pedocs is addressing this problem with its newly drafted author agreement, which will have to be accepted by submitting authors in a double opt-in procedure in the future.[92] Thus, producers will have to confirm in a contract (author agreement) that they have the rights to publish the document in question with pedocs. They e-mail the contract to pedocs, are notified that the contract was received by the repository, and have to confirm again that they intended to submit the contract. This procedure cannot offer the same confirmation of the identity of the depositor as, for example, the creation of user accounts in combination with digital signatures. Yet depositors do close a contract with pedocs by submitting the agreement, in which they ascertain that they are who they claim to be – i.e. the person holding the necessary rights to publish the submitted document with pedocs –, and this does offer some (if limited) assurance concerning the authenticity of the source.

In the process of conceptualizing and developing pedocs, a further problem was identified just recently, which also touches upon the question of authenticity or, more precisely, the question of how accurate a representation of the submitted “original” object the pedocs record and the document attached to it are. This question concerns publications that were digitized by pedocs (or a third party service provider) and that underwent OCR in the digitization process. Depending on the quality of the original printed document, OCR may misrecognize characters, which will lead to an inaccurate representation of the original text. According to Tanner, Muñoz and Ros,

92 Currently authors have to accept the following shorter agreement when submitting their documents to the repository through the web interface: “Ich übertrage dem Deutschen Institut für Internationale Pädagogische Forschung (DIPF) sowie der Deutschen Nationalbibliothek in Frankfurt bzw. Leipzig und der zuständigen Sondersammelgebietsbibliothek das Recht, das/die übermittelte/n Dokument/e elektronisch zu speichern und in Datennetzen öffentlich zugänglich zu machen. Ich übertrage dem DIPF ferner das Recht zur Konvertierung der übertragenen Datei zum Zwecke der Langzeitarchivierung unter Beachtung der Bewahrung des Inhalts. Die Originalarchivierung bleibt erhalten” (http://www.pedocs.de/uni/index.php – 11.10.2009).

> [t]he majority of OCR software suppliers define accuracy in terms of a percentage figure based on the number of correct characters per volume of characters converted. This is very likely to be a misleading figure, as it is normally based upon the OCR engine attempting to convert a perfect laser-printed text of the modernity and quality of, for instance, the printed version of this document. In our experience, gaining character accuracies of greater than 1 in 5,000 characters (99.98%) with fully automated OCR is usually only possible with post-1950's printed text, whilst gaining accuracies of greater than 95% (5 in 100 characters wrong) is more usual for post-1900 and pre-1950's text [...]. (2009, no pag.)

As pedocs' staffing is not adequate for intellectual control of digitization and OCR results, some of the digitized material might not be an entirely accurate representation of the original source, which might, for example, cause publications not to be findable in the database due to spelling errors in the title. The approach taken to this problem by pedocs is to mention it in the policy document, which since its last actualization informs users and producers that "[i]n specific cases, pedocs can use scanning and OCR software to digitize documents that are provided in print only. Please note that the results from these procedures may contain errors depending on the printed master document" (pedocs Guidelines).

To further address this problem, pedocs might want to consider to add a note to the records of digitized documents, pointing out to users that what they are viewing is the result of a digitization process with OCR. This measure would help to ensure the authenticity of the digital object in that "provenance" information added to a record leads to greater transparency concerning the origin and creation of the digital object users are viewing (cf. Factor et al. 2009).[93] A further, if somewhat more taxing possibility is the inclusion of the pre-OCR version of the digitization in addition to the version which underwent OCR: Thus, the image file could be attached to the record for the digitized document along with the OCR-version, and with a description explaining the relationship between the two files. This practice would would give users the opportunity to compare for themselves how strongly the OCR-version deviates from the mere image. However, as the number of files to be archived would be doubled for digitized publications, this does not seem a feasible method.

pedocs is still working on finding the best possible solution to the problem by testing different OCR software. Part of the challenge is to also to find a solution which complies with accessibility standards for webpages ("Barrierefreiheit"). According to Dr. Julia Kreusch of the pedocs team, in the normal use case, what user will see when accessing a publication is the image file, which will – normally – be an authentic reproduction of the original print version of the publication. Hence, a "provenance note" as just suggested may not go to the heart of the problem.

It was stated above that one way of protecting the integrity and authenticity of files during submission as well as the information entered into the submission form by authors is the use of secure connections to prevent someone from interfering with the

93 Since mid-October 2009, documents published in pedocs receive a cover sheet (see below), which can also contain a note saying that the publication was digitized by pedocs (see, for example, http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0111-opus-16178 – 24.10.2009).

transmission. However, pedocs does not secure connections (e.g. SSL or SSH protocols) in the upload process at the moment.

Ingest: Generate AIP

After the metadata have been completed and/or modified, and the uploaded file has – where necessary – been converted to PDF, the SIP is transformed into a pre-stage AIP, i.e. it is saved in the pedocs file system on the DIPF servers in Berlin.

In accordance with the requirements of the criteria catalogs, pedocs records receive a Uniform Resource Name (URN) allowing them to be permanently addressed and identified. As an institution not organized in a library association, DIPF was allocated a sub-namespace and is hence able to assign URNs within this namespace. Such URNs are structured in the following manner: *urn:nbn:de:[four digit number]-[unique production number][check digit].*[94] While the actual "object" to which the URN is assigned is the record for the published document, this document itself can then be addressed via the URL embedded in the record, which is noted in the metadata along with the URN. This URL has the structure http://www.pedocs.de/volltexte/[Year]/[ID]/ and contains an identification number which is unique within pedocs. Both the URN and the URL are then registered with the DNB, which is also notified should changes to the URL occur.[95]

Although URNs were assigned by the pedocs software from the outset, these are being registered with the DNB only since October 2009. Currently, newly assigned URNs are registered immediately, whereas previously assigned URNs will be registered retrospectively. It is for this reason that the URN field is not currently visible in all records, as the field was suppressed as long as the URN was not functional. In the future, all metadata records and the cover sheets added to the documents will contain a reference to the URN. As outlined above, the TRAC criteria also require that existing persistent identifiers are included in the metadata. No such field exists in the pedocs metadata schema, nor is it planned to implement one or submit previously assigned identifiers to the DNB.[96]

The pedocs software allows the attachment of more than file to a record; however, **structural metadata** to indicate, for example, in which order these files should be viewed is not added. Thus, in the previously cited record with the ID 807, it is not obvious to users that the file "Dokument 2.pdf" (logically) precedes "Dokument 1.pdf" and in fact contains explanatory notes relevant to understanding the content of the main file. pedocs is aware of this problem (which also occurs in Qucosa, the other OPUS installation considered in this work) and is working on a solution which will allow staff to determine the order in which files are displayed in the record. However, the problem might also simply be solved by adding short descriptions to files where more than document is attached to a record.

94 See www.persistent-identifier.de – 11.10.2009.
95 See the DNB homepage (http://www.d-nb.de/netzpub/erschl_lza/np_urn.htm – 03.11.2009) for further information about URNs at the DNB and the DNB's associated strategy.
96 According to Kreusch, an existing persistent identifier might be included in the field for general remarks. This field, however, will not be submitted to the DNB.

One of the shortcomings of the software on which pedocs was originally based is its inadequacy in identifying and representing the different kinds of relationships that might exist between digital objects (a requirement explicitly voiced by the nestor criteria: see 2008, 12.1), which can be expressed by means of **structural or context metadata**. In order to improve the current situation, pedocs is planning to make the hierarchical relation existing between journal titles, issues, and the individual articles composing them transparent with the help of a (metadata-based) browsing functionality. Thus it will become possible to browse journal titles and issues to access all related articles published in pedocs. These relations will be established via the journal title, volume, and issue. Further information about journal titles is also available by way of the recently implemented links to the ZDB (Zeitschriftendatenbank).[97]

While the possibility to browse from an article to the journal (issue) and vice versa might primarily be a matter of convenience for users, the impossibility of establishing relationships between different versions of an object has implications for long-term preservation in that it might limit or prevent the effective use of the preserved object in the future. Currently, if a new version of an object is ingested into the repository, there is no way of linking these two versions, or of indicating conveniently that one replaces the other.[98] However, work is currently carried out by the pedocs team to implement versioning. Thus, it is planned to add a metadata field to records into which the ID of a related document record can be entered. In addition, a field will exist in which the relation can be further specified by means of Dublin Core (dc.relation and qualifiers).

Example: The need for versioning in pedocs

The pedocs record with the ID 807[99], which was already cited above, has two documents attached to it; the second of these (Dokument 2.pdf) provides context and background information about the first, which is the main document containing a study/synopsis of German elementary school curricula in the subject “Sachkunde” between 1992 and 2009. In these preliminary remarks the author explicitly asks users to create updated versions of the main document as the curricula change:

> Sollte festgestellt werden, dass Lehrpläne einzelner Länder nicht mehr aktuell sind, wird hier ausdrücklich darum gebeten, aktuelle Daten individuell einzuarbeiten und die Synopse in passender Weise zu ergänzen bzw. zu überarbeiten. Das dabei neu entstandene Dokument sollte dann […] im Open Access Netzwerk als folgende Version eingestellt werden, mit einem Hinweis auf die Aktualisierung […]. Da sich die Datenbank Pedocs bereit gestellt hat, alte und neue Version parallel zu veröffentlichen, wird dann möglich sein, die Veränderungen, die mit den neuen Lehrplänen einhergehen, nach zu verfolgen […]. (Efler-Mikat 2009, Dokument 2.pdf, II)

While currently it is difficult (if not impossible) for future users to conveniently find either later or

97 http://dispatch.opac.d-nb.de/LNG=DU/DB=1.1/ – 31.10.2009.

98 The question of what constitutes a new version of an already submitted (and, for example, hence warrants the creation of a new record with an URN assigned to it) is by no means easily answered. See Brace's “Versioning in Repositories” (2008) and the homepage of the Version Identification Framework (VIF) project (http://www.lse.ac.uk/library/vif/) for considerations about versions and version control in repositories (see in particular http://www.lse.ac.uk/library/vif/Framework/SoftwareDevelopment/linking.html for the issue of linking records and how it is addressed in DSpace, Fedora, and EPrints) – 03.11.2009. The question of versions and editions of AIPs is also raised in the OAIS version 2 candidate, which now contains definitions of AIP Editions and AIP Versions (see OAIS 2009, 1-7).

99 http://www.pedocs.de/volltexte/2009/807/ – 24.10.2009.

previous versions of this document – especially, as different versions might not be published by the same author – after the implementation of the changes described above, links will exist between these documents which will facilitate navigation between versions.

Additional context information, e.g. why the content object was produced, might be contained in the abstract in some cases but is not captured otherwise. In that such context information can, in most cases, only be provided by the original author of a resource, there is no opportunity to add it at any later point (e.g. by the DNB).

Rights metadata: All records contain a link to copyright information[100], which is generally the same for all documents in the repository if not stated otherwise. The copyright information refers users to German copyright law (Urheberrechtsgesetz), which among others allows the printing and saving of documents for private or educational use. Additional rights metadata will only be added in the rare case of exceptions.

As mentioned above, pedocs uses JHOVE to generate a small set of **technical metadata** based on format identification and validation. Thus, file format, format version and checksum will be submitted to the DNB. At the DNB, a fuller set of technical and other preservation metadata will be generated and saved. Thus, the DNB has defined a set of metadata relevant to the preservation of digital objects (LMER – Long-term preservation Metadata for Electronic Resources) and will use this metadata in the preservation of the pedocs collection. Capable of being integrated into METS and working with links to format registries, the LMER metadata set comprises the following elements:

> *lmerObject*: This includes metadata that corporately refer to all sub-files of the document. It also comprises the URN as a Persistent Identifier to establish a unique reference to the bibliographic metadata.
> *lmerProcess*: Within these metadata, all technical changes to the object or to any file of the object are recorded [...].
> *lmerFile*: For each file that belongs to the object, metadata describing its characteristics are stated here. These metadata are composed of general fields that are common to all file types, and of specific fields (e.g., the frame rate for videos) […].
> *Metadata Modification*: Within these metadata, all changes to the LMER metadata themselves are recorded. (Die Deutsche Bibliothek 2005, 8)

Where file type-specific metadata are required, these are added in lmerFile by mean of a special field (xmlData), which “has been assigned to assimilate any XML metadata in another schema,” a “modular approach […] made possible by the namespace concept of XML [...]” (Die Deutsche Bibliothek 2005, 5).

Fixity information for the digital object(s) contained in the (pre-)AIPs produced by pedocs is generated in the form of checksums, which will be discussed below (see Archival Storage).

Ingest: Generate Descriptive Info

pedocs currently supports three document types: Articles in a journal, essays in a collection, and monographs. The metadata for the different document types consists of a

100http://www.pedocs.de/doku/urheberrecht.php?la=en – 03.11.2009.

set of shared metadata which is the same for all types, and document type-specific metadata recorded in the case that the document to be submitted already has been published elsewhere. The content of all ingested documents is described by means of a controlled vocabulary. Authors can search for appropriate subject headings by means of an implemented search tool and add them to the submission form.[101] Although authors are also given opportunity to add free descriptive keywords, these are usually replaced by descriptors/subject headings from the controlled vocabulary by pedocs staff. Particularly where abstracts are included, the descriptive information available appears sufficient to locate and analyze (see OAIS 2002, 4-30) digital objects stored in the repository.

Via the OAI interface, pedocs metadata can be mapped onto the Dublin Core Element Set, of which eight elements are currently used by pedocs. In addition, documents are classified by means of DDC; however, since only the DDC main classes are used, all documents receive the same notation, i.e. 370. This is not displayed in the records and can – obviously – not be used for the purpose of browsing.

2.2.2 JUWEL Ingest

Ingest: Receive Submission

Digital objects reach the repository via the author submission form (Web Submission User Interface [UI]) or by email to the repository staff, who then ingest the submitted files into the repository. In addition, campus press publications are ingested into JUWEL in an automated batch ingest process[102] which can either be initiated by an administrator or be scheduled to be carried out automatically as a so-called cron job at designated times.

Currently, most publications stored in JUWEL are in PDF format, although sometimes the “original” from which the PDF created is archived along with the PDF. As of yet no policy exists which file formats can and will be preserved (i.e. curated) in JUWEL and/or the long term archive with which it might cooperate one day. However, according to Wagner it seems unlikely that proprietary formats such as .doc will be included in preservation efforts beyond bitstream preservation if at all. In contrast, the use and support of open formats such as LaTeX seems much more likely.

While the “JUWEL-Policy”[103] states that authors submitting material to the repository have to give permission that their works can be stored, copied, transformed into other formats, and made accessible via JUWEL[104], the policy does not specify that documents

101 http://www.pedocs.de/schlagwortsuche.php?set=01&status=call – 03.11.2009. The subject terms offered are an excerpt from the controlled vocabulary used for the FIS Bildung database.

102 In addition to the Web Submission User Interface (UI), a “batch item importer” is available in DSpace, “which turns an external SIP (an XML metadata document with some content files) into an 'in progress submission' object” (Tansely et al. 2006, 16). Initiating such a batch ingest is not possible for users, who are thus required to use the Web Submission UI in all cases.

103 http://juwel.fz-juelich.de:8080/dspace/help/policy.jsp – 03.11.2009. Hereafter cited as JUWEL Policy.

104 Note that the policy contains the deposit agreement rather than more general policy information. “Mit der Annahme dieser Bestimmung wird der Zentralbibliothek das nicht-ausschließliche Recht eingeräumt, die Ressource zu speichern, zu vervielfältigen, weltweit zugänglich zu machen und bei Bedarf gedruckte und elektronische Kopien anzufertigen […]. Mit der Annahme dieser Bestimmung räumt der Veröffentlichende der Zentralbibliothek das Recht ein, die Ressource bei Bedarf (z. B. Migration, Barrierefreiheit, bessere Zugänglichkeit, Erschließung) in andere elektronische und physische Formate zu überführen und diese

submitted must be free from DRM or other security or protection measures (e.g. password protection, etc.). Thus, for example, the record accessible at http://hdl.handle.net/2128/3140 (25.10.2009) links to an encrypted campus press publication protected against changes by password. As already explained above, this is highly problematic if the document is to be preserved in a long term archive. Thus, it seems advisable not only to modify the policy accordingly, but also to take measures to ascertain that the repository indeed has complete control over these digital objects as otherwise long-term preservation might not be possible or considerably more difficult to realized for these.

Ingest: Quality Assurance

Integrity: DSpace versions up to 1.5 carry out a format identification which, however, is merely based on the filename extension which is checked against a (locally installed) format registry.[105] Thus, as Larry Stone, commenting on problems with the current method, explains in a paper delivered at the Open Repositories Conference 2008,

> [t]he data model for technical metadata (the BitstreamFormat object) has been essentially unchanged from the inception of DSpace through release 1.5. Although its design anticipated the use of external format registries, it was never completed. There are some other problems with the current data model and implementation:
>
> 1. Formats are identified by arbitrarily-assigned descriptive names such as “Adobe PDF” which have no meaning outside of DSpace. This impedes any attempt to share format descriptions with any other application.
> 2. There is no provision for collecting more extensive format technical metadata, such as standards documents, that would help future preservationists interpret obsolete formats.
> 3. A Bitstream's format is only identified by comparing its filename extension to entries in the format registry. This method is prone to errors, ambiguous results, and outright failures [...]. (2008, 2)

Thus, currently neither a thorough format identification nor a format validation, which would have to be performed by an external tool such as JHOVE, are carried out for submissions to JUWEL. In consequence, the question of whether a file is functional or not is at the moment merely answered by means of opening it, which clearly is no sufficient procedure if digital objects are to be preserved for the long term.

Additionally, the DSpace submission workflow gives authors the opportunity to download their publications immediately after upload in order to verify that the MD5 checksum generated during the ingest process is the same as the one of the original

---

gemäß Absatz 1 zu verwerten” (Juwel Policy).

105 Plans exist to implement an automatic format identification service into DSpace making use of external format registries as well as a new data model for technical metadata in a future DSpace release. These efforts are explicitly related to considerations concerning digital preservation. As Stone explains, “[t]o address digital preservation problems and take advantage of tools currently being developed, DSpace needs more fine-grained format classification, as well as globally-recognized identifiers for formats […]. Both of these are provided by a preservation-minded external format registry such as PRONOM or GDFR” (Stone 2008, 5.2). In addition, the importance of format validation has been recognized, and according to Stone the new data model “is ready to record validation; successfully validated Bitstreams are marked with a Confidence value of 'VALIDATED'” (Stone 2008, 5.2).

As outlined in the DSpace wiki, future projects will address the following preservation applications: “Detecting and notifying administrators of obsolete formats in the archive,” “Format migration and normalization (migration on ingest),” and “Data format validation (some of which is already implemented in pending JHOVE integration work” (DSpace Wiki 2008a).

document, and that hence the document's bitstream is unchanged (note that this is an indicator for both the authenticity and the integrity of the digital object).

Finally, the completeness of SIPs is also ascertained by means of intellectual control by the repository staff who check, edit, and add metadata as part of the DSpace workflow. This editing process becomes necessary in particular because JUWEL will not reject submissions due to incomplete metadata. However, while it would certainly be possible to implement such a "rejection function," this might keep potential depositors from submitting publications. Hence to edit metadata after submission is the more "user-friendly" procedure. As depositors log in before submitting, even in the case of incomplete metadata, submissions can always be linked to a producer.

As outlined above, integrity can also be endangered by viruses. Files incoming by email are automatically scanned for malware; similarly, files submitted on CD ROM or other storage devices will be scanned automatically by the respective computer. Since publications are uploaded only from within the Forschungszentrum's LAN, the computers from which the upload happens are equally protected, and as all in all the Forschungszentrum maintains a very high IT-security standard, the danger posed by viruses/malware is very low.

Authenticity: DSpace offers customizable authorization and authentication functions. According to the DSpace 1.4.1 system documentation, "[a]uthentication is when an application session positively identifies itself as belonging to an E-Person and/or Group. In DSpace 1.4 it is implemented by a […] 'stack' of authentication methods" (Tansley et al. 2006, 15).[106] The JUWEL authentication procedure requires registration by email upon which an account is created (users submit their name, first name, e-mail address, and phone number in order to receive an account).

All communication taking place within JUWEL is encrypted with the help of a DFN (Deutsches Forschungsnetz – German Research Network) certificate so that all submitted information – e.g. passwords – is protected.[107] Thus, file upload, which takes place in the "Mein JUWEL" (My DSpace) area, to which depositors have access after the log in procedure, too, is protected by means of a secure connection. Accordingly, the batch ingest process uses the HTTPS protocol. Users submitting files to the repository staff via email can be identified by means of their email addresses and are not required to have a JUWEL account. It follows that the sources of the digital objects stored in JUWEL are known with a very high degree of certainty, clearly an advantage of an institutional repository with a community of producers/depositors to which access is very restrictive and which can hence be unambiguously identified at all times.

106 No published documentation for DSpace 1.4.2 was found to be available. The minimal differences between versions 1.4.1 and 1.4.2 are documented in the Change History section of subsequent documentations.

107 The DFN provides a public key infrastructure making use of advanced certificates based on the X.509 standard. The service is primarily directed at universities and other research institutions who want to use secure communication. See http://www.pki.dfn.de/ for further details (in German) – 25.10.2009.

As explained in the DSpace Documentation, “DSpace's authorization system is based on associating actions with objects and the lists of EPeople who can perform them” (Tansley et al. 2006, 15). Thus files can only be uploaded by authorized and authenticated users. The same is true for other possible actions in DSpace, i.e. READ, ADD, REMOVE, WRITE (ibid. 15-16), and it follows that an unauthorized modification of collections, items, or bitstreams – which would affect the authenticity and possibly integrity of documents – is prevented by means of the resource policies.

Ingest: Generate AIP

According to the DSpace data model such information packages – called items – are composed of one or more so-called bundles which are themselves composed of one or more bitstreams (also called data files). Each bitstream has a bitstream format, name, size, a checksum, and a bitstream ID as well as an (optional) user description. Each item additionally has metadata in Dublin Core qualified and a handle. Typically, one item is described by one metadata record.

DSpace can be set to automatically create persistent identifiers for every item ingested into the repository as well as for communities and collections. Thus,

> DSpace uses Handles primarily as a means of assigning globally unique identifiers to objects. Each site running DSpace needs to obtain a Handle 'prefix' from CNRI […]. Since it's usually the item that is being preserved, rather than the particular bit encoding, it only makes sense to persistently identify and allow access to the item, and allow users to access the appropriate bit encoding from there. (Tansley et al. 2006, 18)

The Handle prefix of JUWEL is 2128 so that a typical address for a resource stored in JUWEL will look as follows: http://hdl.handle.net/2128/[Local Name].[108] The identifiers are placed prominently on the top of each item record and marked as the preferred link for citation of the record and the document(s) it identifies. Any unique or persistent identifiers assigned to the document before it was submitted to the repository (e.g. ISBN or ISSN numbers, but also other [persistent] URIs such as DOIs or URNs) can be entered into the field dc.identifier and are hence preserved along with the other metadata. In following this practice, JUWEL complies with the respective TRAC criterion.[109]

In addition, DSpace also assigns identifiers (Sequence IDs) to each bitstream. However, these are unique only within a local DSpace installation so that collisions could occur in a DSpace to DSpace use case.[110]

---

108 The Local Name is assigned by the Handle Naming Authority and in the case of JUWEL consists of an object identification number unique within JUWEL. See pages 63-66 of the Paradigm Workbook for an introduction to the Handle System. Further information can also be obtained from http://www.handle.net/ – 25.10.2009.

109 According to the Dublin Core element description, identifiers entered into this field provide “unambiguous reference to the resource within a given context. Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Examples of formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN)” (DCMI 2005b). See http://hdl.handle.net/2128/1610 for an example of a JUWEL records with more than one persistent identifier – 25.10.2009.

110 See DSpace Wiki 2008a. See also the DSpace System Documentation on “Bitstream 'Persistent' Identifiers” (Tansley et al. 2006, 19).

Structural metadata: In DSpace it is possible to express relations between bitstreams belonging to one item by “bundling” them together and by allowing one of the bitstreams in the bundle to be designated as the “primary” or “preferred” one. The latter option can be used in case different versions of a document exist – e.g. in PDF and LaTeX format. The former option is used particularly for HTML files, which might, for example, come with image files embedded in the HTML page (see DSpace Wiki 2008a). In this manner, an HTML entrance or index page can, for example, be distinguished from image files belonging to it. However, it is obvious that this method will quickly cease to be feasible if the webpage to be archived is of even average complexity (see also below). All in all, as stated in the DSpace System Documentation, “[s]tructural metadata in DSpace is currently fairly basic [...]. Additional structural metadata can be stored in serialized bitstreams, but DSpace does not currently understand this natively” (Tansley et al. 2006, 13).

On the item level, in DSpace relations between items can primarily be expressed by means of the qualifiers available for the Dublin Core element “Relation.”[111] JUWEL uses this option primarily for series titles, which are identified as part of a series by means of the field dc.relation.ispartofseries.[112] In contrast, the relation between journal articles and the issue/journal of which they form part is expressed in textual form (string) in the fields dc.identifier.citation and dc.identifier.issn. While this is the preferred form of expressing the relation between a document and its “host” according to the DC-Library Application Profile[113] adhered to by DSpace, Dr. Wagner rightly questioned the adequateness of this method as it strongly limits the possible uses to which the data can be put.

Although the majority of publications ingested into JUWEL consist of a single PDF document, as mentioned above, the repository does contain items consisting of more than one bundle/bitstream. Depositors submitting more than one file are asked to enter a brief description of each file[114] – however, browsing the repository no example was found where such a description was actually given. For example, the record which can be retrieved at http://hdl.handle.net/2128/461 (25.10.2009) contains download links for two documents, a PDF and a ZIP file (see below for a brief discussion of problems associated with container formats). No description was entered for either file so that it is unclear which might be the preferred version, or how the two versions differ.[115]

In order to link two items which are different versions of the same (conceptual) object, Dublin Core qualified elements can be used to indicate which version is the most recent, and which document succeeds which. While it would therefore be possible to express versioning by means of Dublin Core, this is not implemented in JUWEL because works

111 See DCMI 2004 for applicable qualifiers.
112 See, for example, http://hdl.handle.net/2128/3603 – 25.10.2009.
113 Thus, the use of dc.relation with qualifiers (e.g. ispartof or haspart) is recommended only “[w]hen documents in hand are parts of 'host documents' (e.g. journal, monographic series) and when there is no citation information in DC identifier” (DCMI 2004).
114 See http://juwel.fz-juelich.de:8080/dspace/help/help_en.jsp – 03.11.2009. Hereafter cited as JUWEL Help.
115 Thus, merely the file names Bibliothek_13_Buch.pdf and Bibliothek_13_CD.zip hint that the second file might contain files belonging to a CD included with the print version of the book.

submitted to the repository are – in terms of their content – considered as final versions. As the DC relation qualifier “isversionof” is used only if “[t]he described resource is a version, edition, or adaptation of the referenced resource,” and if the changes are “substantive changes in content rather than differences in format” (DCMI 2005a) the use of the respective qualifiers to dc.relation (i.e. isversionof, hasversion, isreplacedby, Replaces) is indeed not required.

Another way of creating relations between digital content objects in DSpace is the hierarchical structure of communities, sub-communities and collections it uses. While this does not express structural aspects so much, it does transport certain information regarding the **context** of a publication. JUWEL currently contains only one community and primarily orders digital objects into collections which represent the institutes or organizational units of the Forschungszentrum by (or in cooperation with) whom they were published (cf. JUWEL Help). According to Dr. Wagner, work is being carried out to reorganize the community and collection structure to improve access to and contextualization of publications. Except for information sometimes recorded in abstracts, this is the only (implicit) context information users receive about the published documents.

DSpace automatically records certain **technical metadata**. The information recorded was, according to the DSpace wiki, selected to aid “preservation and administration” of items in the repository and includes “unique identifier, checksum, checksum type, mimetype, file size, creation date and file path originally assigned to the file” (DSpace Wiki 2008b), not, however, file format and version. As briefly mentioned above, DSpace identifies the mime type of a given digital object and records it in the metadata. However, the JUWEL installation frequently records and displays two mime types for PDF files, namely “application/pdf” on the one and “text/plain” on the other hand.[116] This is possibly a result of recording the mime type of the license text for the respective object. Although this is not highly problematic it might lead to confusion or misunderstandings on part of the users as the metadata record becomes ambiguous in this respect.

Problems with container formats
As already mentioned briefly, browsing the repository showed that while it contains mostly PDF files, in some cases a container format has been used to publish items which have several bundles and/or bitstreams associated with them. For example, the record which can be retrieved at http://hdl.handle.net/2128/466 (25.10.2009) allows users to download the respective publication – the IFF Scientific Report 00-01, which is published in form of a website – as a ZIP file. No description for the ZIP file was entered, and it is entirely unclear how the conceptual object it contains is structured and how it is to be “reconstructed” by means of the files compressed in the folder. Moreover, the record lacks not only structural metadata but also technical metadata for the files contained in the ZIP folder. In fact, not even the ZIP format was identified by the system correctly, as the mime type is given with application/octet-stream, which is the designation for unknown formats. While it is certainly not desirable or feasible to archive complex websites without containers in the repository, from the perspective of long-term preservation, the ZIP folders saved on the JUWEL server are “black boxes” as far as structure, formats, and versions of the files they contain are concerned.[117]

116 See, for example, http://hdl.handle.net/2128/3226 – 25.10.2009.

Rights metadata: Dublin Core allows the recording of rights metadata in the field dc.rights. Although JUWEL makes use of this possibility, it does not do so consistently. Thus, records exist which do not contain any rights information except for the disclaimer at the bottom of each record (see, for example, http://hdl.handle.net/2128/3585 – 25.10.2009).[118] In other cases, the field dc.rights was used (e.g. http://hdl.handle.net/2128/3229; note that the rights information is only visible in the full item display ["Langanzeige"]). Additional rights information – concerning, however, not the ways in which the publication can be used by repository users, but the rights the author had to publish the document in the repository – is given in the field dc.description. The rights information given in this field is taken from the SHERPA/RoMEO database on publishers' copyright and self-archiving policies and will – where possible and applicable – be collected automatically in the future by means of a script which will extract the information from the database.[119] It seems that dc.description is not the best place to store this information, as the field is reserved for a description of the content of the resource.[120]

Provenance metadata: DSpace automatically documents certain steps of the submission and ingest process by means of so-called provenance messages (see Tansley et al. 2006, 16-17). Thus, for example,

> [w]hen the Batch Ingester or Web Submit UI completes the InProgressSubmission object, and invokes the next stage of ingest (be that workflow or item installation), a provenance message is added to the Dublin Core which includes the filenames and checksums of the content of the submission. Likewise, each time a workflow changes state (e.g. a reviewer accepts the submission), a similar provenance statement is added. This allows us to track how the item has changes since a user submitted it. (Tansley et al. 2006, 16)

A further provenance message will be added once the ingest process is concluded by the so-called Item Installer, which adds, among others, a bitstream checksum.[121] The provenance information just mentioned is stored in the field dc.description.provenance which is hidden from OAI-harvesters because it also contains "private information about the submitter and workflow reviewers of the item, including their e-mail addresses" (Tansley et al. 2006, 119). However, due to a bug the provenance information is displayed in some DSpace installations (not JUWEL), and can look as follows:

dc.description.provenance	Submitted by Peter Maher (pmaher@mit.edu) on 2008-01-11T18:23:31Z No. of bitstreams: 1 4672-07.pdf: 235959 bytes,

117 For example, the ZIP file which can be retrieved at http://hdl.handle.net/2128/461 (see example above) alone contains HTML, JPEG, GIF, PDF, and SWF files as well as installation files for Adobe Acrobat Reader (EXE). See http://hdl.handle.net/2128/559 for another example of a ZIP file – 25.10.2009.

118 The following static copyright statement appears at the bottom of each item record (simple and full display): "Alle Dokumente und Ihre [sic] Metadaten sind urheberrechtlich geschützt."

119 http://www.sherpa.ac.uk/romeo/. See http://www.sherpa.ac.uk/romeo/api.html for information on the API which makes the extraction of data possible – 03.11.2009. A JUWEL record which uses dc.description for SHERPA/RoMEO information can be retrieved from http://hdl.handle.net/2128/3220 – 25.10.2009.

120 See DCMI 2005b. According to Dr. Wagner it is possible that the cases in which dc.description is used for rights information result from an import of records created before dc.rights existed into JUWEL.

121 Certain relevant data about provenance is also captured and stored by the History System still available in DSpace 1.4. The "History subsystem is explicitly invoked when significant events occur (e.g., DSpace accepts an item into the archive). The History subsystem then creates RDF data describing the current state of the object" (DSpace Manual, 21). However, the History system has been removed from later DSpace releases in the meantime because it was not working properly.

	checksum: 837d19b699bdba3d1d9a5fef34e604e4 (MD5)
dc.description.provenance	Approved for entry into archive by Peter Maher(pmaher@mit.edu) on 2008-01-11T18:24:09Z (GMT) No. of bitstreams: 1 4672-07.pdf: 235959 bytes, checksum: 837d19b699bdba3d1d9a5fef34e604e4 (MD5)
dc.description.provenance	Made available in DSpace on 2008-01-11T18:24:10Z (GMT). No. of bitstreams: 1 4672-07.pdf: 235959 bytes, checksum: 837d19b699bdba3d1d9a5fef34e604e4 (MD5)[122]

As observed by the NLM working group which tested and evaluated the DSpace software (see NLM 2009), the software failed to completely fulfill the (provenance-related) requirement P2-16 ("Demonstrate the creation of an audit trail of all actions including who, when, how, what and where for Archive Storage and Data Management") in that "[n]o provenance for record update and no email/screen conformation for delete/withdrawal" existed in the respective log file (NLM-DRESWG 2009, Appendix C, P2-16). In addition, note 10.7 of Appendix C explains that

> DSpace tracks every action as an entry in a text-based log file but the log file doesn't reveal the specific action taken. The History file, on the other hand, records more specific details than the log file but some of the entries don't seem related to the actual action. For example, adding a subject and modifying an item type are recorded as updated bit streams and updated bundles in the history file. No tools are provided to access the history file. (ibid., 10.7)

These errors were very likely among the reasons why the History System was disabled in later DSpace releases (see above).

Further provenance information could be recorded in the field dc.provenance, dc.which can contain statements "of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity and interpretation. The statement may include a description of any changes successive custodians made to the resource" (DCMI 2005b). This field is, however, not currently used in JUWEL.

As already mentioned, **fixity information** in the form of MD5 checksums is created. These will be discussed in further detail below (Archival Storage).

Ingest: Generate Descriptive Info

During the submission process via the Web Submission UI, users enter relevant descriptive metadata about the publication they want to submit (see JUWEL Help document for a detailed description of the different steps and the information collected). In addition, where available and applicable, descriptive metadata is also imported into JUWEL out of the publication database VDB and the bibliographic database of the campus press in MARC format (data exchange via OAI interface). Although the publications stored in JUWEL therefore have a minimum of descriptive metadata (title, author/editor, publisher information among others), in the majority of cases resources are neither assigned subject headings nor classification information. In very few cases, objects receive a DDC notation indicating the main class which, however, is too broad to

122 http://dspace.mit.edu/handle/1721.1/40086?show=full (full item view for http://hdl.handle.net/1721.1/40086) – 18.10.2009.

offer more than a very general orientation (e.g. 600 Technology; see long item display for http://hdl.handle.net/2128/3220 – 25.10.2009). In addition, some (but not the majority of) records contains abstracts to support the discovery and selection of resources. This lack of subject indexing, which also means that JUWEL does not comply with the respective DINI criteria at the moment, is problematic for a number of reasons: on the one hand it makes the discovery and selection of documents increasingly difficult, as users either have to trust that they find what they need by means of a simple keyword search or by browsing the collection of the institute which they believe to have published something on the topic they are interested in. As pointed out by Dr. Wagner, while this might be tolerable as long as one stays within the boundaries of JUWEL, it becomes an immense drawback as soon as users access the collection not via the local search functionality but search for open access publications via meta-search engines such as BASE.[123] It is particularly in these search engines that a subject-based search makes much more sense than a full-text search, which will often return too many hits to be useful. Thus, as soon as users carry out such subject searches in BASE or similar engines, no hits from JUWEL will be returned. In addition, from the perspective of long-term preservation, the lack of metadata describing the content of a resource by means of a controlled vocabulary is a considerable deficit in that such metadata will help to ensure that the designated community will be able to understand the digital objects in the repository. Thus, in particular as the amount of publications stored in the repository increases, the use of classification and controlled vocabulary for indexing will become indispensable.

2.2.3 Qucosa Ingest

Ingest: Receive Submission

Digital publications are primarily submitted to Qucosa through the web interface ("Eingabeassistent")[124], which allows authors to upload their publications and related files in six steps. In some cases files are also submitted via email and uploaded by the repository staff. In addition, submission of CD-ROMs and URLs is possible as well (see Qucosa FAQ[125]).

Both the Qucosa FAQ and the Qucosa Publication Guidelines[126] outline requirements on the submitted files meant to guarantee that the repository receives full control over the submitted files. Thus, according to the publication guidelines, submitted files must among others fulfill the following criteria: "Die Dateien sollen identifikationsfrei sein und keinen Sicherheitsbeschränkungen unterliegen." This aspect is explained in more detail in the Frequently Asked Questions, which state that files must be free from any DRM or other technical measures designed to limit access to the file or protect the file from being copied (password protection, encryption, etc.):

123 http://www.base-search.net/ – 03.11.2009.
124 http://www.qucosa.de/veroeffentlichen/eingabeassistent/ – 18.10.2009.
125 http://www.qucosa.de/faq/ – 03.11.2009. Hereafter cited as Qucosa FAQ.
126 http://www.qucosa.de/veroeffentlichen/ – 03.11.2009. Hereafter Cited as Qucosa Guidelines.

> Für die Langzeitverfügbarkeit der auf Qucosa publizierten Dokumente ist die Abgabe von ungeschützten Dokumenten notwendig, d. h. Dokumente dürfen weder verschlüsselt noch mit Kennwortschutz versehen eingestellt werden und müssen Ausdrucke und Vervielfältigungen zulassen. Denn wenn eine Datenmigration erforderlich ist, stellen zugriffs- oder passwortgeschützte Dokumente eine Hürde für die verarbeitende Software dar und machen eine Langzeitarchivierung unmöglich. (Qucosa FAQ)

In addition, the FAQ suggest the submission of files in PDF/A format. As authors might not read the publication guidelines or the FAQ in all cases, it seems advisable to include information on these issues in the author agreement.[127]

Thus, if authors follow the guidelines and FAQ, Qucosa obtains full physical/technical control over the digital objects as required by nestor and TRAC. However, similarly to JUWEL, submitted files are currently not being checked for the presence of DRM measures so that it is not entirely certain that the files ingested into the repository are indeed free from any technical limitations.

While technical control over the submitted files is crucial in order to apply long-term preservation strategies, in order to carry out preservation measures the repository also needs to have the right to copy and alter the submitted files (i.e. change their file format). Accordingly, via the author agreement the SLUB receives

> the non-exclusive (simple) right, to copy, to permanently store and to publish these on the internet [...]. SLUB Dresden is allowed to rename and convert these publication(s) into other data formats, if technical developments would require it. No other content changes will be made during conversion.
>
> SLUB Dresden is allowed to transfer the publication(s) as electronic files, including all metadata, to the German National Library, the DFG Special Interest Collection Library, and if applicable to other libraries and archives. With the transmission, the aforementioned rights are transferred to these institutions. (Qucosa author agreement)

Authors are required to print and sign the author agreement, and send it to the SLUB Dresden before the submission is published. In addition, at the end of the submission process via the web submission form, depositors need to accept the following agreement:

> Hiermit übertrage ich der SLUB Dresden das Recht, das übermittelte Dokument elektronisch zu speichern und in Datennetzen öffentlich zugänglich zu machen. Ich versichere, dass mit einer derartigen Veröffentlichung keine Rechte Dritter verletzt werden. Meine Rechte zur Verwertung in einer anderen körperlichen Form bleiben davon unberührt. (Qucosa Eingabeassistent)

As suggested above, both agreements should contain a statement making it a submission requirement that the submitted files are free from DRM or other technical protection measures and/or a note that long-term preservation cannot be guaranteed if depositors do not adhere to this rule.

Ingest: Quality Assurance

Integrity: Currently, only a virus scan is carried out to check the integrity of the submitted files. Once, koLibRI is put into use to produce SIPs for the long term archive, the software will carry out format identification, characterization and validation with the

127 http://www.qucosa.de/fileadmin/groups/qucosa/PDF/Formular_Archivierungs-_und_Nutzungsrechte.pdf; English version available at http://www.qucosa.de/fileadmin/groups/qucosa/PDF/Formular_Archivierungs-_und_Nutzungsrechte_englisch.pdf – 18.10.2009.

help of JHOVE. In this context, the question remains at which point the SIPs will be created and at which point format characterization and validation will hence be carried out. A possible scenario according to Dr. Kluge is that koLibRI will create SIPs twice a month. In establishing this routine it will be crucial to consider how files submitted to Qucosa will be treated in the time between submission by the depositors and their processing by koLibRI. For example, it has to be considered whether it is advisable to make these files publicly accessible through Qucosa before their integrity was checked by JHOVE. Thus, if an error is detected by JHOVE, it might be easier to correct it – e.g. by replacing the file – if it is still unpublished. Any other procedure might lead to conflicts with the DNB's URN policy, according to which a new URN has to be assigned if, for example, the MD5 checksum of a publication can be shown to have changed (see above).

As with pedocs and JUWEL, the “completeness and correctness” of the submitted files is additionally ascertained by means of intellectual control, especially with regard to metadata and the author agreement – thus, no publication takes place until the signed agreement has been received, and Qucosa staff will correct or add metadata before the digital object is added to the repository. As authors are requested to submit a certain amount of descriptive metadata (e.g. keywords/subject terms) it might be helpful to advise them – e.g. in the FAQ or publication guidelines – that their submitted metadata will be subject to quality control and may be modified in the process.

Authenticity: As authors are not required to identify themselves by creating an account, they confirm that they indeed have the right to publish the work in question, and that hence they are who they claim to be (i.e. the author of a publication) only through the author agreement. As mentioned above, this author agreement has to be printed and signed, and hence exists in analog form. The procedure is therefore different from the one chosen by pedocs, which works with digital agreements (see above).

Qucosa does not use secure protocols at the moment so that any information entered into the web submission form is submitted in unencrypted form.

Ingest: Generate AIP

As in the other two repositories considered here, (pre-)AIPs are created in Qucosa, which will later form the basis for the koLibRI SIPs and the AIPs for long-term archiving. These pre-stage AIPs are saved in the Qucosa file system. The information packages currently contain a persistent identifier (URN) and a set of descriptive metadata. All in all, the assignment and use of URNs in Qucosa is quite similar to their use in pedocs. Thus, URNs are assigned to the record for the digital object (the so-called “front door”); the digital object(s) described and referenced by the record is/are addressable via the URL included in the record. URNs and URLs are submitted to the DNB, where any changes to the URL are registered as well.[128] The URN is displayed as preferred form of citation in the

128 A considerable number of the URNs are at the moment not resolved and result in a DNB error message (see, for example, http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-24790 – 18.10.2009). According to Kluge, the problem lies with the DNB, where an interruption of the technical service provided occurred

record for a title; although a workflow exists during which URNs can be integrated into the PDF-document (see below), this workflow is not frequently used at the moment so that once the publication has been saved or printed by a user, metadata and digital object are separated. Like pedocs, Qucosa does not record previously assigned persistent identifiers.

Structural metadata: In contrast to the output created by the digitization center at the SLUB Dresden, the digital objects stored in Qucosa are generally not of a highly complex nature and hence do not require extensive structural metadata. However, as the Qucosa policy document states, not only “traditional text documents” (which themselves, it should be noted, may very well have a complex structure), but also multimedia objects consisting of sound, image, film, or computer animated elements can now be published in Qucosa: “Neben traditionellen Texten können inzwischen auch Multimediadokumente mit Ton, Bild, Film, Computeranimation etc. publiziert werden.”[129] Thus it is likely that in the future more complex digital objects will be published in the repository, and that structural metadata will be required to capture the logical structure of these objects and to make this structure understandable for current and (ideally) future users. Already at this point, Qucosa is used to store articles from the SLUB Kurier, a magazine published by the SLUB containing information about recent developments and activities in/of the library. While this certainly is not a highly complex digital object, and most articles will be understandable taken by themselves, the use of structural or context metadata to identify articles belonging to the same issue and the sequence of articles in the issue would certainly help to “enabl[e] more efficient navigation of resources” (DPIP 2008). Currently all of this information is available – somewhat implicitly – in the field “source” (“Quelle”), which includes information about the magazine issue and article page numbers. However, as this field cannot be searched individually and can hence only be addressed through the full text search, this information is not available for the purpose of navigation immediately and conveniently.

As already outlined in the discussion of pedocs and JUWEL, even if the structure of journal or magazine issues is maybe not as relevant in the context of long-term preservation as other information (e.g. descriptive metadata), to be able to express the relation between different versions or editions of a digital object is crucial. Although versioning is according to Kluge not relevant for the materials ingested into Qucosa at the moment, it might become an issue in the future. Thus, as the policy requires that changed objects are ingested into the repository as new documents, which receive their own URN, the existence of different – e.g. corrected – versions of a publication is very likely in the future. As discussed above it will be crucial to identify previous and current versions of a document, for example, by indicating that object B replaces object A. Otherwise, the relation between similar but not identical objects in the repository might not be

recently.

129 http://www.qucosa.de/ueber-qucosa/ – 03.11.2009. Hereafter cited as Qucosa Policy.

reconstructable in the future. It will be impossible or at least difficult for future users to judge, for example, which of several versions of an object stored in the repository is the "authoritative" one – future users will only see that similar but not identical objects exist, but they will not be able to tell for sure whether one of them replaces another.

As in pedocs and JUWEL, in Qucosa, too, it is possible to attach more than one file to a record; as in the other repositories, the relationship between these files is not explicitly specified. Thus, for example, the record which can be retrieved at http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-24013 (a three-part musical score by Manfred Weiss)[130] contains links to several files the order of which seems to be determined by the order in which they were uploaded – in consequence, the first part of the composition is listed second. Here it might be advisable to use structural metadata to indicate file sequence, in particular as the main document is supposed to contain explanations about any additional documents uploaded.[131] Ideally, authors themselves should add a brief explanation of the relationship between the files uploaded.

If **context** information is given for the publications, it is recorded in the abstract (see for example, the abstract for the musical score just cited).

No **rights metadata** is currently included in the records themselves. Although each record contains a link to a copyright notice ("Hinweise zum Urheberrecht"), the link takes users to the repository's welcome page, which does not contain information about rights. The Qucosa imprint ("Impressum") does contain a note on copyright, but it is not entirely clear whether this is also meant to extend to the documents published through Qucosa, or only to the layout and immediate contents of the web page. Thus, a page containing rights/copyright information for these publications should be created and linked from each record.

Qucosa does not generate **technical metadata** at the moment as this will be among the tasks to be carried out by koLibRI. The Qucosa software itself only carries out a format identification based on mime type and indicates the file format in the download link. See below for a discussion of **fixity information**.

Ingest: Generate Descriptive Info

As in the other two repositories, authors are required to provide information about themselves and the submitted publication. Additional information is provided by repository staff (see table):

	Provided by author	**Added by repository staff**
Publication (formal characteristics)	Document type Language Main title Subtitle / Additional title (optional) Translation of titles (optional)	Date of publication online

130 This URN was not functional on October 25, 2009. See http://www.qucosa.de/recherche/frontdoor/?tx_slubopus4frontend[id]=2401 instead – 25.10.2009.

131 While this is according to Kluge a requirement (or at least recommendation), it is not mentioned in the policy or FAQ documents, which should be updated accordingly.

	Source (where applicable)	
Publication (content)	Abstract Index terms	DDC classification RVK classification
Author information	Last name First name Date of birth (optional)* Place of birth (optional)* Sex*	
Author institution	Place Name Department	
Additional information on theses/dissertations	Date of submission to the faculty Date of defense Name of adviser Names of reviewers	Date of publication in print

*Fields are not displayed in user view of record.

As shown in the table, Qucosa publications receive DDC and RVK notations in addition to index terms describing the content object. Compliance with Dublin Core is also given and the relevant metadata are harvested via the OAI interface.

2.3 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 fulfill orders. (OAIS 2002, 4-1-4-2)
>
> Technological responsibilities center on receiving material and maintaining the physical bits. Objects are brought into the system and stored. The system is monitored for errors, with media replaced as needed or when superseded. Access to the data and accompanying metadata is provided in response to access requests. Disaster recovery plans are planned and implemented as needed.[132]

The functions comprised in the Archival Storage functional entity primarily center on the physical preservation of bitstreams without altering the Content Information and PDI of the Archival objects (see OAIS 2002, 4-7-4-8). Thus, the Replace Media function

> may perform ‘Refreshment’, ‘Replication’, and ‘Repackaging’ that is straightforward. An example of such ‘Repackaging’ is migration to new media under a new operating system and file system, where the Content Information and PDI are independent of the file systems. (OAIS 2002, 4-7-4-8).[133]

At all times during Archival Storage, the bits stored on the media instances have to remain uncorrupted, that is, the (physical) integrity of the bitstreams must be protected, a function that is crucial also in traditional, non-preservation-centered repositories if these are to be taken seriously as reliable and trusted providers of scholarly information by the scholarly

132 http://www.icpsr.umich.edu/dpm/dpm-eng/foundation/oais/storage.html – 25.10.2009

133 In essence, refreshment implies the replacing of one “media instance” with another “instance of the same type by copying the bits on the medium used to hold AIPs […],” while a replication may involve “the same or [a] new media-type instance” and may “require changes to the Archival Storage mapping infrastructure” (OAIS 2002, 5-4). See OAIS 5-4-5-9 for further details about these different “migration types,” in particular also about “Transformations,” which “require some changes to the Content Information or PDI” and which result in new versions of the original AIPs (OAIS 2002, 5-6).

community. According to the OAIS reference model, the Error Checking function on the one hand draws on (hard- and software related) error logs and on the other depends on "PDI Fixity Information […] [for] assurance that the Content Information has not been altered as the AIP is moved and accessed. Similar information is needed to protect the PDI itself" (2002, 4-8). In addition, the bitstreams are protected by the Disaster Recovery function, which

> provides a mechanism for duplicating the digital contents of the archive collection and storing the duplicate in a physically separate facility. This function is normally accomplished by copying the archive contents to some form of removable storage media (e.g., digital linear tape, compact disc), but may also be performed via hardware transport or network data transfers. The details of disaster recovery policies are specified by Administration. (OAIS 2002, 4-8)

All three criteria catalogs considered here contain criteria relevant to the Replace Media, Error Checking and Disaster Recovery functions. The nestor and TRAC criteria catalogs primarily focus on the Replace Media and the Error Checking functional entity, although often these are subsumed under somewhat more general criteria regarding IT-infrastructure and security. In the nestor catalog, the Replace Media function is subsumed under criterion 8, dealing with the planning of technical long-term preservation measures (see also Preservation Planning below). This criterion, requiring that the "strategic plans (see 4.4) are specified on object level" (nestor 2008, 8; my translation), is fairly broad (subsuming measures for bitstream preservation, protection of authenticity as well as interpretability), and among others contains requirements concerning the physical preservation of the bitstream (uncorrupted bitstreams, refreshment of storage media, etc.). In contrast, TRAC contains a separate criterion exclusively focusing on storage media and hardware change (2007, C1.7). Although DINI does not address the question of hardware change explicitly, it requires frequent technical maintenance of the system and its components (see 2007, 2.5.1).

The Error Checking function is touched upon in the nestor catalog in criteria 6.2 and 7.2, requiring the repository to secure the integrity and authenticity of digital objects during Archival Storage on the one hand by "specifying procedures which secure the authenticity of the objects during the performance of long-term preservation measures or which alternatively record the degree of authenticity" (nestor 2008, 7.2; my translation). On the other hand, nestor requires repositories to determine the necessary degree of physical redundancy in storage (OAIS Disaster Recovery; see also nestor 2008, 10.3), the necessary quality of storage media, and to regulate logical and physical access to servers in order to secure the integrity of data objects (see nestor 2008, 6.2; see also Administration: Physical Access Control).

While the TRAC criteria do not draw a strict terminological distinction between authenticity and integrity, the commentaries to criteria B2.12 and B4.4 evoke both concepts. Thus, in particular B4.4 requires repositories to "have Fixity Information for AIPs and [to] make some use of it" (TRAC 2007, B4.4), e.g. by providing and utilizing MD5

checksums. In particular, TRAC requires repositories to demonstrate that “the Fixity Information (checksums, and the information that ties them to AIPs) are stored separately or protected separately from the AIPs themselves, so that someone who can maliciously alter an AIP would not likely be able to alter the Fixity Information as well” (TRAC 2007, B4.4).[134]

DINI recommends – but does not require – that hashsums are used to document the integrity of the archived information packages (see 2007, 2.5.2).[135] In addition, disaster preparedness in the form of scenarios (“Havarieszenarien”) is recommended in DINI 2.5.1.

2.3.1 pedocs Archival Storage

The **authenticity** of content objects contained in the pedocs (pre-)AIPs is secured primarily by means of the repository software, which prevents the archived documents from being altered. Thus, every change to an object requires the creation of a new record, i.e. a new version of the object with a new URN and ID is created automatically while the original is retained. In addition, CRC, MD5, and SHA-1 checksums are created for all digital objects stored on the pedocs server. These are included in the pedocs records, where they are visible for users. With these checksums it becomes possible to detect changes to the digital objects' bitstream and thus to monitor both integrity and authenticity of the stored files. However, no regular checks of the hashsums are currently carried out in pedocs as these are meant to be performed upon ingest into the long term archive by the DNB. This approach is problematic because users will access not the files archived by the DNB, but the ones on the pedocs server. While every user has the possibility to check the integrity of a file downloaded with the help of the checksums provided, it might be worth considering the implementation of a program similar to the DSpace checksum checker (see below) to make sure that alterations to the checksum are detected quickly.

Any changes made to the metadata after ingest will lead to the assignment of a new entry in the field “Datum letzte Änderung” (“Date last changed”) and to the creation of a history entry which allows to track which metadata information was changed when, by whom, and how. In this manner, any changes made to the metadata are made transparent and can hence be retraced.

The bitstreams of pedocs publications are preserved by means of backup and mirroring procedures. Thus, the data of the running system is mirrored to a different server

134 The scenario of someone maliciously altering both a digital object and the checksum belonging to it were considered as unlikely and somewhat theoretic by all repository representatives interviewed for this study, and in fact it seems that whether a collection is likely to be attacked in this manner, strongly depends on the content and character of the collection.

135 Currently, neither of the three repositories has a policy defining how digital objects will be treated whose checksum has changed. This is something that should be considered, and for which rules and policies should exist. Thus, in order to allow users to judge how authentic a digital object they are viewing is, and to what extent it really is what it appears or purports to be, changes of the checksum – and hence the bitstream – should be recorded in the metadata if the previous, unaltered version of the digital object cannot be recovered.

and in addition copied to two different servers and written on a tape drive from one of these every night. About once per week, all data is written on DVD for archiving. All hardware change is planned and carried out by the IT-department in accordance with manufacturer recommendations.

2.3.2 JUWEL Archival Storage

As required by both TRAC and nestor criteria, DSpace creates MD5 checksums for every bitstream ingested into the repository and can be set so that the checksum checker continuously loops the repository and reports altered check sums in the respective log. The checksum checker is activated in JUWEL and the logs are frequently reviewed by IT-staff. Although the use of checksums cannot prevent that digital objects stored in the repository are altered, they guarantee that such alterations, which might affect both integrity and authenticity of the object, can be detected. In that checksums are part of the metadata stored in a database, they are stored separately from the digital objects to which they refer as required by TRAC B4.4.

The fact that the checksum is part of the metadata again leads to the question of metadata changes in general. As outlined above, DSpace's authorization system prevents unauthorized persons from editing metadata and hence the number of people who have access to metadata and can edit it is limited and clearly defined. In addition, extensive logging procedures take place on a level below the OSI application layer, where DSpace runs. Thus, changes to the metadata (authorized or unauthorized) are recorded in various logs and can hence be detected and traced. Nonetheless, however, as already pointed out, in order to make *authorized* changes to metadata transparent, JUWEL should consider working with time stamps and history information like pedocs to document the nature of the changes, e.g. in dc.provenance.

JUWEL is currently comprised of two servers – a production server and a mirror server used as a fallback and test server (see Hinz 2008). The operating system runs on a RAID 1 system while the DSpace installation and data are secured by means of a RAID 5 system (4 x 146 GB; see Hinz 2008). Due to the mirroring routine, two identical servers exist at all times. Backups of the servers are made by the Jülich Supercomputing Centre (JSC) on magnetic tape every night. The replacement of storage media is regulated in the extensive IT- and security guidelines of the Forschungszentrum (see below).

2.3.3 Qucosa Archival Storage

In order to make it possible to detect whether the bitstream of a digital object changed after ingest into the repository, Qucosa uses MD5 and SHA512 checksums. Users can retrieve the hashsums by way of links in each metadata record and use them to verify, for example, that the bitstream of the document they downloaded has not been altered since it was ingested into the repository. However, similarly to pedocs, the

checksums are currently not checked again after they were created, and hence the integrity of the stored information packages is not actively monitored and cannot be detected other than by chance. For the future it is thus recommendable to implement a monitoring function as outlined above in order to be able to detect changes to the bitstream of the digital objects. The same is true for the metadata records of information packages. Thus, currently changes made to the metadata after ingest into Qucosa are not recorded or documented in any way. While according to an internal policy metadata should not be altered after ingest, it is (and must be) still possible to change them. As checksums are part of the metadata, however, this means they can be altered along with them. As already argued, it is thus highly advisable to document such changes in the manner outlined above.

The separation of checksum from digital objects required by TRAC is realized in Qucosa so that the parallel (unauthorized) changing of checksum and digital object is similarly impeded as in JUWEL and pedocs.

The bitstreams of Qucosa content are secured by means of extensive backup and mirroring procedures. These include a RAID system as well as continual mirroring. In addition, contents are archived on magnetic tape (for important files, there is a daily backup routine). Backups are not only stored in the SLUB Dresden but also by Campus IT in their own buildings which are somewhat removed from the SLUB buildings. Thus, a spatial distribution of backups is achieved. Storage media and hardware are replaced according to manufacturer recommendations.

2.4 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 2002, 4-2)
>
> The Data Management database holds Descriptive Information, used by Access in searching and displaying objects. It also holds system information about the objects in the archive. The database administrator(s) must maintain and update the database [...].[136]

As rightly observed in the Digital Preservation Management tutorial by Cornell University Library, the Data Management functional entity “provides the glue for the system by capturing and managing all of the metadata needed to operate the system.”[137] Thus, this function is needed to ensure that the correct metadata is linked to the digital objects stored in the repository so that these objects can be addressed, identified, and retrieved. All three criteria catalogs primarily address the need to create and maintain referential

136 http://www.icpsr.umich.edu/dpm/dpm-eng/foundation/oais/data.html – 25.10.2009.
137 Ibid.

integrity between AIPs and their metadata. Thus, TRAC criterion B5.3 requires that "[e]very AIP must have some descriptive information and all descriptive information must point to at least one AIP, such that the integrity can be validated" and suggests the following evidence: "Descriptive metadata; persistent identifier/locator associated with AIP; documented relationship between AIP and metadata; system documentation and technical architecture; process workflow documentation" (TRAC 2007, B5.3). Criterion B5.4 requires that this referential integrity must also be *maintained*.[138] To achieve referential integrity, the nestor catalog suggests that persistent identifiers are used to identify digital objects and all their parts (including metadata), and/or that metadata and content object are stored in one file or directory (cf. nestor 2008, 12.7). Similarly, DINI criterion 2.8 requires that a permanent link is created between metadata and document, e.g. by means of persistent identifiers or containers. In practice, these functions are primarily performed by the repository software; more particularly, the DBMS managing the tables containing the information.

2.4.1 pedocs Data Management

As just pointed out, a primary concern of the Data Management functional entity is the referential integrity between archived information packages and their associated metadata (cf. TRAC 2007, B5.3). This can be achieved by saving objects and their metadata together in one container. pedocs does not currently use such containers but builds its information "packages" virtually, so to speak, by linking metadata and object via the relational database tables.

In addition, a cover sheet will be used in the future to permanently link metadata and digital objects linked by including them in the same file. Thus, each document uploaded into pedocs will automatically (via a PHP script) receive a cover sheet added to the PDF as its first page during the Ingest process. The workflow for this procedure is currently being tested, and a number of documents have already received cover sheets (see for example http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0111-opus-18568 – 01.11.2009). It is planned that by the end of 2009 all documents contained in pedocs will have cover sheets containing core bibliographic metadata, the URN, copyright/use information, as well as the logo of the publisher (where applicable) by whom the work was first published. In addition, if the work was digitized by pedocs, this is also stated. This practice is clearly described in the pedocs policy document so that submitting authors are aware of the fact that the files they submit will be modified in this manner. With this cover sheet, the problem that currently the PDF documents are separated from their metadata once they have been opened, printed, and/or saved on a user's computer is solved.

138 Suggested evidence includes "[l]og detailing ongoing monitoring/checking of referential integrity, especially following repair/modification of AIP; legacy descriptive metadata; persistence of identifier/locator; documented relationship between AIP and metadata; system documentation and technical architecture; process workflow documentation" (TRAC 2007, B5.4).

2.4.2 JUWEL Data Management

The structure of JUWEL information packages is currently defined by the DSpace standard configuration. For AIPs, this implies in particular the structuring of items into bundles and bitstreams outlined above. The actual bitstreams are placed in the so-called “Bitstream Store” which is either the file system on the server or provided by a storage manager (Storage Resource Broker, SRB). A bitstream's location in the file system is identified by means of a “38-digit internal ID […] used to determine the exact location (relative to the relevant store directory) that the bitstream is stored in [...]” (Tansley et al. 2006, 84). All of these components are linked by means of ID numbers which function as keys in the relational databases.[139] In addition, DSpace is capable of creating information packages for export containing, among others, an XML file with the metadata, the item's handle, the license text files as well as the content document(s).

According to the DSpace wiki, among the changes planned for the DSpace 2.0 release is also the implementation of an AIP Asset Store. Thus,

> [i]n the current architecture, all metadata is in a relational database, and all content bitstreams are in the file system on the server. This makes certain preservation-related activities complex, including backups, auditing, and replication/distribution. The proposed new asset store stores metadata and content together as standards-based Archival Information Packages OAIS terminology. This does not replace the current relational database in DSpace. Although these AIPs are the authoritative version of information in the system, the relational database and search indices are still used for performant access; however, these are considered caches of the information in the AIPs […].
>
> An AIP consists of:
> - A core metadata METS document, conforming to the DSpace AIP METS profile [...].
> - Zero or more bitstreams. These may be content files (which are in the file manifest of the METS document), or additional metadata (referred to by the METS document).
> - A checksum for the METS. (DSpace Wiki 2009)[140]

While this Asset Store DSpace on the one hand makes an effort to further OAIS-implementation by creating actual AIPs rather than storing the information only in the relational databases. At the same time however, this might lead to storing the checksum and the digital object which it describes in one container, which would not be in accordance with the TRAC catalog of criteria. In addition, as the relational database is to remain functional, it might become more complex to establish the integrity of the information stored in the repository: thus, not only referential integrity of the database has to be monitored, but in addition it has to be ascertained that database and asset store do not deviate from each other in terms of the information they contain.

As mentioned in above quotation, a METS profile will be available for DSpace AIPs. Similarly, a METS profile for SIPs has been proposed in order to answer to “a need to prepare a METS profile or profiles that will govern the creation of the three types of content 'packages' defined by the [OAIS] reference model” (DSpace Wiki 2008b). Thus,

139 A graphical visualization of the relational database is available at http://dspace.cvs.sourceforge.net/dspace/dspace/docs/image/db-schema.gif?view=markup – 30.10.2009.

140 Please note that the page is currently marked as requiring an update. It is, however, not transparent which information exactly needs to be updated.

although these features have not yet been implemented, in the future DSpace might be able to provide information package descriptions.

In the NLM's test results for the DSpace software it is recorded that “[d]ata integrity checks (checksum) [are carried out] for data transfer but not for version upgrades and format migration (7.4.4.). *Also no referential integrity checks* (7.3.1.2)” (NLM-DRESWG 2009, Appendix C, P2-2; emphasis added). This suggests that DSpace does not meet TRAC criteria B5.3 and B5.4.

2.4.3 Qucosa Data Management

Similarly as the other repositories, Qucosa primarily maintains referential integrity by means of its technical architecture and, more specifically, the relational database. As outlined above, once koLibRI is implemented, it will create information packages all of whose elements are included in one container for submission to the long term archive. This, however, will not affect how information is stored in Qucosa.

Additionally, one of the TU Dresden institutes submitting publications to Qucosa has implemented a routine similar to the one now used in pedocs, in the course of which the URNs assigned to the publications by Qucosa are included in the PDF files so that here, too, metadata and content object are permanently linked as the URN will always direct the user who downloaded or printed a publication to the metadata record associated with it in Qucosa (see, for example, http://nbn-resolving.de/urn:nbn:de:bsz:14-ds-1234432867166-41007 – 01.11.2009).

2.5 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 2002, 4-2)

In the Administration functional entity, “the organizational infrastructure meets the technological infrastructure,”[141] and while all functional entities of the OAIS model play a crucial role in the long-term preservation of an archive's holdings, it seems justified to describe Administration as the entity at the core of these efforts in that among others it coordinates the overall activity of the archive, negotiates submission agreements, develops strategies and policies, and initiates migration processes (Archival Information Update function).

141 http://www.icpsr.umich.edu/dpm/dpm-eng/foundation/oais/administration.html – 25.10.2009.

Before any SIPs are accepted, the repository needs to negotiate and conclude agreements with the producers submitting their material for inclusion in the repository. All three criteria catalogs considered here contain criteria making the existence of a contract or agreement between producer and archive a requirement. Thus, according to the relevant nestor criterion, among others “the character and extent of the submission, the digital repository's obligation to archive the submitted object, conditions of use, and – if applicable – cost” (2008, 3.1; my translation) have to be regulated in an agreement. In addition, the commentary to TRAC criterion A5.2 mentions the repository's “need to be able to work with and potentially modify digital objects to keep them accessible.” The DINI catalog of criteria contains an extensive set of requirements and recommendations concerning legal (copyright) aspects of depositing material (especially if previously published) in the repository. But while its treatment of copyright questions in connection with self-archiving or open access publishing via the repository is quite elaborate, the catalog does not contain a criterion making it a requirement that the repository obtain all necessary rights to carry out long-term preservation actions (regardless of whether the archived document was previously published elsewhere or not). On the one hand this reflects DINI's focus on *all* kinds of (open access) repositories, regardless of whether or not they are concerned with questions of long-term preservation. It seems, however, that the inclusion of such a requirement would help repositories to prepare their collections for potential long-term preservation measures at a later point so that the option to begin offering long-term preservation services is not precluded from the outset merely because the repository has insufficient rights.

The Archival Information Update function, which “provides a mechanism for updating the contents of the archive” (OAIS 2002, 4-11), is addressed by TRAC and the nestor criteria catalog, both of which require the implementation of preservation strategies and point to migration and emulation as possibilities. As none of the repositories currently has long-term preservation strategies implemented and will relegate related activities to a long-term preservation service provider, this function will not be treated in the following.

The aspect of Physical Access Control, an OAIS function “which provides mechanisms to restrict or allow physical access (doors, locks, guards) to elements of the archive, as determined by archive policy” (OAIS 2002, 4-11) is covered by nestor and DINI (see nestor 2008, 6.2 and DINI 2007, 2.5.1), but not considered by TRAC. All of the repositories considered here have established sufficient barriers limiting physical access to their servers; hence this function, too, will not be discussed again in the following.

A central task of the Administration functional entity is the establishing of standards and policies, and it is here (as well as in the Preservation Planning entity treated below) that a considerable gap between “traditional” repository tasks and long-term preservation-related requirements becomes apparent. According to OAIS, the standards to be established in the Administration functional entity “include format standards,

documentation standards and the procedures to be followed during the Ingest process" (2002, 4-11; emphasis omitted). In addition, this function

> provides approved standards and migration goals to Preservation Planning […,] develop[s] storage management policies (for the Archival Storage hierarchy), including migration policies to assure that archive storage formats do not become obsolete, and database administration policies. It will develop disaster recovery policies. It will also determine security policies for the contents of the archive [...]. (OAIS 2002, 4-11; emphasis omitted)

The nestor and TRAC criteria require the repository to develop and establish a number of standards and policies, including a (mission) statement in which the repository takes responsibility for preserving the repository content, a definition of its designated community/target group, as well as a definition of the stored digital object's significant properties, to be preserved over the long term. As Knight explains in the InSPECT work package 3.3, "significant properties are the characteristics of the Information Object, encoded in a digital object that must be reproduced, even if there are changes to the hardware and software in which the Information is created and managed" (2008, 4). Similarly, in an earlier version of the work package, Wilson defines significant properties as "the characteristics of digital objects that must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the objects" and divides them into five different categories:

> - content, e.g. text, image, slides, etc.
> - context, e.g. who, when, why.
> - appearance, e.g. font and size, colour, layout, etc.[142]
> - structure, e.g. embedded files, pagination, headings, etc.
> - behaviour, e.g. hypertext links, updating calculations, active links, etc. (Wilson 2007, 8; emphasis omitted)

As Knight points out, "the significant properties are closely linked with the need to maintain the authenticity […] and integrity […] of the Record" (2008, 4). They also serve to focus preservation efforts by making it clear which properties and characteristics of a digital object are considered as most important and hence need to be preserved as complete as possible, and which of its properties might have lesser weight and are dispensable.[143]

In accordance with the above observation that the distinction between traditional and preservation-centered repositories becomes particularly evident in the field of standards and policies, DINI's treatment of policies shows that it is not primarily concerned with long-term preservation. Thus, although a number of policies also relevant to long-term preservation are required and recommended (e.g. concerning rights and duties of authors and repository, the definition of selection/collection criteria, subject indexing, or the deletion of digital objects from the repository), the extent to which the repository should commit itself to long-term preservation remains fairly open. DINI requires the repository to

142 In Knight's 2008 version of the work package (3.3, v1), the appearance category has been replaced by the term "rendering."

143 In this context, Knight also draws attention to the fact "that the level of significance attributed to each property is subjective and may change over time as it is tailored to the needs and capabilities of the institution and the activities it performs" (2008, 4).

guarantee that documents are “archived” for a specified period of time depending among others on the content and technical quality of the digital objects (see DINI 2007, 2.2) and that documents will be available (as outlined in DINI 2007, 2.5) for at least five years. These requirements therefore merely concern bitstream preservation measures: the digital objects have to be available for five years, but it is not a requirement that they remain accessible and interpretable over this period. This limitation to bitstream preservation is understandable in that repositories seek (and should receive) certification which do not aim at becoming a long term archive.[144] However, the DINI criteria should make an attempt at defining more precisely what is meant by “archiving” (see DINI 2007, 2.2 “Archivierungszeiträume”) and by “availability” (see DINI 2007, 2.8 “Dokumentverfügbarkeit” and 4.8, which uses both “Langzeitverfügbarkeit” and “Langzeitarchivierung” without explaining the differences between these concepts).

In addition to the policies mentioned so far, TRAC requires that the repository establishes a mechanism for update and review of policies in answer to change[145], as well as “written policies that specify the nature of any legal permissions required to preserve digital content over time” (TRAC 2007, A3.3). The nestor criteria catalog, on the other hand, contains an additional criterion that the repository must have specified selection criteria for the digital objects it accepts as submissions – an important aspect that is not covered by TRAC (see, however, DINI 2007, 2.2, which requires that selection/publication criteria have to be specified in the repository's policy).

As specified in the OAIS model, “[t]he Audit Submission function will verify that submissions (*SIP or AIP*) meet the specifications of the Submission Agreement [...]. The audit process must verify that the quality of the data meets the requirements of the archive […]” (2002, 4-11-4-12), and therefore complements the Quality Assurance procedures carried out in the Ingest functional entity. The TRAC catalog in particular specifies the need for repositories to establish the quality of a digital object submitted to be included in the repository by evoking the concepts of integrity and authenticity.[146] In contrast, neither nestor nor DINI address this issue as explicitly. As discussed above, the nestor catalog requires that integrity and authenticity of digital objects are secured at all times from Ingest onwards. This, one can assume, also implies that the completeness and correctness of submissions has to be audited – this requirement is, however, not made explicit. DINI, on the other hand, contains a criterion demanding that only documents are stored on the server which meet the selection and publication criteria of the repository

144 On the other hand, for example in the light of the German Research Foundations recommendation that according to good scholarly practice data created and used in research projects should remain available for at least ten years, it seems questionable whether a guarantee that digital objects remain available for a minimum of only five years is enough to make these objects usable for scholarly purposes.

145 Arguably, the review and update of policies is implied in the nestor requirement to engage in long-term planning (nestor 2008, 4.4) – however, seeing how important it is to keep policies current and to adapt them to technological, organizational or other changes, mentioning this point explicitly seems warranted.

146 Note, however, that TRAC strictly speaking only requires the quality of the *AIP* (but not the SIP) to be checked “at the point it is generated” (2007, B2.11). Although the SIP's completeness and correctness is the prerequisite for the completeness and correctness of the AIP, this criterion should be modified to include SIPs.

(see DINI 2007, 2.5.1, 4.5.1). This also suggests (somewhat more obviously) that some kind of quality control has to be carried out before a digital objects is ingested into the repository.

2.5.1 pedocs Administration

Administration: Negotiate Submission Agreement

pedocs enters contracts with two different kinds of depositors – individual authors (or their representatives), and publishers who agree to the publication of previously published titles (predominantly out of print monographs and journals), often in combination with a commitment of pedocs to digitizing the material to be (re-)published. Thus, two different kinds of contracts exist – an author agreement and contracts/deposit agreements with publishers – specifying maintenance, access, and withdrawal of submitted documents. As outlined above, in the future authors will have to accept the author agreement via a double opt-in email procedure.[147] With the agreement, authors will transfer all rights necessary for long-term preservation through the DNB and the right to publish their work openly accessible via pedocs to DIPF.

Administration: Establish Standards and Policies

As required by the criteria catalogs considered here, pedocs has defined its policies, collection and preservation objectives as well as general goals in a policy document (pedocs Guidelines) which also fulfills the purpose of a mission statement. The policy not only specifies collection guidelines and principles but also demonstrates pedocs' commitment to preserving the digital objects accepted into the repository for the long term in cooperation with the DNB. The policy outlines selection criteria, taking into account content and form of documents, the document type (among others pedocs excludes websites, legal texts, user manuals, lecture materials such as PowerPoint slides or handouts, advertising material, and leaflets from its collection), as well as file format. As the policy states clearly, pedocs

> aims to provide documents on a long-term basis, and it collaborates with the German National Library in order to achieve a long-term storage of its documents [...]. Each text is assigned an individual address (URL), allowing for immediate access. Each document is furthermore assigned a Persistent Identifier (URN) for its long-term, global and unambiguous referencing, independent from location. Texts and their descriptions are registered with the German National Library for long-term archiving, where they are catalogued and permanently archived. These procedures are irreversible once they have been completed.[148] (pedocs Guidelines)

The designated community is defined in the pedocs policy, albeit somewhat implicitly and in a very general fashion. Thus, it is stated in section I:

147 Currently, work is carried out to implement a new metadata field which will make it possible to link a record and a document with a certain author agreement. Thus, if one author agreement is modified or replaced, it becomes possible to trace for which objects in the repository a given agreement is valid.

148 The policy states clearly that publications cannot be deleted once the publication process is completed. However, a record can be suppressed from showing up in the database in exceptional cases as laid down in §42 UrhG (see http://bundesrecht.juris.de/urhg/__42.html).

> pedocs is a scholarly open access document server for scientifically relevant publications in the field of educational research and educational science [...]. pedocs offers authors in the disciplines of educational science and education as well as related domains an organisational and technical framework for publishing scientifically relevant documents in electronic form. (pedocs Guidelines)

Certainly, the authors mentioned constitute one major portion of the designated community of pedocs, as often producers will also be users of the service. However, it would be desirable to have more detailed information about the community/communities for whose use pedocs is primarily intended – now and in the future. At this point in time it might seem comparably clear at whom the services offered by pedocs are addressed and what the knowledge base of the designated community or communities is, it will become more difficult in the future to determine the knowledge base available to a designated community not precisely defined. In fact, the policy of FIS Bildung is somewhat more precise here by stating that its services are addressed at both educational research and training, as well as at educational practice (see FIS Bildung Policy).[149]

Although the pedocs policy is frequently reviewed and adapted, there are currently no mechanisms or policies governing this review process.

Significant properties: In that it is crucial in long-term preservation to define an object's (or object class's) significant properties, both the nestor and TRAC criteria catalog require the repository to identify these properties. This definition has not yet been completed for pedocs. Currently the policy only states that the content of the documents submitted and published will be preserved, but that this might not be possible for the layout (see pedocs Guidelines, Section VII), thus indicating that content is considered more important than appearance/rendering. In the future, a more detailed list of significant properties to be preserved for different document types will be added.

Administration: Audit Submission

In order to ascertain that the files published in pedocs adhere to the criteria outlined in the policy, submissions are subject to both intellectual and automatic control. As outlined above, JHOVE is used to identify and validate file format. In addition, the repository staff determine by means of intellectual control whether documents are appropriate for publication in pedocs with regard to form, content, and quality (in particular also quality of metadata). The completeness of the Submission Information Packages to be stored on the server as (pre-)AIPs is primarily checked and guaranteed by the software – thus, for example, a record cannot be saved without having filled in metadata in the required fields; a URN is added automatically so that published records will always have a URN.

2.5.2 JUWEL Administration

Administration: Negotiate Submission Agreement

149 http://www.fachportal-paedagogik.de/fis_bildung/fis_policy_e.html – 11.10.2009.

The Forschungszentrum Jülich is a GmbH which holds all rights to the publications of its staff. It follows that technically a deposit agreement with individual authors is not strictly necessary. Nonetheless authors submitting material to JUWEL have to accept a deposit agreement ("Lizenbestimmungen") as part of the submission procedure, which regulates the following:

1. Non-Exclusive Publishing Rights, giving the Central Library the "non-exclusive right to save, copy, provide access to the resource worldwide and to create print and electronic copies as required"
2. Third-Party Rights: the depositors have to confirm that no third-party rights are infringed by publication with JUWEL
3. Compliance with the Publication Guidelines of the Forschungszentrum Jülich
4. Transformation into Other Formats (electronic or physical), which is granted e.g. for the purpose of accessibility or long-term preservation
5. Transfer to the German National Library (DNB), whereby the DNB is to be granted the same rights to use the submitted resource as JUWEL. (see JUWEL Policy)

Similarly, an agreement with the campus press exists – however, again this is not an agreement with a third party as the press is integrated into the Central Library which also hosts JUWEL.

In addition to the deposit agreement, the publication process is also governed by the above-mentioned Publication Guidelines of the Forschungszentrum Jülich, which regulate how researchers employed at the Forschungszentrum publish the results of their work and among others outline quality control mechanisms.

Administration: Establish Standards and Policies

As a service offered by the Central Library of the Forschungszentrum Jülich, JUWEL is governed not only by its own policy but in addition by the Open Access policy of the Forschungszentrum, which is itself derived from the Open Access policy of the Helmholtz Association.[150] Neither policy is linked from the JUWEL pages, which makes it more difficult for users to place the repository and its services in context, and to understand what its mission is in relation to the Central Library and the Forschungszentrum. For JUWEL, no policy document as such exists – thus, as mentioned above, the document labeled "JUWEL-Policy" contains the deposit agreement. In order to make the purpose and objective of JUWEL more transparent – both with regard to Open Access and long-term preservation as well as its intended role in the Forschungszentrum and the designated communities beyond it –, it would be advisable to draft a new policy document which fulfills these functions.

Currently, the repository only takes responsibility for bitstream preservation (see JUWEL FAQ). This corresponds to the DSpace preservation service level for known

150 The Forschungszentrum's policy is available at http://www.fz-juelich.de/zb/index.php?index=758 and the policy of the Helmholtz Association is stated under http://oa.helmholtz.de/index.php?id=137 – 18.10.2009.

formats.[151] Although the possible need to transform resources into different “electronic or physical formats” is mentioned in the author agreement and in the Frequently Asked Questions, no further, more detailed information on preservation policy (including future plans) exists. In this sense, the repository's policy documents do not yet wholly reflect a “commitment to the long-term retention of, management of, and access to digital information” as required in TRAC (2007, A1.1).

The JUWEL policy documents do not define the repository's designated community/communities explicitly. In part, and in particular with regard to the producers, these are co-extensive with the designated community of the Central Library. In addition, as the documents published via JUWEL are to be made available worldwide (e.g. by inclusion in Open Access search engines such as BASE or OAIster), researchers working in the respective disciplines belong to the designated communities of JUWEL.

The repository has developed criteria for the selection of digital objects to be included in JUWEL, and the types of documents which will be accepted are defined in the Frequently Asked Questions.[152] In addition, both the FAQ and the JUWEL Help page make clear that publishing with JUWEL is only possible if at least one author of the publication is employed at the Forschungszentrum or if the work was written at the Forschungszentrum. The FAQ pages also contain information about accepted file formats. Additionally, which documents can be accepted is regulated by the publication guidelines of the Forschungszentrum.

As long-term preservation strategies are not yet implemented or planned, no significant properties for the digital objects stored in JUWEL have been defined so far.

Administration: Audit Submission

As mentioned above, JUWEL will not reject submissions with incomplete or incorrect metadata automatically as no checks are carried out whether the necessary fields were actually filled in by the submitting authors. However, as the JUWEL ingest workflow includes all three steps offered in DSpace, repository staff will check metadata and edit or add metadata where necessary.

As already discussed, no comprehensive format identification and validation procedures have been implemented in JUWEL.

2.5.3 Qucosa Administration

Administration: Negotiate Submission Agreement

151 DSpace distinguishes between unsupported, known, and supported file formats (see Bass et al. 2002). According to the MIT definition of support levels, a known format is one that “is recognized, and the hosting institution will promise to preserve the bitstream as-is, and allow it to be retrieved. The hosting institution will attempt to obtain enough information to enable the format to be upgraded to the 'supported' level” (Tansley et al. 2006, 12). Note that the support level can be defined by the host institution according to its requirements and preservation goals. The file formats accepted and supported by JUWEL are outlined under http://juwel.fz-juelich.de:8080/dspace/help/formats.jsp – 18.10.2009.

152 Thus, among others, JUWEL accepts journal articles, reviews, abstracts, monographs, chapters of books or proceedings, and doctoral or habilitation theses. The guidelines for the VDB specify each document type.

The procedure in which authors submit a signed deposit agreement in print to the SLUB Dresden in addition to accepting an agreement during the online submission procedure was already explained above (see Ingest). It goes without saying that Qucosa should take care that the signed agreements are archived with the necessary precaution as they form the legal base of both the publication and the preservation process.

Administration: Establish Standards and Policies

According to the Qucosa Policy, the repository serves the following purpose:

> Qucosa dient der Publikation, dem Nachweis und der langfristigen Archivierung von Dokumenten aus Wissenschaft und Wirtschaft. Das von den wissenschaftlichen Bibliotheken im Freistaat Sachsen getragene Angebot ist Teil der internationalen Open-Access-Bewegung. (Qucosa Policy)

This policy statement contains a commitment to long-term preservation which is further specified in the FAQ: “Die Publikationen sind ohne zeitliche Beschränkung verfügbar und für die langfristige Archivierung vorgesehen” (Qucosa FAQ). According to this statement, Qucosa currently guarantees the long term accessibility of bitstreams and is planning to implement long-term preservation measures in the future. Should more concrete statements about or commitments to long-term preservation be made in the future, it might be worth considering the definition of different preservation service levels (e.g. in analogy to the DSpace levels; see Bass et al. 2002) to indicate which file formats will be (attempted to be) preserved and which might be too difficult to preserve. This seems advisable especially as the repository also accepts dynamic content in proprietary file formats, which might pose considerable problems for long-term preservation.

For the selection of content, Qucosa has outlined a set of criteria defining which documents will be accepted for publication (stated in the Qucosa Policy and the Publication Guidelines). Thus, the following criteria have to be fulfilled for a document to be accepted for publication:

> Ein über Qucosa zu veröffentlichendes elektronisches Dokument erfüllt folgende Bedingungen:
> 1. Es ist zur Verbreitung in der Öffentlichkeit bestimmt.
> 2. Sind Aktualisierungen am jeweiligen Dokument notwendig, wird das geänderte Dokument als neue Version gespeichert.
> 3. Das Dokument entspricht den von Qucosa vorgegebenen Veröffentlichungsparametern. (Qucosa Policy)

The “publication parameters” mentioned above are stated in the publication guidelines and include the document type[153], the absence of DRM tools/measures, as well as the affiliation of the author. Although the last criterion is not mentioned explicitly, and although this rule may be subject to exception, Qucosa is currently a service open only to depositors from Saxon institutions.

153 Document type, however, is a fairly non-distinctive criterion in that it also contains a category “Other publications.” Thus, currently, no publication type is explicitly excluded.

While the preferred file format is PDF[154], other formats – open and proprietary – are accepted as well so that file format is currently not a selection criterion. In contrast to pedocs, dynamic content/elements is/are not excluded.

The designated community of Qucosa is at this point defined only very broadly and moreover implicitly. Thus, it is stated in “Über Qucosa” that the repository is a service focusing on scholarly documents and documents from the business sector (“Dokument[e] aus Wissenschaft und Wirtschaft”). As the list of accepted document types shows, other publications included are musical scores[155], which will be of interest to a very specific designated community that, however, will probably be very different from the designated community of publications from the business sector, for example. Finally, public administration institutions are also included in the list of institutions registered for submitting material to Qucosa, and their publications will be of interest to a very heterogeneous group of users.

For the purpose of monitoring the repository's designated community or communities, this implicit definition deriving from the published document types is far too imprecise. On the other hand, however, the question is whether it is feasible and possible to define the designated communities for a service like Qucosa, which is intended to collect and preserve material from and for the entire federal state of Saxony, in a detailed fashion.

Significant properties are not currently identified or defined for Qucosa publications. Although the significant properties are something that will have to be specified when the cooperation with a service provider for long-term preservation is brought under way, how this is to be achieved is according to Dr. Kluge still unclear. Thus, in his opinion the concept of significant properties seems to theoretic and abstract, in particular as it is entirely unclear to us today which properties will be regarded as significant by future users and depositors (see also Knight 2008, 4). Thus it is worth considering, according to Kluge, a flexible and ad-hoc specification of significant properties together with the designated community whenever it becomes clear that a format is in danger of becoming obsolete and if migration tests show that that information will be lost.

Administration: Audit Submission

A document is only published after intellectual control by a member of the repository staff. This quality control currently primarily concerns the associated descriptive metadata. In the future, with the implementation of koLibRI, quality control will also extend to technical aspects such as file format validity, etc.

154 In the file upload dialog, the required format of what in DSpace would be called the primary bitstream is given as PDF. However, the Qucosa software currently does not reject files in other formats. As Qucosa also accepts content such as audio, film, etc. it seems problematic to limit the preferred bitstream to PDF – this is a decision which should be reconsidered.

155 In fact, the German Composers Association (Deutscher Komponistenverband) recommends the publication of musical scores via Qucosa. See http://www.komponistenverband.de/content/view/471/117/ – 18.10.2009.

2.6 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. (OAIS 2002, 4-2)

The basis of the Preservation Planning functional entity in the OAIS model is provided by the "Monitor Designated Community" and "Monitor Technology" functions, which seek to detect changes – e.g. in the designated community's "service requirements and available product technologies[156]" or with regard to "digital technologies, information standards and computing platforms […] to identify technologies which could cause obsolescence [...]" (OAIS 2002, 4-13). These functions are of utmost importance to long-term preservation in that they guarantee that an archive's or repository's holdings remain both understandable and accessible to the designated community. While both functions are covered by the nestor and TRAC criteria catalogs, each catalog takes a somewhat different approach to them.[157] Thus, TRAC includes four criteria which can be linked to the "Monitor Technology" function, and one criterion which refers directly to the "Monitor Designated Community" function in that it requires the repository to demonstrate "that it systematically and routinely seeks feedback from stakeholders to monitor expectations and results, and that it is responsive to the evolution of requirements" (TRAC 2007, A3.5). The nestor criteria catalog contains a similar criterion, requiring a repository to take measures which secure the understandability and accessibility ("Interpretierbarkeit," understood as both intellectual and technological interpretability of an object) of the digital objects (content and metadata) to the designated community, and to ascertain frequently that interpretability is still given (see 2008, 2.2). In addition, criteria 4.4 and 4.5 in the nestor catalog address the two monitoring functions in a more general manner, e.g. by requiring the repository to "react to substantial changes" (2008, 4.5; my translation).[158] Including reaction to substantial changes in technology in addition to changes in organization and society, criterion 4.5 makes it necessary for the repository to observe closely how technology, its designated community's knowledge base and service requirements evolve, something, it should be noted, which is only possible if the designated community was (and can be) defined with sufficient precision. A need to monitor technology also derives

156 "Such requirements might include data formats, media choices, preferences for software packages, new computing platforms, and mechanisms for communicating with the archive" (OAIS 2002, 4-13).

157 The DINI catalog does not contain criteria relevant to Preservation Planning and will accordingly not be considered in the following.

158 In addition, the explanatory notes to criterion 4.4 contain an explicit reference to the "Monitor Technology"/"Monitor Designated Community" functions as the basis for any long term planning (see nestor 2008).

from nestor criterion 8, making it prerequisite for repositories to engage in long term planning of their technical long-term preservation measures.

Closely linked to the two monitoring functions in the OAIS model is the function "Develop Preservation Strategies and Standards," which

> is responsible for developing and recommending strategies and standards to enable the archive to better anticipate future changes in the Designated Community service requirements or technology trends that would require migration of some current archive holdings or new submissions.[159] (OAIS 2002, 4-14)

This includes the specification of file formats which are supported by the repository and considered as suitable for long-term preservation.

Finally, the Preservation Planning functional entity contains the function "Develop Packaging Designs and Migration Plans," which "develops new IP designs and detailed migration plans and prototypes, to implement Administration policies and directives" (OAIS 2002, 4-14). Thus, one of the main tasks of this function is the design of Information Package templates, i.e. of detailed specifications of how SIPs, AIPs, and DIPs are structured and which elements they contain. Both nestor and TRAC require the definition of SIPs and AIPs, but differ somewhat in their requirements concerning Archival Information Packages. TRAC primarily repeats the elements of an AIP according to the OAIS model whereas nestor is on the one hand more general than TRAC in simply requiring the definition of AIP structure (without explicit recourse to OAIS); on the other hand, nestor is more specific than TRAC in demanding that suitable file formats as well as location where the AIP is saved are defined (cf. 2007, 10.1) and in suggesting the use of open formats as well as the use of XML and METS for a description of AIP structure. Thus nestor puts a stronger emphasis on *how* the structure of AIPs is to be defined rather than spelling out what the structure is supposed to look like.

2.6.1 Preservation Planning: pedocs, JUWEL, and Qucosa

With regard to Preservation Planning, all three repositories considered in this study are in a very similar situation as none of them is or is planning to become the long-term archive in which its collections will be preserved. In all three cases this means that although some preservation planning activities will be carried out in cooperation with the long term archive, many activities will remain the sole responsibility of the latter.

Nonetheless, all institutions involved in the preservation cooperation (i.e. both the repositories and the long-term archives with which they intend to cooperate), will have to monitor their respective designated community/-ies, their knowledge bases, and the technologies these communities use, understand, and have access to. Similarly, both the repositories and the long-term archives will have to be able to detect changes in technology relevant to the software they use, while in addition the long term archive will

159 In addition, "[t]his function receives reports from the Monitor Designated Communities and Monitor Technology functions, and […] sends recommendations on system evolution to Administration" (OAIS 2002, 4-14; emphasis omitted).

also have to monitor file formats, Representation Information (cf. TRAC 2007, B3.2), etc. in order to protect the digital assets it preserves from the danger of obsolescence.

Preservation Planning: Develop Preservation Strategies and Standards

To the extent that the Develop Preservation Strategies and Standards function is also "responsible for developing and recommending strategies and standards *to enable the archive to better anticipate future changes in the Designated Community service requirements*" (OAIS 2002, 4-14; emphasis added), this function will also have to be addressed by the three repositories, even if they are not the long-term archive – in particular as access to the preserved digital objects will always take place through them. As outlined above, regular and well-structured communication between the repository and its digital preservation service provider will be necessary to accomplish this. On the other hand, strategies spelling out, for example, which steps are to be taken if a file format is in danger of becoming obsolescent, have to be developed by the long-term archive.[160]

As already pointed out above, all three repositories must, moreover, have concepts, standards, and strategies for bitstream preservation and protection, both to ensure that these bitstreams remain uncorrupted in the time elapsing between submission to the repository and ingest into the long-term archive and in order to ascertain that users access authentic and uncorrupted publications.

Preservation Planning: Develop Packaging Designs and Standards

The development of packaging designs and standards, too, has to be undertaken by the repository and long-term archive in cooperation wherever information packages are submitted from one to the other. Thus, cooperation agreements will have to be drafted which specify exactly the design of the SIPs submitted by the repository to the long-term archive (e.g. the DNB in the case of pedocs), including required elements and metadata. Thus, for example, it is already clear that pedocs SIPs will most likely include a certain set of descriptive and technical metadata (file format and format version) together with the content object when submitted to the DNB.[161] In all cases, the design of the SIPs to be submitted to the long-term archive will strongly depend on the software used to build them (be that the repository software or an external tool such as koLibRI), whereas the design of AIPs to be stored in the long-term archive will be determined by (the requirements and capacities of) the latter (including the definition of formats, storage locations, etc. as outlined in nestor 2008, 10.1).[162] In this context, as required by TRAC, a catalog of minimum requirements to be met by information packages if their content is to be

160 Note that this does not mean that such steps cannot also consist in discussing concrete actions with the repository whose collections have to be migrated.

161 The DNB has defined a core set of metadata for electronic monograph publications deposited to the DNB (in German) which can serve as an example for a possible SIP design. See http://www.d-nb.de/netzpub/info/pdf/metadaten_kernset_extern.pdf – 25.10.2009.

162 One of the most striking differences between the pre-stage AIPs currently stored on the repositories' servers and the AIPs preserved by the long term archive will certainly be the use of preservation metadata, including extensive technical metadata.

preserved for the long term, will have to be created. In turn, repositories will have to have a policy which regulates how files will be treated that are rejected by the long-term archive, e.g. for reasons of format validity, but which were ingested into the repository and are accessible through it. In addition, the design of information packages disseminated by the long-term archive for re-ingest into the repository has to be agreed upon.

Finally, repositories and long-term archives will have to define an exchange format for the SIPs, e.g. the UOF generated by koLibRI.

2.6.2 pedocs Preservation Planning

Preservation Planning: Monitor Designated Community, Monitor Technology

The DIPF makes available "scientific infrastructure and research services to researchers, practitioners, administrators and policy-makers in the field of education."[163] It can therefore be assumed to be in close contact with the researchers and practitioners who form the designated community for pedocs and the other (electronic) services comprised in the German Education Portal. All activities of the Portal and its modules are monitored by an Advisory Board, consisting of both LIS professionals and researchers from the field of education. In that members of the Advisory Board belong to the designated community of pedocs, communication with representatives of this community – prerequisite for the detection of "substantial changes," for example, in the community's knowledge base or in the way the community publishes – is facilitated. Currently, plans exist at pedocs to seek feedback about the services offered by the repository from Advisory Board members and other users. It seems crucial that these plans are put into practice in the near future, e.g. in the form of regularly repeated usability tests and surveys. Thus, surveys (ideally carried out with users and non-users) might be used to find out more about both the designated community's knowledge base, the technology its members use/are capable of using, and its ideas about scholarly communication.[164] Interestingly, pedocs itself can become a source of information about scholarly communication-practices of its designated community, as a series of articles on digital publishing trends in education science from the current issue of Erziehungswissenschaft (20, 2009) also published on pedocs shows (see, for example, http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0111-opus-18510 – 01.11.2009).

The responsibility for monitoring technology changes will lie partly with the DIPF IT-department in Berlin and partly with the DNB. Thus, as mentioned above, particularly monitoring RI and file formats obsolescence will fall within the responsibility of the DNB. However, pedocs will be required to provide the DNB with information concerning the technologies used and accepted by its designated community; this will not necessarily be within the scope of the DNB's monitoring efforts but is required information if, for example,

163 http://www.dipf.de/en/institute/mission/leitbild-1 – 03.11.2009.
164 Note that according to TRAC A3.5, policies should exist that govern how feedback from the designated producer and user communities is sought and addressed.

one wants to make sure that target formats of migrations are in fact used by and understandable to the designated community. In addition, the DIPF IT-department has to make sure that all components of the system on which pedocs is running are kept up to current standard. For this activity no documented procedures or policies exist – it is rather the case that the IT-staff, who work with the hard- and software on a daily basis and of which one person is exclusively responsible for pedocs, keep an eye on the hard- and software components during their activities.

As the repository software itself is by now too far removed from its original base OPUS, bugfixes for this software are no longer monitored or implemented.

Preservation Planning: Develop Preservation Strategies and Standards

In addition to the responsibilities outlined for all three repositories above, pedocs needs to develop strategies and standards for the preservation of the pedocs author agreements, which exist only in digital form. Currently, the agreements are merely saved and backed up in the file system in Frankfurt – no concept or strategy for their preservation exists at the moment and will thus have to be developed at a later point.

2.6.3 JUWEL Preservation Planning

Preservation Planning: Monitor Designated Community, Monitor Technology

The primary designated community of JUWEL is composed of the researchers working at the Forschungszentrum and of external researchers working in the same disciplines worldwide. As the library staff is constantly in touch with the researchers working at the Forschungszentrum, providing services crucial to their work, it can be assumed that changes in the designated community – e.g. with regard to scholarly communication/publication culture, knowledge base, or the technology used and available – will be detected. This seems even more likely as the campus press of the Forschungszentrum Jülich is also located within the Central Library so that significant changes in publication culture can easily be communicated. Thus, even though no user surveys were carried out so far, according to Wagner, feedback from users easily reaches JUWEL staff through these channels.

Both hardware and software requirements are monitored and evaluated by campus IT and IT-security officers (cf. IT Security Guidelines; see below), who take care that hard- and software are up to current standard. In particular, this includes a monitoring of developments with regard to the DSpace software which is regularly updated following a conservative update policy (thus, for example, as no critical problems exist with the current version, JUWEL has not updated to DSpace 1.5 as of yet).

2.6.4 Qucosa Preservation Planning

Preservation Planning: Monitor Designated Community, Monitor Technology

It seems that of all three repositories, Qucosa is in the most difficult position when it comes to defining and monitoring its designated communities. Thus, as explained above, the latter are of a highly heterogeneous nature including among others researchers, scholars, and students of Saxon universities, as well as depositors/users from vocational colleges, the business sector and from public administration. It seems that especially documents published by public administration will be of interest to a “wider public,” making it particularly difficult to define – let along monitor – their designated user community precisely. It follows that a close and detailed monitoring of all communities that might have an interest in the publications provided through Qucosa is impossible to accomplish with limited financial and human resources. While this does not mean that the SLUB Dresden should refrain from any monitoring activity, a workable solution has to be found – e.g. by limitation to few “model communities.” Thus faculty and students of the TU Dresden (possibly limited to the disciplines most strongly represented in the repository) might serve as such a model community. In addition, a routine could be established in which other institutions whose members submit publications to Qucosa or existing advisory councils monitor their communities and notify Qucosa should significant changes occur.

Similarly as in the case of the other two repositories, surveys should be carried out among users and producers in the future to gather information about the technology they use, for example, or about the usability of Qucosa from their perspective (e.g., the understandability of the metadata).

In that Qucosa is managed by department for information technology of the SLUB Dresden, technological developments are closely monitored by the staff who will react should significant changes, e.g. in hardware or software technology, occur. However, no documented policies exist for this monitoring activity.

2.7 Common Services and Requirements

In addition to the functional entities described and explained above, the OAIS model defines a set of (strongly IT-centered) common services meant to support the operation of the long term digital archive, for example by providing security services “to protect sensitive information and treatments in the information system” (OAIS 2002, 4-4). Similarly, the criteria catalogs introduced above contain a number of criteria which concern the entire repository rather than a single functional entity or sub-entity, addressing the necessary IT-infrastructure, quality management as well as the overall organization of the long-term archive (see Appendix A).

Thus, nestor criterion 4 requires the long-term archive to be organized in a manner allowing it to fulfill the goals it set itself, pointing to adequate and financial and human resources among others (see nestor 2008, 4.1 and 4.2). While currently all three repositories seem adequately equipped with staff and resources (although, clearly, the question of adequacy in this context is one of “those 'would you like more money?'

questions" cited above) and are well-integrated into the larger structures and processes of the organizations of which they form part[165], Dr. Kluge of the SLUB Dresden pointed to the general difficulty that publicly funded institutions have to comply with the requirement of long-term strategic planning (see nestor 2008, 4.4, 4.5) – particularly with regard to financial and human resources. Thus, which budget they receive is largely out of their control and is hence predictable at best for fairly short time spans. This situation is aggravated, according to Kluge, by the growing importance of project-based work in universities and other institutions whose members deposit to Qucosa or other repositories. Thus, with the beginning of such projects, Qucosa and other repositories may – quite unexpectedly – be faced with requirements and demands which cannot be planned and calculated ahead very well.

Similarly, it is extremely difficult to plan ahead for the case that financial or organizational support of long-term preservation is no longer given because the institution in question is no longer able to provide that support. While the three repositories presented in this study are aware of this possible threat, no concrete succession plans exist as of yet, as currently other, more pressing issues and questions need to be addressed. However, in particular the question of if and how the repositories' digital collections will remain accessible if the repository itself does not exist anymore needs to be addressed, as this is the primary function provided – possibly exclusively – by the repository. It is for these reasons that, according to Kluge, a centralized nationwide or even international solution for long-term preservation of digital materials has to be sought and is to be preferred over an institutional solution. Although the kopal project was certainly a first step towards such a cooperative solution, considerable development and standardization work will have to be carried out in the future, in which the National Libraries will continue to play a highly important role.

Both the nestor and the TRAC catalog require the repository to perform comprehensive quality management activities, e.g. by documenting processes, elements, and responsibilities (see nestor 2008 5, 5,1, 5.2 and TRAC 2007, A3.6, C1.8, B1.8, B4.5). Quality management procedures also play a role in the DINI criteria, especially with regard to IT-infrastructure.[166] To some extent, the repository software and its architecture as well as the various logs it writes, help to support quality management. Therefore it is highly important that the repositories make use of the possibilities offered by these logs, which seemed to be the case for all three repositories considered here. In addition, the definition of responsibilities and workflows is also supported by the respective software, which, for example, allows the definition of roles (e.g. administrator, user) and the rights

165 Thus, consider, for example, the integration of pedocs into the Fachportal Pädagogik and its connections with FIS Bildung, or JUWELs links with the VDB as well as the campus press. Strong (organizational) links also exist between Qucosa and the digitization center at the SLUB Dresden.

166 See Dobratz et al. 2008 for a discussion of the "Use of Quality Management Standards in Trustworthy Digital Archives."

associated with them, or has defined workflows for submission, storage, and dissemination.

On a more general level, responsibilities are clearly defined by the institutional structure in which the repositories are embedded. In addition, in all three repositories there was a concern not to create “monopolies of knowledge,” but to document processes and procedures so that they are commonly known – e.g. by using wikis or other forms of documentation. It seems important that these activities are adhered to and followed through, ideally supported by a policy or guideline.

Finally, as already mentioned, all criteria catalogs contain requirements concerning a secure IT-infrastructure (see, among others, nestor 2008, 13-13.2, 14; TRAC 2007, C1.1-C1.2, as well as DINI 2007, 2.5). In all three repositories, the IT-infrastructure seemed adequate to guarantee that services could be offered according to the repositories' set goals, including the protection of integrity and authenticity of the digital collections. In all cases, only a very limited number of persons had access to the servers both physically and logically, and encrypted communication for server log-in was used, for example, by pedocs and JUWEL. JUWEL in particular benefits from the extensive IT-security measures in place at the Forschungszentrum Jülich, on whose campus highly sensitive data is produced, stored, and dealt with, e.g. in the field of nuclear power plant technology.[167] It follows that – as so often – sources of potential threats to the data stored in the repositories are primarily to be sought within the institutions rather than the outside; that is, data is endangered by human error or hardware defects much more than potential intruders from the outside.

167 The IT-security guidelines and concepts applied in the Forschungszentrum Jülich are listed under http://www.fz-juelich.de/jsc/index.php?index=1026 – 02.11.2009 (please note that some files are not accessible from off-campus). In particular, the “IT Security Rules for Baseline Protection” and the “IT Security Guideline of Forschungszentrum Jülich” document this high standard of security measures. Overall, the IT security measures are in accordance with the requirements and standards outlined by the Federal Office for Information Security (BSI).See https://www.bsi.bund.de/ – 02.11.2009.

3. Conclusion

This study looked at three German institutional and subject repositories in the process of planning or implementing cooperative models for the long-term preservation of their digital collections. Thus, while none of the repositories aims to become a long-term archive itself, it is the goal of all three to build collections of digital objects suitable for long-term preservation in an archive, for which they will act as a content provider, and to provide access to these collections in uncorrupted, authentic form – a goal equally important for "traditional" repositories which seek to become trusted providers of scholarly information. As outlined above, although these repositories therefore are not necessarily full OAI-Systems meeting all requirements for compliance outlined in the reference model, they share important responsibilities with the preservation service provider in all of the OAIS functional entities.

During their own **Ingest** procedure, repositories must ascertain that they receive authentic and uncorrupted files from their producers, and that they have sufficient technical and legal control over these digital objects to include them in preservation workflows. Only if this foundation is laid by the repositories, preservation action can be taken in the future. Both pedocs and JUWEL were found to contain individual publications protected by DRM or similar technical restrictions. While no examples of such protected documents were found in Qucosa, it is likely that such cases exist nonetheless as submitted files are not checked for the presence of DRM measures. Of all three repositories, currently only pedocs uses a tool to detect possible technical restrictions.

Different procedures and techniques were employed to protect the integrity and authenticity of accepted files during ingest. However, only JUWEL uses a secure protocol in combination with a log-in procedure for submission and upload by authors. While the danger of someone interfering with the submission procedure was generally regarded as very low, pedocs and Qucosa might nonetheless want to consider using secure protocols to protect the information authors/depositors enter into the web submission forms.

During the generation of the (pre-)AIPs, only very little – if any – structural and context metadata was created. In particular with regard to the latter it seems advisable to consider developing a policy outlining which context metadata might be required for which types of publication, and how it should be collected. Whereas structural metadata (expressing, for example, hierarchical relations between journal articles, issues, and titles) might be generated retrospectively, context metadata can in many cases only be provided by the author/producer him- or herself. Another problem in this context concerned the question of versioning, which was partly not regarded relevant by repositories (JUWEL, Qucosa), but which might become an issue once objects migrated by the long-term archive are re-ingested into the repository and replace or complement the "original" digital object stored there. Accordingly, in pedocs the decision was made to implement a

versioning functionality which allows the linking of digital objects via their IDs and to describe the nature of the relationship with the help of Dublin Core metadata.

Different approaches were also taken to authenticating the source of material submitted to the repositories. Of all three repositories, as an institutional repository with a closed and clearly defined community of (potential) depositors, JUWEL is certainly in the most comfortable position, whereas both Qucosa and pedocs are in a more difficult situation as their communities of potential depositors are much bigger and in many cases much further removed (geographically, institutionally) from the repository, making communication a greater challenge, for example. Thus, although Qucosa is also a university-based institutional repository, potential depositors also come from a large number of different and very heterogeneous institutions throughout Saxony. In consequence, and to compensate for this lack of “closeness” between repository and depositors, Qucosa and pedocs use contracts in the form of author agreements to ascertain that depositors are who they claim to be and hence have the right to deposit the material in question.

In the functional entity of **Archival Storage**, the question of how each repository protected the integrity and authenticity of its digital objects is of particular interest and relevance – not only because the repositories must be able to protect their digital assets in the time span elapsing between submission by depositors and harvesting by or transmission to the long-term archive, but also because users accessing the digital collections must be certain at all times that they are viewing authentic, uncorrupted documents. In this context it was recommended that pedocs and Qucosa implement a tool or program similar to the DSpace checksum checker in order to make sure that any changes of checksums are detected as quickly as possible. In addition, all three repositories should develop a policy spelling out which steps are to be taken if the checksum of an object is found to have changed, and how this object will be treated both in relation to access and long-term preservation.

Of particular interest was also the question of metadata changes. In all three repositories metadata can be changed only by authorized staff. For reasons of transparency it seems nonetheless important that such changes are documented, e.g. with a time stamp and a history note explaining the nature of these changes. Such a documentation procedure has been implemented in pedocs and should be considered by the other repositories as well – in particular as the checksums, indicating the integrity and authenticity of the digital objects, are part of the metadata.

In the **Data Management** functional entity, none of the repositories worked with actual information packages in the sense of actual containers; instead, (virtual) AIPs are managed by means of the relational databases and DBMS of the repository software. Although a new DSpace functionality is currently being developed which will allow for the creation and storage of actual AIPs, whether this is a useful and necessary feature for a

repository focused primarily on the provision of access services is questionable as the relational databases seem sufficient to fulfill all necessary tasks. Thus, while the creation of actual packages might be required when exporting data to the long-term archive, it does not seem a necessity for the repositories' data management.

Administration: All three repositories have some kind of policy document expressing their commitment to long-term preservation in various degrees, depending on how far they have progressed in planning and implementing a long-term preservation workflow. It seems, however, that JUWEL in particular should consider to combine policy information contained, for example, in its Frequently Asked Questions or the Help page, in a single policy document reflecting both its commitment to Open Access and addressing relevant questions of long-term preservation. For all three repositories it seemed that additional policies or guidelines governing workflows and processes, often already existing in implicit form, should be “spelled out” (i.e. written down and made explicit) for the purpose of quality assurance.[168]

Among the most difficult tasks to be carried out by the repositories in the Administration functional entity appeared to be the definition of their designated communities (depositors and users) and their respective knowledge bases. While pedocs serves only to the community of a single (albeit strongly subdivided) discipline, both JUWEL and Qucosa cover a wide range of different disciplines; on the other hand, JUWEL has the advantage of a relatively closed community of depositors, whereas Qucosa collects submissions from a very heterogeneous group of Saxon institutions in higher education, business, and administration, and pedocs – as a subject repository – accepts submissions from authors in the field of educational science and research regardless of their affiliation. This difficulty to define the designated communities of the repositories has direct implications for P**reservation Planning** in that only a defined community can be monitored. While it seemed that both JUWEL and pedocs already have mechanisms and communication channels established (although these, it should be noted, are not necessarily “institutionalized” in all cases) allowing them to receive feedback from at least portions of their designated communities, Qucosa is in a much more difficult situation when it comes to monitoring its designated communities. As suggested, Qucosa might therefore attempt to focus on a “model” designated community, e.g. the users and depositors from the TU Dresden for which it serves as an institutional repository, while at the same time depending on the other submitting institutions to monitor the designated communities for their own publications.

In conclusion, it seems that the repositories considered in this study will face two major challenges in the future. On the one hand, the relation between the repositories and the long-term preservation service provider will have to be defined thoroughly, including

168 Some guidance as to institutional preservation policies is provided by Beagrie et al. (2008), or the OpenDOAR Policies Tool, which covers policies on submission, metadata, and preservation among others. See http://www.opendoar.org/tools/en/policies.php – 03.11.2009.

communication channels, responsibilities, and workflows – in particular those concerning the submission of information packages to the long-term archive and the reingest of objects into the repository after they were, for example, migrated by the long-term archive.[169] The second, possibly even more significant challenge for the repositories consists in balancing the sometimes conflicting requirements of long-term preservation and user-orientation. Thus, as Hitchcock et al. observe, "IRs are some way from being able to impose on authors content creation rules to support preservation" (2007; no pag.), and the same is certainly the case for subject repositories. It seems similarly true that for institutional and subject repositories "the focus on Long-Term [sic] preservation could be viewed to sideline other, perhaps more central, business requirements and could act as a barrier, rather than an enabler, particularly if preservation activity might slow repository population or incur additional costs" (Allinson 2006, 12). These appear to be valid observations and concerns in that a repository which "deters" authors from submitting material by overly elaborate requirements (e.g. concerning file format and metadata) may risk losing support both from the financing body and the user community. It follows that despite the utmost importance of considerations relating to the long-term preservation of the digital assets collected by institutional and subject repositories, it is crucial for the latter to also focus on the needs and concerns of their user communities. Only by doing so can repositories contribute to the continuing growth and use of their digital collections, thus ensuring that they themselves do not become obsolete in the future.

What is true for repositories – namely, that they should not allow long-term preservation concerns to interfere with usability more than absolutely necessary, so as not to create any barriers to using the repository and its collections – is, in slightly different form, also true for repository software. Thus, although the concern with long-term preservation is evidently growing in the developer communities, neither software studied and discussed in this work already provides a ready-to-use suite of long-term preservation features. Instead, such features often have to be built and implemented, sometimes with considerable programming effort, which proves to be an obstacle to entering long-term preservation. In consequence, repository managers not yet interested in long-term preservation or not yet aware of its importance might find in the future that a considerable part of their collections is not suitable for preservation, e.g. because information that could only have been provided by authors is missing, because files are protected by DRM, because complex relations between digital objects cannot be expressed, or because it is not clear if the digital objects are still authentic and uncorrupted. Only if the available software as a matter of course supports repository managers in preparing their collections for implementing long-term preservation measures and strategies, for example, by

169 In this context, emulation as a preservation strategy seems to pose a particular challenge as it is not certain whether the repository will be capable of providing an environment in which the emulated digital objects can be accessed and viewed. Although this might not concern textual documents so much, it might become an issue where multimedia objects and applications are accepted for publication in the repository, as is, for example, the case for Qucosa.

collecting and documenting relevant information and by monitoring the integrity and authenticity of the digital objects, more of them will feel confident to rise to the challenge of long-term preservation.

Works Cited

Allinson, Julie (2006): OAIS as a Reference Model for Repositories. An Evaluation. Version 0.5. Digital Repositories Programme Support Team. UKOLN, University of Bath. http://www.ukoln.ac.uk/repositories/publications/oais-evaluation-200607/Drs-OAIS-evaluation-0.5.pdf. Last accessed on 29.10.2009.

Altenhöner, Reinhard (2007): "kopal goes live". Nutzungsmodelle und Perspektiven eines Langzeitarchivs digitaler Informationen. Vorstellung der Projektergebnisse. http://kopal.langzeitarchivierung.de/downloads/kopal-goes-live_kopal-Nutzungsmodelle_und_Perspektiven_Altenhoener.pdf. Last accessed on 18.10.2009.

Association for Library Collections and Technical Services (ALCTS) (2007): Definition of Digital Preservation. http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.pdf. Last accessed on 29.10.2009.

Ball, Rafael (2003): Open Access: Die Revolution im wissenschaftlichen Publizieren? Vortrag von Dr. Rafael Ball im Rahmen des FZJ-Kolloquiums am 30. April 2003. Forschungszentrum Jülich, Zentralbibliothek. http://www.fz-juelich.de/zb/datapool/page/534/Vortrag%20Open%20Access.pdf. Last accessed on 17.10.2009.

BASE: Bielefeld Academic Search Engine. Universitätsbibliothek Bielefeld. http://www.base-search.net/. Last accessed on 03.11.2009.

Bass, Michael J. et al. (2002): DSpace: A Sustainable Solution for Institutional Digital Asset Services. Spanning the Information Asset Value Chain: Ingest, Manage, Preserve, Disseminate. Hewlett-Packard Company, Massachusetts Institute of Technology. Cambridge, MA. http://libraries.mit.edu/dspace-mit/technology/functionality.pdf. Last accessed on 02.11.2009.

Beagrie, Neil et al. (2008): Digital Preservation Policies Study. Part 1: Final Report October 2008. Charles Beagrie Limited. http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf. Last accessed on 02.11.2009.

Becker, Hans-Georg (2008): Vertrauenswürdige digitale Langzeitarchivierung an einer Universitätsbibliothek. Master's Thesis. Cologne University of Applied Sciences, Institute of Information Science. Unpublished.

Bildungsserver Blog (2009): Verlage beim Workshop "Open Access Erziehungswissenschaften." 25.08.2009. http://blog.bildungsserver.de/?p=269. Last accessed on 03.11.2009.

Bodleian Library (2007): Paradigm Workbook on Personal Digital Archives. University of Oxford. Also available online: http://www.paradigm.ac.uk/workbook/index.html. Last accessed on 02.11.2009.

Brace, Jenny (2008): Versioning in Repositories. Implementing Best Practice. In: Ariadne 56 (2008). http://www.ariadne.ac.uk/issue56/brace/. Last accessed on 02.11.2009.

British Library of Political and Economic Science (2008): Version Identification Framework (VIF) Homepage. http://www2.lse.ac.uk/library/vif/. Last updated on 04.03.2008. Last accessed on 03.11.2009.

Bundesministerium der Justiz: JURIS. Gesetze im Internet. http://bundesrecht.juris.de/. Last accessed on 03.11.2009.

CARPET (Community for Academic Reviewing, Publishing, and Editorial Technology) Project Homepage. http://www.carpet-project.net/. Last accessed on 03.11.2009.

Center for Research Libraries (CRL); OCLC Online Computer Center, Inc (Ed.) (February 2007): Trustworthy Repositories Audit and Certification. Criteria and Checklist. Version 1.0. http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf. Last accessed on 30.10.2009.

Consultative Committee for Space Data Systems (CCDS) (2002): Reference Model for an Open Archival Information System (OAIS). CCSD 6 50.0-B-1 Blue Book, Issue 1. Washington, DC. http://public.ccsds.org/publications/archive/650x0b1.pdf. Last accessed on 30.10.2009.

Consultative Committee for Space Data Systems (CCDS) (2002): Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-P-1.1 Pink Book, May 2009. Version for public examination to identify errors. http://cwe.ccsds.org/moims/docs/MOIMS-DAI/Draft%20Documents/OAIS-candidate-V2-markup.pdf. Last accessed on 02.11.2009.

Cornell University Library: Digital Preservation Management. Implementing Short-Term Strategies for Long-Term Problems. Tutorial. http://www.icpsr.umich.edu/dpm/dpm-eng/eng_index.html. Last accessed on 03.11.2009.

Deutsche Initiative für Netzwerkinformation e.V. (DINI) Homepage. http://www.dini.de/. Last accessed on 03.11.2009.

Deutsche Initiative für Netzwerkinformation e.V. (DINI) (2007): DINI-Zertifikat: Dokumenten- und Publikationsservice 2007. Version 2.1. http://nbn-resolving.de/urn:nbn:de:kobv:11-10079197. Last accessed on 02.11.2009.

Deutsche Initiative für Netzwerkinformation (DINI) e.V.: Open Access Netzwerk. http://www.dini.de/projekte/oa-netzwerk/. Last accessed on 03.11.2009.

Deutsches Institut für Internationale Pädagogische Forschung (DIPF) Homepage. http://www.dipf.de/de . Last accessed on 03.11.2009.

Deutsche Nationalbibliothek: Persistent Identifier. Eindeutiger Bezeichner für Digitale Inhalte. http://www.persistent-identifier.de/. Last accessed on 03.11.2009.

Deutsche Nationalbibliothek: URN-Service. http://www.d-nb.de/netzpub/erschl_lza/np_urn.htm. Last accessed on 03.11.2009.

Die Deutsche Bibliothek (2005): LMER. Long-term preservation Metadata for Electronic Resources. Version: 1.2. Leipzig, Frankfurt, Berlin. http://www.d-nb.de/standards/pdf/lmer12_e.pdf. Last accessed on 30.10.2009.

DigiCULT Project (Ed.) (2002): DigiCULT. Integrity and Authenticity of Digital Cultural Heritage Objects. Thematic Issue 1. http://www.digicult.info/downloads/thematic_issue_1_final.pdf. Last accessed on 02.11.2009.

Digital Curation Centre (DCC) Homepage. http://www.dcc.ac.uk/. Last accessed on 03.11.2009.

Digital Curation Centre; Digital Preservation Europe (2007): DCC and DPE Digital Repository Audit Method Based on Risk Assessment, v1.0. http://www.repositoryaudit.eu/download. Last accessed on 30.10.2009.

Digital Curation Centre; DigitalPreservationEurope; nestor; Center for Research Libraries: Core Requirements for Digital Preservation Repositories. http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re. Last accessed on 30.10.2009.

DigitalPreservationEurope (DPE) Homepage. http://www.digitalpreservationeurope.eu/. Last accessed on 03.11.2009.

DigitalPreservationEurope (DPE): Planning Tool for Trusted Electronic Repositories. http://www.digitalpreservationeurope.eu/platter/. Last accessed on 30.10.2009.

Digital Production and Integration Program (DPIP) (2008): Best Practices for Structural Metadata. Version 1. Yale University Library. http://www.library.yale.edu/dpip/bestpractices/BestPracticesForStructuralMetadata.pdf. Last accessed on 30.10.2009.

Digital Repository Audit and Certification Wiki. http://wiki.digitalrepositoryauditandcertification.org/bin/view. Last accessed on 30.10.2009.

Dobratz, Susanne et al. (2008): The Use of Quality Management Standards in Trustworthy Digital Archives. In: The British Library Board (Ed.): iPres 2008. Proceedings of The Fifth International Conference on Preservation of Digital Objects. Joined Up and Working: Tools and Methods for Digital Preservation. The British Library, London. 29-30 September: 205-212. http://www.bl.uk/ipres2008/ipres2008-proceedings.pdf. http://nbn-resolving.de/urn:nbn:de:kobv:11-10092248 (unpaginated). Last accessed: 02.11.2009.

Dobratz, Susanne; Schoger, Astrid (2005): Digital Repository Certification: A Report from Germany. Also published in: RLG DigiNews 9.5(2005). http://www.rlg.org/en/page.php?Page_ID=12081. http://nbn-resolving.de/urn:nbn:de:kobv:11-10063181. Last accessed on 30.10.2009.

DRIVER (Digital Repository Infrastructure Vision for European Research) Homepage. http://www.driver-community.eu/. Last accessed on 03.11.2009.

DROID (Digital Record Object Identification) Project Page. http://droid.sourceforge.net/. Last accessed on 03.11.2009.

DSpace Homepage. http://www.dspace.org/. Last accessed on 03.11.2009.

DSpace Wiki (2008a): BitstreamFormat Renovation. http://wiki.dspace.org/index.php/BitstreamFormat_Renovation. Last updated on 08.04.2008. Last accessed on 30.10.2009.

DSpace Wiki (2008b): DSpaceMETSSIPProfile. http://wiki.dspace.org/index.php/DSpaceMETSSIPProfile. Last updated on 02.05.2008. Last accessed on 25.10.2009.

DSpace Wiki (2009a): AssetStore. http://wiki.dspace.org/index.php/AssetStore. Last updated on 09.02.2009. Last accessed on 30.10.2009.

DSpace Wiki (2009b): ContributionGuidelines. http://wiki.dspace.org/index.php/ContributionGuidelines. Last updatd on 15.04.2009. Last accessed on 03.11.2009.

Dublin Core Metadata Initiative (DCMI) Homepage. http://dublincore.org/. Last accessed on 03.11.2009.

Dublin Core Metadata Initiative (DCMI) (2004): DC-Library Application Profile. http://dublincore.org/documents/library-application-profile/. Last accessed on 30.10.2009.

Dublin Core Metadata Initiative (DCMI) (2005a): Using Dublin Core – Dublin Core Qualifiers. http://dublincore.org/documents/2005/11/07/usageguide/qualifiers.shtml. Last accessed on 25.10.2009.

Dublin Core Metadata Initiative (DCMI) (2005b): Using Dublin Core – The Elements. http://dublincore.org/documents/usageguide/elements.shtml. Last updated on 28.08.2006. Last accessed on 30.10.2009.

EDItEUR: ONIX Homepage. http://www.editeur.org/8/ONIX/. Last accessed on 03.11.2009.

Efler-Mikat, Daniela (2009): Synopse der Lehrpläne der deutschen Bundesländer für das Fach Sachunterricht in der Grundschule. Dokument 2.pdf. http://www.pedocs.de/volltexte/2009/807/. Last accessed on 30.10.2009.

Factor, Michael et al. (2009): Authenticity and Provenance in Long Term Digital Preservation. Modeling and Implementation in Preservation Aware Storage. Presentation given at the First Workshop on the Theory and Practice of Provenance. San Francisco, California, USA. http://www.usenix.org/events/tapp09/tech/full_papers/factor/factor.pdf. Last accessed on 30.10.2009.

Fachportal Pädagogik. The German Education Portal (English Homepage). http://www.fachportal-paedagogik.de/start_e.html. Last accessed on 03.11.2009.

Fish, Sands Alden (2008): BitstreamFormat Renovation. DSpace Gets Real Technical Metadata. Presentation given at Open Repositories Conference 2008. Southampton, UK. http://mit.edu/sands/www/bfr/Sands%20Fish%20-%20Bitstream%20Format%20Renovation.ppt. Last accessed on 30.10.2009.

Forschungszentrum Jülich Homepage. http://www.fz-juelich.de/portal/. Last accessed on 03.11.2009.

Open Access Model of the Forschungszentrum Jülich. http://www.fz-juelich.de/zb/index.php?index=758. Last accessed on 03.11.2009.

Funk, Stefan et al. (2007): kopal Library for Retrieval and Ingest. Documentation. v1.0. Project kopal. German National Library / Goettingen State and University Library. http://kopal.langzeitarchivierung.de/kolibri/koLibRI_v1_0_documentation.pdf. Last accessed on 30.10.2009.

Gesetz über Rahmenbedingungen für elektronische Signaturen (SigG) (16.05.2001). Signaturgesetz vom 16. Mai 2001 (BGBl. I S. 876), das zuletzt durch Artikel 4 des Gesetzes vom 17. Juli 2009 (BGBl. I S. 2091) geändert worden ist. http://www.gesetze-im-internet.de/sigg_2001/index.html. Last accessed on 30.10.2009.

Harnad, Stevan et al. (2004): The Green and the Gold Roads to Open Access. In: Nature Web Focus: Access to the Literature. Nature Publishing Group. http://www.nature.com/nature/focus/accessdebate/21.html. Last accessed on 30.10.2009.

Harvard University Library; NARA; OCLC: Global Digital Format Registry. http://www.gdfr.info/index.html. Last accessed on 03.11.2009.

Herb, Ulrich; Kersting, Anja; Leidinger, Tobias (2008): Vernetzung von fachlichen und institutionellen Open-Access-Repositorien. Pilotversuch zum Austausch von Metadaten zwischen KOPS, dem institutionellen Repository der Universität Konstanz, und PsyDok, dem fachlichen Repository der Saarländischen Universitäts- und Landesbibliothek im Bereich Psychologie. In: Bibliotheksdienst, 42.5(2008): 550-555. http://www.zlb.de/aktivitaeten/bd_neu/heftinhalte2008/DigitaleBibliothek010508BD.pdf. Last accessed on 30.10.2009.

Hinz, Waldemar (2008): JUWEL. Open Access Server des Forschungszentrums Jülich. DSpace-Anwender-Workshop, 19.11.2008. Göttingen. http://hdl.handle.net/2003/25943. Last accessed on 30.10.2009.

Hitchcock, Steve et al. (2007): Digital Preservation Service Provider Models for Institutional Repositories. Towards Distributed Services. In: D-Lib Magazine, 13.5/6(2007). doi:10.1045/may2007-hitchcock. Last accessed on 30.10.2009.

IBM: Implementation Services. Digital Information Archiving System (DIAS). http://www-935.ibm.com/services/nl/dias/is/implementation_services.html. Last accessed on 24.10.2009.

IBM: Implementation Services. Preservation Manager. http://www-935.ibm.com/services/nl/dias/is/implementation_services.html. Last accessed on 24.10.2009.

Illinois Digital Environment for Access to Learning and Scholarship Wiki: Digital Preservation. https://services.ideals.uiuc.edu/wiki/bin/view/IDEALS/Internal/IdealsPreservation. Last accessed on 03.11.2009.

JISC (Joint Information Systems Committee) Homepage. http://www.jisc.ac.uk. Last accessed on 03.11.2009.

JISC (Joint Information Systems Committee): Digital Preservation and Curation. http://www.jisc.ac.uk/whatwedo/topics/digitalpreservation.aspx. Last accessed on 03.11.2009.

JISC Repositories Ideascale Page. http://jiscrepository.ideascale.com/. Last accessed on 03.11.2009.

Jantz, Ronald (2009): Letter to the Editor. Re: Authentic Digital Objects. In: International Journal of Digital Curation, 4.2(2009): 8-11. http://www.ijdc.net/index.php/ijdc/article/view/114/101. Last accessed on 24.10.2009.

JSTOR; Harvard College (2009): JHOVE - JSTOR/Harvard Object Validation Environment Homepage. http://hul.harvard.edu/jhove/. Last updated on 25.02.2009. Last accessed on 03.11.2009.

JUWEL. JUelicher Wissenschaftliche Elektronische Literatur. Zentralbibliothek des Forschungszentrum Jülich. http://juwel.fz-juelich.de:8080/dspace/. Last accessed on 03.11.2009.

Kingsley, Danny (2008): Those who don't look don't find: Disciplinary Considerations in Repository Advocacy. Preprint of article published in: OCLC Systems and Services: International Digital Library Perspective, 24.4(2008). http://hdl.handle.net/1885/46229. Last accessed on 29.10.2009.

Knight, Gareth (Feb 2008): Framework for the definition of significant properties. Work Package 3.3, v.1. http://www.significantproperties.org.uk/documents/wp33-propertiesreport-v1.pdf. Last accessed on 20.07.2009.

Knight, Gareth; Hedges, Mark (2007): Modelling OAIS Compliance for Disaggregated Preservation Services. In: International Journal of Digital Curation, 2.1(2007): 62-72. http://www.ijdc.net/index.php/ijdc/article/view/25/14. Last accessed on 31.10.2009.

kopal Homepage: Kooperativer Aufbau eine Langzeitarchivs digitaler Informationen. http://kopal.langzeitarchivierung.de/. Last accessed on 03.11.2009.

Lavoie, Brian; Dempsey, Lorcan (2004): Thirteen Ways of Looking at…Digital Preservation. In: D-Lib Magazine, 10. 7/8(2004). doi:10.1045/july2004-lavoie. Last accessed on 29.10.2009.

Ludwig, Jens (2007): koLibRI. Workflows und Tools. http://kopal.langzeitarchivierung.de/downloads/kopal-goes-live_koLibRI_Ludwig.pdf. Last accessed on 24.10.2009.

Lynch, Clifford A. (2000): Authenticity and Integrity in the Digital Environment: An Exploratory Analysis of the Central Role of Trust. In: Authenticity in a Digital Environment. Council on Library and Information Resources (Ed.). Washington, DC. http://www.clir.org/pubs/reports/pub92/contents.html. Last accessed on 02.11.2009.

Lynch, Clifford A. (2003): Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. In: ARL – A Bimonthly Report, 226(2003). http://www.arl.org/resources/pubs/br/br226/br226ir.shtml. Last accessed on 02.11.2009.

Marahrens, Oliver (2005): Prüfung und Erweiterung der technischen Grundlagen des Dokumentenservers OPUS zur Zertifizierung gegenüber der DINI anhand der Installation an der TU Hamburg- Harburg. Projektbericht für das Projektstudium an der Virtuellen Fachhochschule, Standort Lübeck. http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:gbv:830-opus-827. Last accessed on 31.10.2009.

Marill, Jennifer L.; Luczak, Edward C. (2009): Evaluation of Digital Repository Software at the National Library of Medicine. In: D-Lib Magazine, 15.5/6(2009). doi:10.1045/may2009-marill. Last accessed on 16.10.2009.

Müller, Uwe et al. (2009): OA Network. An Integrative Open Access Infrastructure for Germany. In: D-Lib Magazine 15. 9/10(2009). doi:10.1045/september2009-mueller. Last accessed on 30.10.2009.

National Information Standards Organization (NISO) (2004): Understanding Metadata. Bethesda, MD. http://www.niso.org/publications/press/UnderstandingMetadata.pdf. Last accessed on 02.11.2009.

National Library of Medicine (2009): Development of a Digital Repository for NLM Digitized Collections and Born-Digital Resources. http://www.nlm.nih.gov/digitalrepository/. Last updated on 02.07.2009. Last accessed on 16.10.2009.

National Library of Medicine Digital Repository Evaluation and Selection Work Group (NLM-DRESWG) (2009): Recommendations on NLM Digital Repository Software. http://www.nlm.nih.gov/digitalrepository/DRESWG-Report.pdf. Last accessed on 16.10.2009.

National Library of Medicine Digital Repository Working Group (NLM-DRWG) (2008): Policies and Functional Requirements Specification. Version 1. http://www.nlm.nih.gov/digitalrepository/NLM-DigRep-Requirements-rev032007.pdf. Last accessed on 16.10.2009.

nestor Homepage: Kompetenznetzwerk Langzeitarchivierung. http://www.langzeitarchivierung.de/. Last accessed on 03.11.2009.

nestor c/o Deutsche Nationalbibliothek (Ed.) (2008): nestor-Kriterien. Kriterienkatalog vertrauenswürdige digitale Langzeitarchive Version II. nestor-Materialien 8. Frankfurt am Main. http://nbn-resolving.de/urn:nbn:de:0008-2008021802.

Neuroth, Heike et al. (Ed.) (2009): nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0. Boizenburg: Verlag Werner Hülsbusch.

OpenDOAR: The Directory of Open Access Repositories. http://www.opendoar.org/. Last accessed on 01.11.2009.

OpenDOAR: Directory of Open Access Repositories: Policies Tool. http://www.opendoar.org/tools/en/policies.php. Last accessed on 03.11.2009.

pedocs. Deutsches Institut für Internationale Pädagogische Forschung (DIPF). http://www.pedocs.de/. Last accessed on 03.11.2009.

PRESERV Homepage: Repository Preservation and Interoperability. http://preserv.eprints.org/. Last accessed on 03.11.2009.

Preservation and Long-Term Access Through Networked Services (PLANETS) Project Homepage. http://www.planets-project.eu/. Last accessed on 03.11.2009.

Qucosa (Quality Content of Saxony). Sächsische Landesbibliothek - Staats- und Universitätsbibliothek (SLUB) Dresden. http://www.qucosa.de. Last accessed on 03.11.2009.

Research Libraries Group; OCLC Online Computer Center (RLG-OCLC) (2002): Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report. Mountain View, CA. http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf. Last accessed on 30.10.2009.

ROAR: Registry of Open Access Repositories. http://roar.eprints.org/. Last accessed on 01.11.2009.

Rothenberg, Jeff (1999): Ensuring the Longevity of Digital Information. http://www.clir.org/pubs/archives/ensuring.pdf. Last updated on 22.02.1999. Last accessed on 29.10.2009.

Rusbridge, Chris (2009a): Repositories and Preservation. 23 February 2009. In: Rusbridge, Chris: Digital Curation Blog. Blog inspired by the Digital Curation Centre to discuss issues relating to the curation and long term preservation of digital science and research data. http://digitalcuration.blogspot.com/2009/02/repositories-and-preservation.html. Last accessed on 29.10.2009.

Rusbridge, Chris (2009b): Repository Preservation Revisited. 9 March 2009. In: Rusbridge, Chris: Digital Curation Blog. Blog inspired by the Digital Curation Centre to discuss issues relating to the curation and long term preservation of digital science and research data. http://digitalcuration.blogspot.com/2009/03/repository-preservation-revisited.html. Last accessed on 29.10.2009.

Salo, Dorothea (2007): Inkeeper at the Roach Motel. (Postprint). Also published in: Library Trends 57.2(2008). http://digital.library.wisc.edu/1793/22088. Last accessed on 29.10.2009.

Scholze, Frank; Summann, Friedrich (2009): Forschungsinformationen und Open Access Repository-Systeme. In: Wissenschaftsmanagement, 15.3(2009): 41-42.

Siermann, Barbara (2008): Long-term preservation for institutional repositories. In: Weenink, Kasja; Waaijers, Leo; van Godtsenhoven, Karen (Ed.): A DRIVER's Guide to European Repositories. Amsterdam: Amsterdam University Press:153-184. http://dare.uva.nl/aup/en/record/260224. Last accessed on 29.10.2009.

Social Science Open Access Repository (SSOAR). http://www.ssoar.info/. Last accessed on 03.11.2009.

Steinhart, Gail; Dietrich, Dianne; Green, Anne (2009): Establishing Trust in a Chain of Preservation. The TRAC Checklist Applied to a Data Staging Repository (DataStaR). In: D-Lib Magazine, 15.9/10(2009). doi:10.1045/september2009-steinhart. Last accessed on 02.11.2009.

Steinke, Tobias (2006): Universal Object Format. An Archiving and Exchange Format for Digital Objects. http://kopal.langzeitarchivierung.de/downloads/kopal_Universal_Object_Format.pdf. Last accessed on 02.11.2009.

Steinke, Tobias (2008): Erfahrungen mit kopal und digitaler Langzeitarchivierung. Vortrag im Rahmen des 97. Deutschen Bibliothekartags, Mannheim. http://www.opus-bayern.de/bib-info/volltexte/2008/615/. Last accessed on 02.11.2009.

Stone, Larry (2008): BitstreamFormat Renovation. DSpace Gets Real Technical Metadata. Presentation given at Open Repositories Conference 2008. Southampton, UK. http://wiki.dspace.org/static_files/a/a3/BSF-or08.pdf. Last accessed on 02.11.2009.

Tanner, Simon; Muñoz, Trevor; Ros, Pich Hemy (2009): Measuring Mass Text Digitization Quality and Usefulness. Lessons Learned from Assessing the OCR Accuracy of the British Library's 19th Century Online Newspaper Archive. In: D-Lib Magazine, 15.7/8(2009). doi:10.1045/july2009-munoz. Last accessed on 02.11.2009.

Tansley, Robert et al. (2006): DSpace 1.4.1. beta 1 System Documentation. http://wiki.dspace.org/static_files/b/be/DSpaceStandard141beta1.pdf. Last accessed on 02.11.2009.

Thibodeau, Kenneth (2007): If you build it, will it fly? Criteria for success in a digital repository. In: Journal of Digital Information, Jg. 8, H. 2. http://journals.tdl.org/jodi/article/view/197/174. Last accessed on 29.10.2009.

Universität Stuttgart (2006): OPUS Dokumentation. http://elib.uni-stuttgart.de/opus/doku/dokumentation.php, zuletzt aktualisiert am 03.02.2006. Last accessed on 03.11.2009.

University of Nottingham: SHERPA/RoMEO. Publisher Copyright Policies and Self-Archiving. http://www.sherpa.ac.uk/romeo/. Last accessed on 03.11.2009.

van der Graaf, Maurits; van Eijndhoven, Kwame (2008): The European Repository Landscape. Inventory Study into the Present Type and Level of OAI-Compliant Digital Repository Activities

in the EU. Amsterdam: Amsterdam University Press. http://dare.uva.nl/aup/en/record/260225. Last accessed on 29.10.2009.

Weenink, Kasja; Waaijers, Leo; van Godtsenhoven, Karen (Ed.) (2008): A DRIVER's Guide to European Repositories : Five studies of important Digital Repository-Related Issues and Good Practices. Amsterdam: Amsterdam University Press. http://dare.uva.nl/aup/en/record/260224. Last accessed on 02.11.2009.

Wilson, Andrew (2007): Significant Properties Report. InSPECT Work Package 2.2, v.2. http://www.significantproperties.org.uk/documents/wp22_significant_properties.pdf. Last accessed on 02.11.2009.

Wilson, Andrew (2009): Letter to the Editor. Authentic Digital Objects. In: International Journal of Digital Curation, 4.2(2009): 4-7. http://www.ijdc.net/index.php/ijdc/article/view/113/100. Last accessed on 16.10.2009.

Winkler, Marco (2008): Langzeitarchivierung von Online-Publikationen digitaler Repositorien. Untersuchung am Beispiel der Publikationssoftware OPUS. Diplomarbeit zur Erlangung des akademischen Grades Diplom-Dokumentar (FH). Fachhochschule Potsdam, Fachbereich Informationswissenschaften. Unpublished.

Appendix A: Mapping of Criteria Catalogs

In the following, an attempt has been made to reorganize the criteria from the nestor, TRAC, and DINI catalogs so as to allow them to be mapped onto the OAIS functional entities Ingest, Archival Storage, Data Management, Preservation Planning, and Administration. In addition, as some criteria are of a more general nature, concerning not single areas in the OAIS functional model but the entire repository or repository system, these have been assigned to a separate category of higher-level requirements, comparable to the “common services” identified in the OAIS model (see above).

In the attempt of mapping, problems sometimes occurred where criteria were either too general or too specific to be assigned to one of the OAIS functional entities; equally problematic were criteria that concerned more than one functional area. In consequence, the relationship between OAIS functional entities and sub-entities and a given criterion is sometimes not entirely straightforward and sometimes only makes a partial match. Other criteria (marked with an asterisk) appear in more than one place in the tables below.

Further problems concerned the relationship between superordinate and subordinate criteria in the nestor catalog, i.e. criteria with full and decimal numbers respectively. For the purpose of this mapping, it was assumed that a criterion with a full number was fulfilled when all sub-criteria were fulfilled. In consequence, these superordinate criteria are listed below only in the “Common Services” table and in exceptional cases, e.g. where either no sub-criteria existed.

Please note that the tables do not accomplish a one to one matching of criteria of the three catalogs in all cases as this would have made the tables unnecessarily complex (particular in the case of the criteria on metadata). Thus, the matching primarily takes place by categorizing criteria under the same OAIS functional sub-entity.

Ingest

nestor	TRAC	DINI
Receive Submission		
9.3 Das dLZA erhält die technische Kontrolle über die digitalen Objekte, um Langzeitarchivierungs-maßnahmen durchführen zu können.	B1.5 Repository obtains sufficient physical control over the digital objects to preserve them.	2.8. Langzeitverfügbarkeit *Mindeststandard* - Die gegebenenfalls zusätzlich zu den eingereichten Originaldateien des Autors erstellten Archivkopien sind frei von Schutzmaßnahmen (DRM), die eine Anwendung von Strategien zur Langzeit-verfügbarkeit (Migration, Emulation) verhindern. *Empfehlungen* - Nutzung von offenen Dateiformaten, die zur Langzeitarchivierung geeignet (z. B. PDF/A, ODF, TXT, HTML, TEX) und frei von Schutzmaßnahmen (DRM) sind.
Quality Assurance		
6.1 Aufnahme (Ingest): Das dLZA sichert die Integrität der digitalen Objekte. 7.1 Aufnahme (Ingest): Das dLZA sichert die Authentizität der digitalen Objekte.	B1.4 Repository's ingest process verifies each submitted object (i.e., SIP) for completeness and correctness as specified in B1.2.[170] B1.3 Repository has mechanisms to authenticate the source of all materials.	2.5 Sicherheit, Authentizität und Integrität 2.5.1 Server *Mindeststandard* - Kontrollierte und nachweisbare Aufnahme von Dokumenten aus technischer Sicht. 2.5.2 Dokumente *Empfehlungen* - Fortgeschrittene digitale Signatur nach § 2 Abs. 2 SigG 2001 wird verwendet.
Generate AIP		
10.2 Das dLZA sorgt für eine Transformation der Übergabepakete in Archivpakete. 12.1 Das dLZA identifiziert seine Objekte und deren Beziehungen eindeutig und dauerhaft. 12.3 Das dLZA erhebt in ausreichendem Maße Metadaten zur strukturellen Beschreibung der digitalen Objekte. 12.4 Das dLZA erhebt in ausreichendem Maße Metadaten, die alle vom digitalen Langzeitarchiv vorgenommenen Veränderungen an den digitalen Objekten verzeichnen. 12.5 Das dLZA erhebt in ausreichendem Maße Metadaten zur technischen Beschreibung der digitalen Objekte.	B2.3 Repository has a description of how AIPs are constructed from SIPs. B2.5 Repository has and uses a naming convention that generates visible, persistent, unique identifiers for all archived objects (i.e., AIPs). B2.6 If unique identifiers are associated with SIPs before ingest, the repository preserves the identifiers in a way that maintains a persistent association with the resultant archived object (e.g., AIP). B2.9 Repository acquires preservation metadata (i.e., PDI) for its associated Content Information.	2.5 Sicherheit, Authentizität und Integrität 2.5.2 Dokumente *Mindeststandard* - Verwendung von Persistent Identifiers, dazu zählen Systeme, die einen Resolver-Dienst besitzen, z. B. urn:nbn oder DOI. 2.6.2 Metadatenexport *Mindeststandard*: - Metadaten sind nach Dublin Core Simple (ISO 15836:2003) strukturiert. *Empfehlungen*: - Metadaten sind nach Dublin Core Qualified strukturiert. - Metadaten sind nach ONIX strukturiert. - Technische und/oder Archivierungsmetadaten […] werden angeboten (z. B. PREMIS, LMER).

170 "B1.2 Repository clearly specifies the information that needs to be associated with digital material at the time of its deposit (i.e., SIP)" (TRAC 2007).

Generate AIP ctd.		
12.6 Das dLZA erhebt in ausreichendem Maße Metadaten, die die entsprechenden Nutzungsrechte und -bedingungen verzeichnen.	B2.7 Repository demonstrates that it has access to necessary tools and resources to establish authoritative semantic or technical context of the digital objects it contains (i.e., access to appropriate international Representation Information and format registries). B2.8 Repository records/registers Representation Information (including formats) ingested.	2.8 Langzeitverfügbarkeit *Empfehlungen*: - Erstellung von technischen Metadaten zur Langzeitarchivierung (z. B. mit dem Tool JHOVE). - Eindeutige Identifizierung des jeweiligen Dateiformats in den Metadaten mit Referenzen zu öffentlich zugänglichen File Format Registries.
Generate Descriptive Info		
12.2 Das dLZA erhebt in ausreichendem Maße Metadaten für eine formale und inhaltliche Beschreibung und Identifizierung der digitalen Objekte.	B5.2 Repository captures or creates minimum descriptive metadata and ensures that it is associated with the archived object (i.e., AIP).	2.6 Erschließung 2.6.1 Sacherschließung *Mindeststandard*: - Verbale Sacherschließung durch freie Schlagwörter oder klassifikatorische Erschließung wird durchgeführt. - Dewey-Dezimalklassifikation (DDC) gemäß der Verwendung in der Deutschen Nationalbibliografie als allgemeine klassifikatorische Erschließung aller Dokumente (entsprechend den DINI-OAI-Empfehlungen) wird angewandt. Empfehlungen: Mindestens ein weiteres normiertes System verbaler oder klassifikatorischer Erschließung (allgemein oder fachspezifisch […]) wird verwendet. Englischsprachige Schlagwörter werden vergeben. Kurzzusammenfassungen / Abstracts in Deutsch und Englisch werden angeboten.

Archival Storage

nestor	TRAC	DINI
Receive Data, Manage Storage Hierarchy		
Replace Media		
*8 Das dLZA betreibt eine langfristige Planung seiner technischen Langzeiterhaltungs-maßnahmen.	C1.7 Repository has defined processes for storage media and/or hardware change (e.g., refreshing, migration).	2.5 Sicherheit, Authentizität und Integrität *2.5.1 Server *Mindeststandard* - Das Betriebskonzept gewährleistet eine angemessene Verfügbarkeit des Systems. - Regelmäßige Wartung des Systems.
Error Checking and Disaster Recovery		
*6.2 Archivablage (Archival Storage): Das dLZA sichert die Integrität der digitalen Objekte. 7.2 Archivablage (Archival Storage): Das dLZA sichert die Authentizität der digitalen Objekte. 10.3 Das dLZA gewährleistet die Speicherung und Lesbarkeit der Archivpakete	B2.12 Repository provides an independent mechanism for audit of the integrity of the repository collection/content. B4.4 Repository actively monitors integrity of archival objects (i.e., AIPs). *C1.2 Repository ensures that it has adequate hardware and software support for backup functionality sufficient for the repository's services and for the data held, e.g., metadata associated with access controls, repository main content. C1.5 Repository has effective mechanisms to detect bit corruption or loss.	2.5 Sicherheit, Authentizität und Integrität *2.5.1 Server *Mindeststandard* - Einsatz einer Technologie zur Sicherung und Wiederherstellung der Server-Software, der Metadaten und der Dokumente mit täglicher Sicherung. - Sichere Installation des Systems und der Software-Komponenten. - Regelmäßige Wartung des Systems. *Empfehlungen* - SSL-Zertifizierung mit vertrauenswürdigem Zertifikat für verschlüsselte Kommunikation wird eingesetzt. 2.5.2 Dokumente *Mindeststandard*: - Archivierung der eingereichten Originaldateien des Autors auch im Ablieferungsformat. *Empfehlungen* - Einsatz eines Verfahrens zum Nachweis der Unversehrtheit der Dokumente (Hash-Wert) sowie Veröffentlichung von Verfahren und Hash-Werten. 2.8 Langzeitverfügbarkeit *Empfehlungen* - Sicherstellung der Langzeitverfügbarkeit, ggf. durch Kooperation mit einer Archivierungsinstitution.

Data Management

nestor	TRAC	DINI
12.7 Der Erhalt der Paketstruktur ist zu jeder Zeit gegeben.	B5.3 Repository can demonstrate that referential integrity is created between all archived objects (i.e., AIPs) and associated descriptive information. B5.4 Repository can demonstrate that referential integrity is maintained between all archived objects (i.e., AIPs) and associated descriptive information	2.8 Langzeitverfügbarkeit *Mindeststandard*: - Dauerhafte Verbindung der Metadaten mit den Dokumenten (z. B. Verbindung über Persistent Identifier oder zusammen in einem Container)
Administer Database, Perform Queries, Generate Report		
Receive Database Update		
	C1.3 Repository manages the number and location of copies of all digital objects. C1.4 Repository has mechanisms in place to ensure any/multiple copies of digital objects are synchronized.	

Administration

nestor	TRAC	DINI
Negotiate Submission Agreement		
3.1 Es bestehen rechtliche Regelungen zwischen Produzenten und dem digitalen Langzeitarchiv.	A5.2 Repository contracts or deposit agreements must specify and transfer all necessary preservation rights, and those rights transferred must be documented. A5.3 Repository has specified all appropriate aspects of acquisition, maintenance, access, and withdrawal in written agreements with depositors and other relevant parties.	2.4 Rechtliche Aspekte[171] *Mindeststandard* Bei Primärpublikation: - Es ist mit den Autoren eine Vereinbarung (Autorenvertrag) zu schließen, in der den Endnutzern die freie elektronische Verbreitung des Dokuments erlaubt wird und deren Bedingungen festgeschrieben sind. (Recht zur elektronischen Speicherung, insbesondere in Datenbanken, und zum Verfügbarmachen für die Öffentlichkeit zum individuellen Abruf, zur Wiedergabe auf dem Bildschirm und zum Ausdruck beim Nutzer [Online-Nutzung], auch auszugsweise) […].
Manage System Configuration		

171 DINI contains an extensive criterion on the questions of rights, predominantly focusing on copyright questions in the context of self-archiving. A statement covering the question of rights in the context of long term preservation is missing – such a statement, containing an outline of the rights that the repository must have in order to preserve the submitted digital object for the long term should be added.

Archival Information Update		
10.4 Das dLZA setzt Strategien zum Langzeiterhalt der Archivpakete um.	B4.1 Repository employs documented preservation strategies. B4.2 Repository implements/responds to strategies for archival object (i.e., AIP) storage and migration.	
Physical Access Control		
*6.2 Archivablage (Archival Storage): Das dLZA sichert die Integrität der digitalen Objekte.		2.5 Sicherheit, Authentizität und Sicherheit *2.5.1 Server *Mindeststandard* Es existiert eine Dokumentation des technischen Systems mit […] Zugangsregelung zum Server - räumlich - auf das System bezogen - personell (Verantwortlichkeit und Vertretung)
Establish Standards and Policies		
1.1 Das dLZA hat Kriterien für die Auswahl seiner digitalen Objekte entwickelt. 1.2 Das dLZA übernimmt die Verantwortung für den dauerhaften Erhalt der durch die digitalen Objekte repräsentierten Informationen. 1.3 Das dLZA hat seine Zielgruppe(n) definiert.	A1.1 Repository has a mission statement that reflects a commitment to the long-term retention of, management of, and access to digital information. A3.1 Repository has defined its designated community(ies) and associated knowledge base(s) and has publicly accessible definitions and policies in place to dictate how its preservation service requirements will be met. A3.2 Repository has procedures and policies in place, and mechanisms for their review, update, and development as the repository grows and as technology and community practice evolve. A3.3 Repository maintains written policies that specify the nature of any legal permissions required to preserve digital content over time, and repository can demonstrate that these permissions have been acquired when needed.	2.2 Leitlinien (Policy) *Mindeststandard* - Der Anbieter des Dokumenten- und Publikationsservice veröffentlicht Leitlinien für inhaltliche Kriterien sowie den Betrieb. Darin müssen die Rechte und Pflichten des Anbieters sowie der Autoren/Herausgeber der Dokumente festgeschrieben sein. - Die Policy muss enthalten: - Festlegungen der inhaltlichen, funktionalen und technischen Qualität der Dokumente, die auf einem Dokumentenserver veröffentlicht werden. - Eine Garantie für bestimmte Archivierungszeiträume in Abhängigkeit von der inhaltlichen, funktionalen und technischen Qualität der Dokumente. 2.5 Sicherheit, Authentizität und Integrität 2.5.2 Dokumente *Mindeststandard* - Ein inhaltlich verändertes Dokument ist wie ein neues Dokument zu behandeln (neuer Persistent Identifier). 2.6 Erschließung 2.6.1 Sacherschließung *Mindeststandard* - Eine Policy zur Sacherschließung muss vorhanden und dem Autor bekannt sein.

Establish Standards and Policies ctd.		
		2.8 Langzeitverfügbarkeit *Mindeststandard* - Definition einer Mindestzeit der Dokumentverfügbarkeit, die 5 Jahre nicht unterschreiten darf, ist in der Policy vorhanden. *Empfehlungen* - Eine Policy zum Umgang mit Löschungen von Dokumenten ist vorhanden.
9.2 Das dLZA identifiziert, welche Eigenschaften der digitalen Objekte für den Erhalt von Informationen signifikant sind.	B1.1 Repository identifies properties it will preserve for digital objects.	
Audit Submission		
6 Das digitale Langzeitarchiv stellt die Integrität der digitalen Objekte auf allen Stufen der Verarbeitung sicher. 7 Das digitale Langzeitarchiv stellt die Authentizität der digitalen Objekte auf allen Stufen der Verarbeitung sicher.	B2.11 Repository verifies each AIP for completeness and correctness at the point it is generated.	2.5 Sicherheit, Authentizität und Integrität *2.5.1 Server *Mindeststandard* - Kontrollierte und nachweisbare Aufnahme von Dokumenten aus technischer Sicht.
Activate Requests, Customer Service		

Preservation Planning

nestor	TRAC
Monitor Designated Community	
2.2 Das dLZA stellt die Interpretierbarkeit der digitalen Objekte durch seine Zielgruppe(n) sicher. *4.5 Das dLZA reagiert auf substanzielle Veränderungen.	A3.5 Repository has policies and procedures to ensure that feedback from producers and users is sought and addressed over time.
Monitor Technology[172]	
*4.4 Das dLZA betreibt eine langfristige Planung. 4.5 Das dLZA reagiert auf substanzielle Veränderungen. *8 Das dLZA betreibt eine langfristige Planung seiner technischen Langzeiterhaltungsmaßnahmen.	A3.4 Repository is committed to formal, periodic review and assessment to ensure responsiveness to technological developments and evolving requirements. C2.1 Repository has hardware technologies appropriate to the services it provides to its designated community(ies) and *has procedures in place to receive and monitor notifications, and evaluate when hardware technology changes are needed*. [emphasis added] C2.2 Repository has software technologies appropriate to the services it provides to its designated community(ies) and *has procedures in place to receive and monitor notifications, and evaluate when software technology changes are needed*. [emphasis added] B3.2 Repository has mechanisms in place for monitoring and notification when Representation Information (including formats) approaches obsolescence or is no longer viable.
Develop Preservation Strategies and Standards	
*4.4 Das dLZA betreibt eine langfristige Planung. *8 Das dLZA betreibt eine langfristige Planung seiner technischen Langzeiterhaltungsmaßnahmen.	B3.1 Repository has documented preservation strategies.
Develop Packaging Designs and Standards	
9.1 Das dLZA spezifiziert seine Übergabepakete (Submission Information Packages, SIPs).	B1.2 Repository clearly specifies the information that needs to be associated with digital material at the time of its deposit (i.e., SIP).
10.1 Das dLZA definiert seine Archivpakete (Archival Information Packages, AIPs).	B2.1 Repository has an identifiable, written definition for each AIP or class of information preserved by the repository. B2.2 Repository has a definition of each AIP (or class) that is adequate to fit long-term preservation needs.

172 Note that while generally DINI does not have any criteria in this functional entity, it could be argued that the requirement that the IT-system has to be subject to regular maintenance has similar implications as "Monitor Technology."

Common Services and Requirements

nestor	TRAC	DINI
4 Die Organisationsform ist für das dLZA angemessen.		
5 Das dLZA führt ein angemessenes Qualitätsmanagement durch. 5.1 Alle Prozesse und Verantwortlichkeiten sind definiert. 5.2 Das dLZA dokumentiert alle seine Elemente nach einem definierten Verfahren.	A3.6 Repository has a documented history of the changes to its operations, procedures, software, and hardware that, where appropriate, is linked to relevant preservation strategies and describes potential effects on preserving digital content. C1.8 Repository has a documented change management process that identifies changes to critical processes that potentially affect the repository's ability to comply with its mandatory responsibilities. B1.8 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (Ingest: content acquisition).[173]	
13 Die IT-Infrastruktur ist angemessen. 13.1 Die IT-Infrastruktur setzt die Forderungen aus dem Umgang mit Objekten um. 13.2 Die IT-Infrastruktur setzt die Sicherheitsanforderungen des IT-Sicherheitskonzepts um. 14 Die Infrastruktur gewährleistet den Schutz des digitalen Langzeitarchivs und seiner digitalen Objekte.	C1.1 Repository functions on well-supported operating systems and other core infrastructural software. *C1.2 Repository ensures that it has adequate hardware and software support for backup functionality sufficient for the repository's services and for the data held, e.g., metadata associated with access controls, repository main content.	2.5 Sicherheit, Authentizität und Integrität *2.5.1 Server *Mindeststandard*: - Das Betriebskonzept gewährleistet eine angemessene Verfügbarkeit des Systems. - Es existiert eine Dokumentation des technischen Systems mit 1. relevanten Versionsangaben und technischen Parametern zu allen Komponenten […] 3. Regelung der Wartung des Systems - Einsatz einer Technologie zur Sicherung und Wiederherstellung der Server-Software, der Metadaten und der Dokumente mit täglicher Sicherung. - Sichere Installation des Systems und der Software-Komponenten. - Regelmäßige Wartung des Systems. *Empfehlungen* - SSL-Zertifizierung mit vertrauenswürdigem Zertifikat für verschlüsselte Kommunikation wird eingesetzt. - Aufteilung der Dokumentation des technischen Systems in einen veröffentlichten und einen internen Teil. - Autonome Überwachungs- und Alarmfunktion bei Ausfall des Servers oder einzelner Komponenten. - Havarieszenarien sind vorhanden.

173 See also TRAC 2007 B4.5, which makes the same requirement for Archival Storage.