DEUTSCHE INITIATIVE
FÜR NETZWERKINFORMATION E.V.

DINI-Certificate
Document and Publication Services
2007

Working Group "Electronic Publishing"

DINI Schriften 3-en
[Version 2.0, September 2006]

DEUTSCHE INITIATIVE
FÜR NETZWERKINFORMATION E.V.

DINI-Certificate
Document and Publication Services
2007

Working Group "Electronic Publishing"

DINI Schriften 3-en

[Version 2.0, September 2006]

September 2006

Imprint
DINI – Deutsche Initiative für Netzwerkinformation e.V.
– Geschäftsstelle –
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen
Papendiek 14
37073 Göttingen
Tel.: 0551-39-3866
Fax: 0551-39-3856
E-Mail: gs@dini.de
www.dini.de

Contents

Summary

Installing and institutionalising document and publication repositories at universities creates the opportunity to archive scholarly publications that were produced at the respective universities and to offer them to a worldwide audience. This new service offered by libraries and computing centres as infrastructure facilities helps disseminate the idea of electronic publishing as a new tool for academic work. DINI, the Deutsche Initiative für Netzwerkinformation (German Initiative for Network Information), supports this to reach a higher level of scientific and scholarly communication both nationally and internationally. To reach this goal it is necessary to create a network of Document and Publication Services. A network of local publication repositories complements the dominant (commercial) publications of publishing houses, and can therefore have a regulating function against the monopolistic tendencies in scholarly publishing.

The German Science Council and the Conference of University Rectors in Germany call for the installation of Document and Publication Services, funding institutions such as the Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research) and the Deutsche Forschungsgemeinschaft (German Research Foundation) support (and fund) this. It is important to develop these services according to international standards and on the basis of proven technologies. It is only in this manner that the worldwide visibility and impact of the scholarly work of the individual scientist and scholar, as well as that of an entire university, can be improved.

This second edition of the 'DINI-Certificate Document and Publication Services' includes current international developments, describes how Document and Publication Services support the open-access publication of scholarly documents, and shows how as Institutional Repositories they can position themselves in their respective institutions, especially with regard to supporting scholars and scientists to make pre and post prints of their documents available that have already been published by publishing houses – free of time or space constraints.

The DINI-Certificate 2007 aims for the following:

- Provide a detailed description of the requirements of certified Document and Publication Services
- Position Document and Publication Services as an Institutional Repository
- Highlight development directions in the areas of developing services and of information exchange

- Offer a visible reference to both users and providers to certify the adherence to defined standards and recommendations by the respective Document and Publication Services

By issuing this certificate, DINI offers a quality control for Document and Publication Services.

A set of mandatory minimum requirements of a service and its provider must be attained, prior to the award of the DINI Certificate. Fulfilling these requirements is essential in order to allow modern scholarly communication. At the same time, DINI makes recommendations with regard to foreseeable developments that might become future requirements.

A working group within DINI verifies the criteria for the DINI Certificate against international standards and developments and updates them accordingly. For this reason, the certificate carries a date stamp (year of version).

In this second version of the DINI Certificate the name of the certificate has been changed from 'Document and Publication Repositories' to 'Document and Publication Services', to clarify that it not only certificates a server, but also server-based services.

1 Aims

With the World Wide Web (www), science has created a new tool for communication, which is increasingly used commercially as well. Despite intensive use of the web by academics in their everyday work, the opportunities the web offers to improve scholarly communication are by no means exploited to their full extent.

Deficits in this regard are most prominent in the dissemination of high-quality scholarly publications. While acceptance of electronic media has risen enormously, the willingness to actively use the technology as a tool for publication is largely limited to commercial publishers that publish electronically, especially in the STM areas. It is in these sciences, however, that the quasi monopoly of the publishers has led to exorbitant prices, more and more often exceeding the abilities of libraries and universities to pay the demanded charges. Building local publication repositories may work as a regulating factor: publishing the great majority of scholarly works on non-commercial repositories would create a communication network that will at least make disproportional profit maximisation more difficult.

Offering publications on local repositories allows for the improvement of publication services in scientific areas of less commercial interest. The publication of

doctoral theses on the internet is a good example of this process. Major prerequisites for success were:

- Development of nationwide standards for the collection of metadata
- Publication on local repositories through the cooperation of libraries and computing centres
- A workflow for the upload to the German National Library (Deutsche Nationalbibliothek), who in turn assumed responsibility for long-term archiving of the theses

The German Science Council and the Conference of University Rectors in Germany recommend the installation of document and publication repositories, the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) and the German Research Foundation (Deutsche Forschungsgemeinschaft) support the implementation of document and publication repositories. National science institutions, such as the Max-Planck-Society, are preparing for their extensive use. Internationally, organisations such as SPARC are strong supporters of this development. The MIT and other universities in the USA and the United Kingdom are offering open-source software to run institutional repositories.[1] In the Netherlands, the DARE project has facilitated widespread introduction of services. Local publication systems are common in Germany. In order to make them even more successful it is necessary to standardise further developments. DINI, with its publication 'Electronic Publishing in Higher Education', laid a first foundation for the introduction of general regulations and standards. The 'DINI-Certificate Document and Publication Services' takes this a step further by clearly describing criteria that will guarantee for services to be in line with national standards and international developments. Repositories fulfilling these criteria will be awarded the DINI Certificate, testifying to their quality. The certified services will be links in a chain of content providers offering their metadata to specialised service providers in a standardised format and procedure in accordance with the OAI model. The service providers in turn may aggregate these metadata and, among other possible functions, serve as connectors to commercial service providers.

If the worldwide movement to implement and operate local (institutional) repositories can be based on the solid foundations of universities' infrastructure insti-

[1] See the *OSI Guide to Institutional Repository Software* v 3.0: http://www.soros.org/openaccess/software/

tutions such as libraries and computing centres in their respective international networks, a new base may be created for a communication infrastructure in the sciences. Without much extra effort, essential investments in information infrastructure could be used to benefit the sciences and the financial situations of the universities much more efficiently.
Certification of services to state their international compatibility within a growing communication network of the sciences for high-quality publications does not have to be restricted to universities. It is desirable that commercial publishers integrate their services into a standardised scholarly communication.

2 Criteria

The following criteria are split into two sections. In the first section minimum standards and requirements are specified, which must be met by the Document and Publication Services and their providers to be eligible for the certificate. As DINI plans continuous updates of the certificate, the second section lists recommendations that, from today's point of view, are possible future minimum standards.

2.1 Visibility of the Services

Minimum standards

- The entire range of services must be available on the internet
- The Document and Publication Services' homepage must be referenced from the institution's homepage

Recommendations

- Registration of the Document and Publication Services at the Directory of Open Access Repositories (OpenDOAR)[2] or the Registry of Open Access Repositories (ROAR)[3]
- Registration of the Document and Publication Services at the Open Archives Initiative[4] as data-provider

[2] http://www.opendoar.org

[3] http://archives.eprints.org/

[4] Directly or via a data provider that aggregates or accumulates data (http://www.openarchives.org)

2.2 Policy

Minimum standards

The provider of the Document and Publication Services publishes a policy regarding content and operation of the repository. This policy must state terms and conditions (rights and obligations) of the provider and of authors and/or editors. The policy must contain:

- Standards for the publications in the repository with regard to content and functional and technical qualities
- A guarantee to archive published documents for a defined time depending on content and functional and technical qualities of the publications
- Procedures for the operation of the document repository; definition of the services offered to the authors and/or editors by the provider of the Document and Publication Services
- Specification of terms and conditions (rights and obligations) of authors and/or editors
- Statement on open access

Recommendations

If the Document and Publication Services is to be part of an institutional open access strategy, a number of elements must be added to the policy. Besides a definition of the institution's understanding of 'open access', the policy must contain statements in three areas:

- Action guideline for authors: the institution may intend to self-archive documents that are published elsewhere (open access 'green') and/or to publish open access (open access 'gold'). It may be mandatory or recommended for the authors to follow this guidelines. Disregard of or adherence to the mandate or the recommendations may be sanctioned or rewarded.
- The action guideline may apply to all forms of publications or to selected publication types (eg mandate or recommendations only for journal articles).
- Depending on the intended publication form and action guideline, the Document and Publication Services must provide different tools ranging from a simple repository with an upload tool to versioning, to document authentication mechanisms, and to automated licensing of genuine electronic publications

2.3 Author Support

Minimum standards

- Offer consultancy services and support via web pages and/or email and/or telephone, as well as person-to-person
- Support of the entire publication process (including technical and legal advice)[5]
- A link to the SHERPA/RoMEO-List[6]

If the Document and Publication Services is part of an institutional open access strategy:

- Provide authors who intend to upload their preprints and/or post-prints with online forms (document upload tools)

Recommendations

- Offer courses on electronic publishing at least once per semester
- Offer courses on structured writing
- Use of helpdesk system
- Provide style sheets or templates
- Offer 'how-to manuals' for download (eg on PDF creation)
- Provide advice on usage and citation of documents
- Provide advice on copyrights
- Provide interfaces and help in English

2.4 Legal Aspects

When hosting publications on a server, with regard to copyrights one must differentiate between a primary publication and the hosting of an author's copy of a document already published elsewhere. In the case of a primary publication, the author in general still holds all copyrights; in the case of a so-called author's copy, the author will usually have transferred certain rights to a third party before publishing on an institutional repository. However, many publishing houses' policies allow hosting the publications on the author's homepage or on an institutional repository (SHERPA/RoMEO), even if contracts exist.

---

[5] These do not necessarily have to be provided for free, especially detailed technical support (PDF creation, conversions, etc). The fees should be transparent to the author/customer.

[6] http://www.sherpa.ac.uk/romeo.php

As a rule, it is easiest if the author still holds all rights at the time of the upload of a document.

Minimum standards

Primary publications

No copyrights or exclusive rights were transferred to a third party (true primary publications on the document and publication repository):

- An agreement (author contract) has to be signed by the author allowing the end-users free electronic distribution of the document and stating the conditions for this (rights to store electronically, especially in databases, and make available to the public for individual retrieval to view on a monitor and print (online use) in part and/or as a whole)

Should an additional publication in eg a journal be intended, the author should make certain that the publishing house supports the primary publication on an institutional repository. The author's copy could be an alternative.

Author's copy

The work is published elsewhere or is in the process of being published elsewhere. Copyrights were or will be transferred to a third party.[7] In this case, different scenarios exist:

- The author published the document under a licence (Creative Commons, DPPL Digital Peer Publishing Licence) that permits hosting the publication on the institutional repository and does not contradict the transfer of copyrights (eg as listed in an author's contract)
- The author's contract lists rights to publish a digital author's copy
- An addendum to the author's contract stipulates terms and conditions for the publication of digital author's copies
- Publication of digital author's copies is covered in the respective publishing houses' policies

For every author's copy the applied scenario of the abovementioned must be stated.

---

[7] If no different agreement exists between author and publishing house, exploitation rights for articles in collections and periodicals will return to the author, permitting the archiving of an author's copy.

Recommendations

- Primary publications: it is recommended that the author publishes under a licence that permits the end-user free digital dissemination of the publication[8]
- Primary publications: implement the offer to the author to license the uploaded documents in the upload process (as realised in the OPUS 3.0[9] licence module)
- In addition to the end-user licence the author may transfer additional rights to third parties (eg publishing houses) in an author contract. This must not contradict the conditions of the licence
- List rights and terms and conditions in the metadata
- Exemption from liability must be formulated in a disclaimer and be part of the contract between author/editor and service provider

Due to the protection of the original publication by copyright laws, publishing according to the so-called 'open access green' paradigm, ie archiving documents published elsewhere, carries consequences for the author as well as for the Document and Publication Services provider. To support the author, the service provider should at least point out the SHERPA/RoMEO list and mention the legal aspects. The service provider should explain to the author that the service provider is not entitled to offer binding legal advice. Cooperation with the law faculty or the legal department of the institution, the DINI Working Group on Copyright[10] or the Aktionsbündnis 'Urheberrecht für Bildung und Wissenschaft' (Coalition for Action 'Copyright for Education and Research')[11] can be useful.

---

[8] Different forms of content (eg text, software, simulations, primary data or multimedia data) require different licensing models. It is for this reason that no specific licence is recommended here. The Creative Commons License (CCL) is widely used (http://creativecommons.org/) (Science Commons will analyse the legal framework for scientific applications in particular), another licence is the Digital Peer Publishing Licence (DPPL) of the Digital Peer Publishing project in North Rhine-Westphalia, Germany (www.dipp.nrw/lizenzen). Where software is concerned, the common licences are listed under www.opensource.org/licenses.

[9] http://elib.uni-stuttgart.de/opus/doku/opus_sw.php/

[10] http://www.dini.de/dini/arbeitsgruppe/arbeitsgruppe_details.php?ID=31

[11] http://www.urheberrechtsbuendnis.de

2.5 Security, Authenticity and Data Integrity

2.5.1 Server

Minimum standards

- The operational concept guarantees adequate availability of the system
- A documentation of the technical system, including:
  1. Relevant information on the version and on the technical parameters of all components
  2. Regulations on access to the server with regard to
     - location
     - system
     - personnel (responsible person(s))
  3. Regulations on system maintenance
- Back-up system to secure repository software, metadata and documents on a daily basis
- Secure installation of the system and the software components
- Regular system maintenance
- Technically controlled and verifiable acceptance of documents

Recommendations

- SSL certification with trusted certificate for encoded communication
- Separation of the technical system documentation into a public and an internal section
- Autonomous controlling and alerting functions in case of a malfunction of the server or one of its components
- Damage-control procedures in place

2.5.2 Documents

Minimum standards

- Use of persistent identifiers (systems applying a resolving service, eg urn:nbn or DOI)
- A document with altered content must be treated as a new document (new persistent identifier)
- Archiving of uploaded original files in their original formats (other formats see 2.8)
- Minimum standards must be documented in the policy or the guidelines of the Document and Publication Services

Recommendations

- Application of a procedure to verify the integrity of documents (eg a hash number) as well as the publication of the procedure and hash value
- Application of advanced digital signature (§ 2 Abs. 2 SigG 2001[12])
- Rendering of archival file-formats for the export of documents into long-term archiving facilities or institutions[13] (see 2.8 and 4.7)
- Presentation of the documents in an established presentation file-format. Should this not be possible, applicable software should be offered for download, or an internet reference be given

2.6 Indexing

2.6.1 Subject indexing

Minimum standards

- An indexing policy must exist and be known to the authors
- Verbal indexing with free keywords or classification terms
- Application of the Dewey Decimal Classification (DDC) in accord with the usage in the German National Bibliography as the general classificatory indexing system for documents (see DINI OAI-Recommendations)[14]

Recommendations

- Use of at least one additional standardised system of verbal or classificatory indexing (general or subject specific, eg German Schlagwortnormdatei, LoC Subject Headings, CCS, MSC, PACS etc)
- Keywords in English
- Short summaries/abstracts in German and English

2.6.2 Export of metadata

Minimum standards

- Metadata are available for free
- Metadata are structured as Unqualified Dublin Core (ISO 15836:2003)

---

12 http://www.gesetze-im-internet.de/bundesrecht/sigg_2001/gesamt.pdf

13 http://www.langzeitarchivierung.de/

14 http://edoc.hu-berlin.de/series/dini-schriften/2005-2-de2/PDF/2-de2.pdf/ (URN: urn:nbn:de:kobv:11-10049220)

Recommendations

- Metadata are structured as Dublin Core Qualified[15]
- Metadata are structured as ONIX[16]
- Offer technical and/or archival metadata including (where applicable) print-on-demand data (eg PREMIS,[17] LMER[18])
- Support import and export of metadata in bibliographical databases (eg reference-management software, consortia, OPAC etc)
- Offer link lists for search engines' indexing robots (eg Google, Fast, Scirus, etc)

2.6.3 Interfaces

Minimum standards

- Web interface for users
- Support OAI-PMH 2.0 in accordance with DINI OAI-Recommendations[19]

Recommendations

- Support use of the OAI-Protocol to exchange complex metadata schemas
- Web-service interface (eg SOAP)[20] is available
- Z39.50[21] or SRU[22] or both interfaces are available

---

15 Currently no binding definition or standard exists for DC Qualified. Cf. http://dublincore.org/documents/abstract-model/

16 http://www.editeur.org/onix.html

17 PREMIS: PREservation Metadata Implementation Strategies (http://www.oclc.org/research/projects/pmwg/)

18 LMER: Langzeitarchivierungsmetadaten für elektronische Ressourcen (Long-term preservation Metadata for Electronic Resources), http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:1111-2005051906 (urn:nbn:de:1111-2005051906)

19 http://edoc.hu-berlin.de/series/dini-schriften/2005-2-de2/PDF/2-de2.pdf/ (urn:nbn:de:kobv:11-10049220)

20 http://www.w3.org/TR/soap

21 http://www.niso.org/standards/resources/Z3950_Resources.html

22 http://www.loc.gov/standards/sru

2.7 Logs and Statistics

Minimum standards

- Every document repository must keep consistent access statistics (web server log files) in accordance with Data Protection Acts.
- Web server log files must be made anonymous for long-term archiving purposes[23]
- A declaration or documentation must be appended to the statistics describing the criteria used to collect or enhance the data contained therein. It must be pointed out that access numbers can generally not be used to compare different document repositories, but solely to compare access to different documents on one repository

Recommendations

- Access to documents by automated agents, robots or similar is filtered out[24]
- Web server log files are edited according to the Counter Code of Practice[25]
- A document's access statistics are attached to the document as dynamic metadata and visible to the end-user[26]

2.8 Long-Term Availability

Minimum standards

- Persistent linking of metadata and documents (eg via a persistent identifier or through storage of metadata and document in one single container)
- Definition of a minimum availability of a document of no less than five years is contained in the policy
- Where archival copies are created in addition to the original files uploaded by the author, these must be free of Digital Rights Management (DRM) constraints that prevent application of long-term availability strategies (migration, emulation)

---

[23] In compliance with the applicable paragraphs of the German laws on the protection of data privacy: § 4 Abs. 4 Nr. 2 TDDSG, § 18 Abs. 4 Nr. 2 MDStV. For an example see: http://www.zendas.de/technik/sicherheit/apache/index.html

[24] Eg http://www.robotstxt.org/wc/active/all.txt

[25] See http://www.projectcounter.org/cop_books_ref.html for orientation, and appendix D http://www.projectcounter.org/cop_books_appendix_d.html

[26] See eg at PsyDok (http://psydok.sulb.uni-saarland.de/) and at MONARCH (http://archiv.tu-chemnitz.de/) where it is linked to from the respective index pages (document splash pages).

Recommendations

- Secure long-term availability, where necessary through cooperation with an archiving institution
- Use of open file formats suited for long-term archiving (eg PDF/A, ODF, TXT, HTML, TEX) and free of protective DRM measures
- Creation of technical metadata for long-term archiving (eg using JHOVE)[27]
- Unambiguous identification of the file format in the metadata with a link to publicly available file format registries
- Policy on the deletion of documents
- Support import and export of object containers that – besides the actual document – contain bibliographical and, for long-term archiving purposes, technical metadata (eg in the Universal Object Format)[28]

3 Awarding and Evaluation

The DINI office or an authorised working group is responsible for the awarding of the 'DINI Certificate Document and Publication Services'. The certificate's seal shows the year of its version. The certificate acknowledges that the certified repository meets the minimum standards of DINI-certified Document and Publication Services.

A fee is charged for the issuing of the DINI Certificate:

- Non-profit organisations
  DINI members 50.00 €, others 100.00 €
- Profit organisations
  DINI members 150.00 €, others 250.00 €

The provider of the Document and Publication Services applies at DINI for certification by completing a form on the DINI website. This form lists the criteria laid down in Section 2. The provider states if and to what extent the Document and Publication Services fulfils the minimum standards of the DINI Certificate. Further explanations/clarifications can be added in designated fields in the form, as well as URLs or other options on how to receive additional information.

An authorised DINI Working Group will verify the submitted data. Access to the services to be certified must be permitted to this group. The provider of the

---

[27] http://hul.harvard.edu/jhove/

[28] http://kopal.langzeitarchivierung.de/index_objektspezifikation.php.en/

Document and Publication Services must be prepared to answer questions from the DINI Working Group. An on-site visit will not normally be required. Additional costs that may emerge during the certification process must be covered by the provider of the Document and Publication Services. The provider of the Document and Publication Services will be informed about possible additional costs beforehand.

The certification process should generally be completed within two months. There is no time limit on the DINI Certificate. As the certificate shows the year of the version, it will always be clear to what standards the Document and Publication Services applies.

The provider of the certified Document and Publication Services is entitled to call the services 'DINI-certified Document and Publication Services', and to show the DINI Certificate's seal on a web page or in other applicable forms. Any misuse of the seal or certificate will be prosecuted in accordance with applicable laws.

The new DINI Certificate 2007 takes current international developments into account. It supports positioning German Document and Publication Services as Institutional Repositories in the sense of trustworthy, future-oriented services that closely observe open-access developments. For this reason a number of new minimum standards as well as recommendations for further development of the Document and Publication Services have been added.

All DINI Certificates awarded to this date retain their validity according to the year of the certificate's version.

Document and Publication Services already certified have the option of updating their DINI Certificate to the 2007 version within one year after its publication in a fast-track procedure. Only the new minimum standards will be checked. The fee charged for the certification will be 50% of the regular fee.

DINI reserves the right to revoke the DINI Certificate, should a breach of the minimum standards of the DINI Certificate occur after it has been awarded.

4 Explanations and Examples

Explanations and examples of how the DINI Certificate's criteria can be fulfilled are listed below.

A local document repository (institutional repository) is to ensure storage, management, provision, archiving, as well as search and retrieval of electronic resources of an institution.

4.1 Visibility of the Entire Service

Easy accessibility and visibility of the Document and Publication Service within the institution for which it is provided is an important objective. The service also has to be integrated in other subject-oriented or institutional services. Registration organisations, such as openarchives.org or OpenDOAR, are general mediators between document and publication services and other services building on these. Of great importance for the technical integration are the metadata and interfaces mentioned in Section 2.6.

4.2 Policy

4.2.1 Example 1: Policy of the Document and Publication Services of Humboldt University Berlin

Document and Publication Server of Humboldt University Berlin – Policy

1 Objectives and Criteria for Content of the Document and Publication Server of Humboldt University

The Document and Publication Server offers the organisational and technical framework to all members of Humboldt University for publishing scientific documents digitally. As a result , Computer and Media Services and the University Library are able to publish highly important scientific documents on the internet under strict quality control.

The digital documents are provided with persistent identifiers and addresses, and are indexed within national and international library catalogues, search engines and other reference tools. The Document and Publication Server provides protection against distortion by using digital signatures and digital time stamps. Furthermore, long-term preservation of digital publications is guaranteed.

Operation and development of the Document and Publication Server are integrated into national and international initiatives and projects such as the 'Networked Digital Library of Theses and Dissertations' (NDLTD) or the 'Open Archives Initiative' (OAI).

2 The Collection Mandate of Humboldt University Library for Digital Documents

The collection mandate of Humboldt University Library consists of collecting, cataloguing and archiving all the scientific documents published by the mem-

bers of Humboldt University. It refers to digitally born documents as well as digital versions of printed documents.
Also included are significant historic documents from the University Library and other institutions, which are digitised due to terms of content, conservational aspects, or the requirements of place-independent use.
Intellectual property rights will be preserved. Publishing the document with the Document and Publication Server does not prevent publication elsewhere, such as in scientific journals or other document servers. According to the recommendations of the German Science Council (Wissenschaftsrat), all scientists at Humboldt University are asked to secure further rights of use when negotiating with publishers. At a later date, possibly after a certain qualifying period, they will be asked to publish their documents with the Humboldt University Document Server as well.
Observation of copyrights and rights of use for third parties are solely with the authors or the editors of the digital documents.

3 Digital Documents
The term 'digital document' as used here is defined as a document based on text and images, that is stored in digital form on a data medium, and that is distributed via computer networks. In the future this term will be extended toward multimedia documents that include audio and video sequences.
Any document to be published with the Document and Publication Server should meet the following requirements:

1. *It is to be distributed for public access.*
2. *It is not a dynamic document. If any changes are made in the document, a new version will need to be stored.*
3. *It fulfils the technical parameters as defined by the Computer and Media Services and the University Library of Humboldt University.*

4. Digital documents to be published and distributed by the Document and Publication Server:

1. *The following categories of digital documents will be stored and distributed:*
   a. *Single publications and publication series of Humboldt University with scientific content, as the public lectures of Humboldt University or the publication series of the University Library*

b. *Single publications and publication series published by staff of Humboldt University, ie collections, conference papers, research reports, journals (e-journals)*
c. *Documents that are mandatory to be published according to examination rules and regulations (theses and dissertations)*
d. *Single publications and publication series of institutions and persons associated with Humboldt University*
e. *Documents of Humboldt University students, eg masters theses and seminar papers, should be recommended by a lecturer*

2. *Distribution of these documents can be restricted to local or temporary use only.*

5 *Technical characteristics of the Document and Publication Server:*

1. *By providing and auditing qualified digital signatures, the digital documents receive a legally valid certificate of authenticity. The assignment of these qualified digital signatures occurs in full accordance with the German Digital Signatures Law (Signaturgesetz – SigG).*
2. *The digital documents are provided with individual and persistent addresses. This enables direct access to the document.*
3. *Retrieving the digital documents is possible via library catalogues, queries of the bibliographical metadata, search terms within the structures of the digital documents, via alphabetical and indexing systems, and via dynamically produced lists and indexes.*
4. *For indexing, storage and archiving of the digital documents, international standards such as the guidelines of the Open Archives Initiative (OAI) are employed and developed.*
5. *When using SGML/XML, an archiving time span of 50 years is guaranteed. Preserving other formats depends on the availability of these formats, their viewer software and their conversion tools.*

6 *Organisational Regulations:*

1. *The Document and Publication Server of Humboldt University is a joint operation between the Computer and Media Services and the University Library.*
2. *The electronic publication is free of charge for Humboldt University staff and members of associated institutions.*

3. *Submitting the digital documents for distribution via the Humboldt University document server takes place in the University Library.*
4. *Additional services that are necessary for publication, such as processing digital documents or conversion into other formats, are conducted by staff of the Computer and Media Services or the University Library Services. Any significant additional work will be charged according to the scale of charges of the Computer and Media Services and the University Library.*
5. *Contact for all questions about the Document and Publication Server is the joint 'Electronic Publishing Group' of the Computer and Media Services and the University Library.*

4.2.2 Example 2: Open-Access Policy of the Georg-August-University Göttingen

The Georg-August-University-Göttingen Open Access Policy

The internet has become an essential piece of the global scientific enterprise as a medium for communication, information and publication. At the same time, the current scientific communication and publication system finds itself in a crisis situation due in part to the role played by commercial publishers. Symptoms of this phenomenon are excessive price increases for scientific journals that overwhelm the library budgets of universities. This results in subscriptions being cancelled and immediate access to relevant information being threatened. To meet this challenge on a global scale activities are underway to provide Open Access to a larger share of the scientific literature. The Chair of the University of Göttingen strongly encourages the scientists of the university to deposit their peer-reviewed and published articles on the university's publication server at the Göttingen University Library whenever this is legally possible. A growing number of publishers allow authors to self-archive already published articles on their institutional repositories.

The online database SHERPA/RoMEO lists which scientific publishers allow self archiving under which conditions. This refers to large publishing houses such as Elsevier, Springer and Wolters-Kluwer. According to Article 38a of the German Intellectual Property Rights Law authors regain the full publishing rights for their articles from collective works or journals one year after publication, unless different contractual agreements have been made. Publications dated

before 1995 may be self-archived without any problems, as online publishing rights were not transferred and remain with the author. If you have questions about how to self-archive and whether you are allowed to do so please contact […][29].

Self-archiving helps authors as well as publishers: research shows that articles with an additional online version are cited more often, which enhances visibility and impact of your scientific works' results. Additionally, higher citation rates improve a journal's impact factor – a fact highly appreciated by commercial publishers as well. At the same time, the university's scientific output receives more recognition by being displayed on an internationally networked institutional repository.

In many disciplines it is common practice to deposit preliminary versions of articles on so-called preprint servers before being submitted to journals or other collective works. One of the most successful examples is the arXiv server, an indispensable platform for many scientists searching for up-to-date research papers. The arXiv server started off as a preprint server for high-energy physics and is today widely used within mathematics, biology and computational sciences. Many scientific publishers accept this way of displaying research on public international servers accessible at no cost to participants. At times, publishers even appreciate the possibility to observe this preliminary publishing activity and the reactions of other scientists and / or scholars.

Scientists and libraries around the world have begun depositing preprints on local institutional repositories. Global databases such as OAIster index documents that have been archived in OAI-interface-equipped (Open Archives Initiative) repositories similar to the Göttingen University server. Platforms like Google Scholar or commercial services like Web of Science or SCOPUS now index preprints that have been deposited on trusted repositories such as the Göttingen University server.

The Institute for Scientific Indexing (ISI) conducts surveys on citation rates within scientific journals. Within the natural sciences the traditional flag-ship publications such as Nature, Cell or Science are still at the top. However, Open Access Journals, such as Public Library of Science Biology (PLoS) or certain journals from BioMed Central (BMC), attain citation-rate values of up to 13.9 and rank among the top journals of their respective disciplines. The business models

[29] Current contact information left out as it is subject to change.

for PLoS or BMC rests on author's fees instead of subscription fees charged to libraries or individual subscribers. For certain journals the Göttingen University Library bears such author's fees in part or whole due to special consortia conditions. Furthermore, in 2002, the large research-funding organisations signed the 'Berlin Declaration', manifesting their support of the Open Access principle. They have also indicated that publication fees can be requested in proposals for research funding.

If you have plans to initiate any new publishing activities you are welcome to use the university's infrastructure, including the University Press Göttingen and the University server, both operated and maintained by the Göttingen University Library. Scientists often serve as editors, referees or as members of editorial and advisory boards for journals within your discipline. Please pay attention to the pricing policy and the access conditions for your readership. If necessary, exert influence towards a better Open Access policy of your publisher.

Links to mentioned institutions

- *ArXiv: http://www.arxiv.org*
- *Berlin Declaration on Open Access: http://oa.mpg.de/openaccess-berlin/berlindeclaration.html*
- *BioMed Central: http://www.biomedcentral.com/*
- *DOAJ (Directory of Open Access Journals): http://www.doaj.org*
- *Electronic Journal Library: http://www.bibliothek.uni-regensburg.de*
- *Google Scholar: http://scholar.google.com/*
- *Institute for Scientific Indexing: http://www.isinet.com/cit*
- *OAI (Open Archives Initiative): http://www.openarchives.org*
- *OAIster (OAI-Harvester): http://www.oaister.org*
- *Open Access Server of Göttingen University: http://webdoc.sub.gwdg.de/*
- *Public Library of Science: http://www.plos.org*
- *SCOPUS: http://www.scopus.com*
- *SHERPA/RoMEO database: http://www.sherpa.ac.uk/romeo.php*
- *University Press Göttingen: http://univerlag.uni-goettingen.de/*
- *Web of Science: http://www.isinet.com/products/citation/wos*

4.3 Author Support

An important goal of electronic publishing at universities is to archive all publications that are created at the respective universities to make their scientific and scholarly outputs more available. For the purposes of long-term archiving and the ability to further edit digital publications, standardised file formats, signatures, metadata etc are advisable.

Working towards this goal begins with advising and supporting authors and editors. The use of templates in word-processing applications is one basis for structural writing in the digital age.

It is therefore recommended to initiate dedicated course programmes for electronic publishing at the universities. To make electronic publishing more attractive for authors, strategic assistance, such as support services, should be developed and offered to authors and editors.

All this should lead to a university-wide electronic-publishing infrastructure. This is a prerequisite to ensure further development of support services and technologies, and to shape and implement technical requirements for the authors to allow long-term archiving and availability of the documents. Examples are the use of DTDs for dissertations according to international standards, the support of WinWord templates for dissertations, or the definition of guidelines for LaTeX, as is current practice at the Humboldt University, Berlin.

4.4 Legal Aspects

Regardless of the type of publication (primary publication or author's copy), the guidelines/policy, as required in the 'DINI Certificate Document and Publication Services', must define rights and obligations of both authors/editors and the service provider, from which a legally binding agreement between the two parties is agreed.

It must be stipulated that authors should grant the following non-exclusive rights to the Document and Publication Services provider:

- The right to store the publication electronically, especially in databases, to make the publication available to the public for individual retrieval, to view on a monitor and/or print (online use) in part and/or as a whole
- The right to notify an archiving institution and transfer the data to that institution
- The right to convert the files for the purpose of long-term archiving while retaining the content's integrity

These rights also apply to the metadata as provided by the author (eg abstracts).

Granting these rights not only allows publication but also secures the long-term availability of the electronic documents on the document and publication repositories.

Providers of Document and Publication Services must explicitly grant the following rights to authors/editors:

- The author/editor may make the document available on personal repositories or other institutions' repositories in part and/or as a whole (right to self-archive). This right is granted for the unaltered publication. Altered versions must be flagged as such
- The author/editor has the right to license the published work following a generally used licence model to stipulate the relation between user and author/editor

It has to be taken into account that in the cases of documents already published elsewhere, certain rights may have been granted to others. The SHERPA/RoMEO list offers instructions on how to proceed.

Third-party rights

- Authors/editors must assure the service provider that no (copy)rights of third parties will be breached by publishing a document or parts thereof (eg pictures)
- Authors/editors must ensure immediate notification of the service provider in cases of doubt or in cases of alleged or actual legal hindrances to the publication of a document or parts thereof
- Questions of liability in cases of damages and/or legal prosecution must be clarified and stipulated between authors/editors and the operator of a repository in a written contract

Print publications

The following items become relevant should universities or university presses offer the extra service of printing a publication.

If publishing and distributing the work in print, universities/university presses should examine if they require the authors/editors to grant them the exclusive right to do so, and should this be the case, request authors/editors to grant them this right. Authors/editors must agree not to publish, transmit or reproduce the work in tangible or intangible form over a given period of time (abstention duty, avoidance of duplicate registration with VG Wort (Germany)). The option of printing an edition in a print-on-demand procedure should be stipulated in the contract.

In the case of a full-fledged print publication (ISBN, distribution via wholesalers and bookstores), a legally binding contract must be signed by authors/editors and universities/university presses. The contract must contain paragraphs regulating (copy)rights and exploitation rights.

4.5 Security, Authenticity and Integrity

4.5.1 Server

In the technical system's documentation, administrator(s) must be named and the system itself must be described. The documentation has to be kept current at all times. Date and author of the last update have to be stated.
Manufacturers and exact names must be given for all components. Hardware components must be described with the relevant parameters (speed, memory, etc). Software components must be listed with their respective version numbers.

Access to server
It has to be listed:
- Which group of persons is granted physical access to the server
- Who may log on to the system (differentiating between users with administrator privileges and users with regular privileges, and if applicable additional sets of privileges)
- Who is responsible for the system and how stand-ins for the persons responsible are organised
- Where the administrator password is stored

Organisation of system maintenance
Maintenance routines for the system, as well as reaction procedures in cases of malfunction have to be organised and listed.

Back-up and recovery
A back-up has to be stored every day, allowing recovery of the entire system without loss of data or functionality. Usage of a central automatic back-up service is recommended.

Reliable installation of the system and the software components
To offer reliable Document and Publication Services a system must be composed only of tested and stable components. There must be no known conflicts in the cooperation of these components. Sub-components of systems that are not necessary for the operation and provision of Document and Publication Services are to be deactivated or deinstalled. Products with updating services are preferable.

Regular system maintenance
An administrator has to be responsible for system maintenance as part of his/her regular duties. Maintenance especially includes security updates, but also software updates and reacting to malfunctions.

Technically controlled and traceable upload of documents
The controlled upload of documents ensures that only those documents that comply with the guidelines of the Document and Publication Services will be uploaded on the server. The service provider has to implement the necessary workflow. The workflow has to permit the verification of the document's upload.

SSL certification
An SSL certification is the prerequisite to guarantee authenticity of the documents for the reader. Certification has to be awarded by a certification authority (CA). Digital signatures warrant a document's integrity.

Segmentation of the documentation
The documentation should be segmented into a publicly available and an internal part (eg containing sensitive data).

Autonomous Monitoring and Alerting Function
Such a system continuously monitors the server's operation as a whole, as well as individual services (web pages, database functions, etc). Should the server or components of the server malfunction, an alerting service will send out an email or an SMS.[30]

'Worst-case' scenarios
For every degree of severity of a server's malfunction or dysfunction – from the failure of individual software or hardware components to the loss of an entire server – an action plan should be prepared to ensure operations can be resumed in the shortest possible time. Possible system failures include technical problems or damage caused by faulty handling or attacks from outside (hacking).

[30] Eg Nagios (www.nagios.org).

4.5.2 Documents

Persistent identifier

A persistent identifier must be assigned to every document. One possible option is URN assignment.

A document-management system (with data replication, data back-up) can be used to organise administration, storage and archiving of digital resources during their entire life cycles. Long-term access to these resources must be ensured through the use of unique, independent-of-location, persistent identifiers. Additionally, persistent identifiers can support harvest (ingest) requests. Persistent identifiers can assist in workflow control, where a clear and unequivocal identification of a resource is required, eg in decentralised workflows, checks for duplicates, authentication mechanisms, alerting systems, or multimedia or modular resources. Persistent identifiers can facilitate re-merging different resource elements that are stored separately in different repositories. In modular resources, identification of individual elements permits a selective merging of modules of a resource in a version without the need to store this version as a whole. Persistent identifiers allow reliable citing of digital resources in digital and non-digital environments.

To use a persistent identifier, its own identifier's schema and its name space have to be known. Within schema and name space, the persistent identifier has to be unique. Both should be public, documented, and listed in a persistent identifier registry (eg the IANA Registry for URN name spaces). For a persistent identifier a resolving service must exist.

The persistent identifier's schema should conform to RFC 1737[31] requirements.

The persistent identifier has to be listed in a resource's metadata, including its schema and the unique value for the respective object. References to related resources should conform to the same schema.

Persistent identifiers are, for example: in libraries the National Bibliography Number (NBN), a sub name-space of URNs (Uniform Resource Names); in publishing houses and increasingly with raw data the Digital Object Identifier (DOI) is used; other examples are Persistent URL (PURL), Handle, Archival Resource Key (ARK), etc.

---

[31] http://rfc.net/rfc1737.html

Examples

1. urn:nbn example:
   urn:nbn:de:bsz:93-opus-15563
2. DOI example without metadata syntax:
   DOI: 10.1045/april2004-dobratz
3. Handle example (D-Space System):
   URI: http://hdl.**handle.**net/1721.1/30592

For additional information on usable systems see:

- NBNs: homepage of EPICUR project: www.persistent-identifier.de
- General requirements on URNs: ftp://ftp.rfc-editor.org/in-notes/rfc1737.txt
- Registration of sub name-spaces: IANA Registry: www.iana.org/assignments/urn-namespaces
- Digital Object Identifier homepage: www.doi.org
- The Handle System homepage: www.handle.net
- URI: http://info-uri.info/
- PURL: http://purl.oclc.org/

Hash method

To secure a document's integrity for the reader, a hash value has to exist for every document. As the requirements on security evolve with ongoing developments in cryptography, the algorithms have to be adapted accordingly. Currently, use of the MD5 Message-Digest algorithm or of the Secure-Hash-Algorithm (SHA-1) can be recommended. The following examples are taken from the MONARCH Document-Service info page at Chemnitz University.[32]

Publications archived in MONARCH are secured with various hash methods and digital signatures. Basic principles of the security system are:

- A hash file exists for every publication in MONARCH
- For every file of a publication, three different hash values are calculated and stored in the publication's hash file
- A digital signature is added to the hash file
- Using the hash file, it can be determined at all times if a publication has been altered. The signature of the hash file itself further guarantees that the hash values have not been changed since their creation

[32] http://archiv.tu-chemnitz.de/integrity.html

Applied hash methods are:
- Secure Hash Algorithm (SHA)
- MD5-Message-Digest algorithm
- RIPE-MD algorithm

Hash files signature:
- The archivists adds a digital signature to the hash file
- For the digital signature the PGP (Pretty Good Privacy)[33] version 2.6.3 is used

Any change to the content of a document leads to a new document. This new document will be assigned its own unique persistent identifier. Versioning allows reference to earlier versions of the document. This reference should be a metadata element to create a link between versions/generations of a document.

Digital signature

The German Digital Signature Law differentiates the following signature levels:

1. 'Digital signatures' are data in electronic form, that are appended or logically linked to other digital data and serve authentication purposes.
2. 'Advanced digital signatures' are digital signatures (see 1.) that:
   - are assigned exclusively to the signature-key holder
   - allow identification of the signature-key holder
   - are created through processes only the signature-key holder controls, and
   - are linked to the data to which they are referring, to allow identification of any alteration of those data.
3. 'Qualified digital signatures' are advanced digital signatures (see 2.) that:
   - are based on a certificate that is valid at the time of their creation, and
   - are created using a secure signature-creation unit (hardware and software).

We recommend the use of at least the 'advanced digital signature'. To facilitate long-term archiving it is advised that the institution accepting the original files of a publication creates files suitable for archival purposes.

The Document and Publication Services provider should sign an agreement with the operator of an archive repository. For the time being, no specific archival file format is recommended. Documents on the document and publication repository should be published in an established presentation file-format. Currently this is the Portable Document Format (PDF). Should another format be used, either

---

[33] http://www.pgpi.org

the necessary viewing software should be offered for download, or an internet address be given, where it is available.

4.6 Indexing

4.6.1 Subject indexing

A written strategy (policy) on the subject indexing of the documents has to be available. Indexing with uncontrolled vocabulary only will not ensure good precision and recall results in the long run.

To enable a unified search across a number of document repositories, international indexing systems should be applied (see DINI-OAI recommendations).

4.6.2 Metadata export and interfaces

We are explicitly not making recommendations with regard to the repository's internal file structure. However, standardised and interchangeable metadata should be supplied.

Standardised administrative and technical metadata should be assigned to the electronic resources to facilitate archiving and secure long-term availability. Recommended standards such as PREMIS or LMER (see 2.6.2) aim at precisely this. Developments in national and international efforts to improve long-term availability of digital documents should be considered. At this point in time, interfaces to long-term archiving systems cannot be listed here because they are still under development.

4.7 Logs and Statistics

Document-access logs can indicate content relevance compared to other documents. Repository-access logs can be the qualitative/quantitative and technological basis for the evaluation of a document repository. Evaluation of a document-access log over the document's lifespan can give an indication of its continuous relevance. To enable comparisons of statistics of different Document and Publication Services, it is essential that standards exist and are adhered to. Today however, suitable standards for access statistics of institutional repositories do not exist. Differing protocols and access technologies create some of the problems of data acquisition and evaluation. No basis exists for a comparative evaluation of documents in different repositories, and of the Document and Publication Services themselves. It is recommended to store access statistics for at least five years.

DINI observes international developments, eg the standardisation project COUNTER (Counting Online Usage of Networked Electronic Resources) and the

project Interoperable Repository Statistics (IRS),[34] and will make new recommendations if necessary. Other approaches, such as the evaluation of link-resolver logs, will also be followed.[35]

4.8 Long-Term Availability

The Document and Publication Services provider has to guarantee the availability of a published document for no less than five years after the publication date (see 2.5). At the same time, long-term availability of these electronic resources has to be ensured. Compliance with recommendations by nestor – Network of Expertise[36] facilitates this. Within their respective financial and technological capabilities, publishing institutions can ensure this independently (OAIS depot system). Alternatively, administrative and organisational structures for external archiving can be created in cooperation with an archiving institution. The applicable minimum standards for the transfer of resources and the necessary metadata for archiving and usage must be kept (see 2.6).

At present, nestor – Network of Expertise does not make concrete recommendations on archival file formats, but used formats should be documented publicly to be applicable for long-term archiving. DINI-certified Document and Publication Services must archive all documents independently or be able to transfer them for that purpose. To ensure long-term archiving the original documents or the archive copies must be archived in an open document format. The original files uploaded by the author are not necessarily in one such archive format. Open formats will presumably be readable in the future, as their source codes are made public (as DIN/ISO or OASIS) and permission is granted indefinitely to develop them further. Examples are the Open Document Format (ODF), ASCII-Text (TXT), Hypertext Markup Language (HTML), Portable Document Format for Archiving (PDF/A) and TeX/LaTeX (TEX). Unambiguous naming of the file format beyond stating a mime type is important for decisions on necessary preservation strategies. References to publicly available databases containing file-format descriptions (File Format

---

34 http://irs.eprints.org/

35 J. Bollen, H. van de Sompel: An architecture for the aggregation and analysis of scholarly usage data, In: *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, 2006, pp 298 – 307, http://doi.acm.org/10.1145/1141753.1141821, http://arxiv.org/pdf/cs.DL/0605113

36 New information on appropriate measures and tools to secure long-term availability can be found on the website of 'nestor – Network of Expertise in Long-Term Storage of Digital Resources'. (www.langzeitarchivierung.de/index.php?newlang=eng)

Registries) are especially helpful. Additional technical metadata (information on coding, colour depth, etc) facilitate later application of long-term archiving techniques such as migration or emulation. These metadata can be automatically generated with a tool such as JHOVE. When exporting the actual object, these metadata should be appended to the object together with other metadata and information eg on earlier migrations. Dedicated data-package formats, eg the Uniform Object Format (UOF) based on METS,[37] are well suited to create a structured frame for the metadata. Copy-protection measures as used in Digital Rights Management (DRM) are completely unusable for archiving purposes if ruling out conversions (migration) and execution in other system environments (emulation).

Authors

Last Name	Institution	Email
Margo Bargheer	Niedersächsische Staats- und Universitätsbibliothek Göttingen	bargheer@sub.uni-goettingen.de
Kim Braun	Carl von Ossietzky Universität Oldenburg, Informations-, Bibliotheks- und IT-Dienste	kim.braun@uni-oldenburg.de
Susanne Dobratz	Humboldt-Universität zu Berlin, AG Elektronisches Publizieren, Universitätsbibliothek	dobratz@cms.hu-berlin.de
Dr Stefan Gradmann	RRZ der Universität Hamburg	stefan.gradmann@rrz.uni-hamburg.de
Prof. Dr Eberhard Hilf	Universität Oldenburg, Institute for Science Networking Oldenburg	hilf@physnet.physik.uni-oldenburg.de

[37] http://www.loc.gov/standards/mets

Dr Wolfram Horstmann	Niedersächsische Staats- und Universitätsbibliothek Göttingen	horstmann@sub.uni-goettingen.de
Prof. Dr Elmar Mittler	Niedersächsische Staats- und Universitätsbibliothek Göttingen	mittler@sub.uni-goettingen.de
Prof. Dr Peter Schirmbacher	Humboldt-Universität zu Berlin, Computer- und Medienservice	schirmbacher@cms.hu-berlin.de
Frank Scholze	Universität Stuttgart, Universitätsbibliothek	scholze@ub.uni-stuttgart.de
Dr Silke Schomburg	Hochschulbibliotheks-zentrum des Landes Nordrhein-Westfalen	schomburg@hbz-nrw.de
Thomas Severiens	Universität Osnabrück, Fachbereich Mathematik/Informatik	Severiens@Mathematik.Uni-Osnabrueck.de
Dr Heinrich Stamerjohanns	International University Bremen	h.stamerjohanns@iu-bremen.de
Tobias Steinke	Die Deutsche Bibliothek	t.steinke@d-nb.de
Bert Wendland	Humboldt-Universität zu Berlin, AG Elektronisches Publizieren, Computer- und Medienservice	bwendland@cms.hu-berlin.de
Christoph Ziegler	Technische Universität Chemnitz, Rechenzentrum	christoph.ziegler@hrz.tu-chemnitz.de

[Version 2.0, September 2006]

Imprint
DINI – Deutsche Initiative für Netzwerkinformation e.V.
– Geschäftsstelle –
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen
Papendiek 14
37073 Göttingen
Tel.: 0551-39-3866
Fax: 0551-39-3856
E-Mail: gs@dini.de
www.dini.de