DEUTSCHE INITIATIVE
FÜR NETZWERKINFORMATION E.V.

# DINI Certificate
# Document and Publication Services
# 2010

Working Group „Electronic Publishing“


DINI Schriften 3-en
[Version 3.0, March 2011]

DEUTSCHE INITIATIVE
FÜR NETZWERKINFORMATION E.V.

# DINI Certificate
# Document and Publication Services
# 2010

Working Group „Electronic Publishing“



This document is available online at
http://nbn-resolving.de/urn:nbn:de:kobv:11-100182800.

DINI Schriften 3-en

[Version 3.0, March 2011]

**Imprint**
DINI – Deutsche Initiative für Netzwerkinformation e. V.

DINI-Geschäftsstelle
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen
Papendiek 14
D-37073 Göttingen
Tel.: +49 551 39-3857
Fax: +49 551 39-3856
E-Mail: gs@dini.de
www.dini.de

# Content

## About DINI

The development of modern information and communication technologies causes a change in the information infrastructures of higher education institutions and other research institutions. This change is a major topic within higher education in Germany, and more than ever requires agreements, cooperation, recommendations, and standards. The Deutsche Initiative für Netzwerkinformation (DINI, German Initiative for Network Information) supports this development.

DINI was founded to advance the improvement of the information and communication services and the necessary development of the information infrastructures at the universities as well as on regional and national levels. Agreements and the distribution of tasks among the infrastructure facilities can significantly extend the range of information technology and of services. Additionally, the joint development of standards and recommendations is a requirement.

DINI is an initiative of three organizations:

- AMH (Arbeitsgemeinschaft der Medienzentren der deutschen Hochschulen; Consortium of German University Media Centers),
- dbv (Deutscher Bibliotheksverband Sektion 4: Wissenschaftliche Universalbibliotheken; German Library Association, Section 4: Academic Universal Libraries),
- ZKI (Zentren für Kommunikation und Informationsverarbeitung in Lehre und Forschung e. V.; Association of German University Computing Centers).

DINI has the following goals:

- Publicize and recommend best practices;
- Encourage and support the formulation, application and further development of standards as well as distribute recommendations regarding their application;
- Register and advertise Competence Centers using modern web-based technologies;
- Improve inter-disciplinary exchange through congresses, workshops, expert conferences etc.;
- Advertise new funding programs and encourage new programs.

## Summary

The global scientific communication system is currently undergoing fundamental changes. Due to the new possibilities that the internet and other information and communication technologies offer, and also to the scientists' and scholars' changing requirements, new distribution and dissemination channels appear in addition to the classical publishing houses. A leading development of the past years is the global Open Access movement that is committed to making scientific information, especially scientific and scholarly publications, available free of charge.

Most scientific and scholarly institutions have reacted to this and installed publication infrastructures in the form of so called document and publication servers, thus creating the possibility to make scientific and scholarly publications available online for a worldwide audience and to archive them. Not only the Wissenschaftsrat (German Science Council) and the Hochschulrektorenkonferenz (Conference of University Rectors in Germany) require this, but the funding organizations such as the Bundesministerium für Bildung und Forschung (BMBF; Federal Ministry of Education and Research) and the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) support this as a current task.

It is important that this development is in accordance with international standards and based on proven technologies. It is only in this manner that visibility and impact of the individual scientist's/scholar's work can be maximized and the research results of universities or other research institutions be globally and adequately recognized.

To accompany the numerous developments in Germany and to define general requirements for publication infrastructures DINI's working group for Electronic Publishing embraced this topic early on and in 2002 published its first recommendations for "Electronic Publishing in Higher Education"[1]. Based on these, the working group formulated criteria and formalized them in the "DINI Certificate for Document and Publishing Services". Following the 2004 and 2007 editions, 2010 is the third edition. The certificate describes technical as well as organizational and legal aspects that should be considered when setting up and running a sustained Document and Publication Service.

[1] Electronic Publishing in Higher Education – Recommendations, 2002, see http://nbn-resolving.de/urn:nbn:de:kobv:11-10045967.

While the 2004 edition's focus was on so called university document servers–meaning mostly university publishing platforms for theses and dissertations–the 2007 edition paid more regard to the global Open Access activities, especially the so-called "green road". The "green road" describes making a publication available online by self archiving them on institutional or disciplinary repositories parallel to or after publishing the same work elsewhere. This concerns most of all preprints and postprints of academic articles, but also other publication types such as monographs, research reports and conference proceedings. Additionally, the DINI Certificate 2007 edition addressed scientific and scholarly institutions of all kinds and emphasized the service character that should drive a Document and Publication Service[2].

This new 2010 edition especially incorporates the following aspects and developments:

- The growing **importance of the "golden road" to Open Access**. This second main strategy in Open Access is the primary publication of scientific articles in Open Access journals, but in principle includes Open Access publication of other document types (e. g. monographs, anthologies) as well. Analog to print publications these primary online publications usually undergo a quality assurance process for the most part in so-called peer-review processes. The German Science Foundation support Open Access publication with a structure-forming funding procedure. The 2010 edition of the certificate for Document and Publication Services also includes the golden road to Open Access along institutional (e. g. university publishing houses) or discipline specific lines.
- The increased demand for **interoperability with comprehensive services**. These services–especially search and catalog but also other added-value services–are especially important for locally provided publications. They are being implemented mostly within service-oriented infrastructures for electronic publications; on a national level this is the case in the Open Access Netzwerk[3] project, on a European level within the DRIVER[4] project. The quality of these services depends on the provided data and their standardization. Against this

---

2 The certificate's first edition of 2004 had the title DINI Certificate – Document and Publication Server.

3 See http://www.dini.de/projekte/oa-netzwerk/. A combined search in all participating Document and Publication Services can be found at http://oansuche.open-access.net/.

4 See http://www.driver-repository.eu/.

background, this certificate edition's guidelines for the OAI interface were amended and adapted to the DRIVER Guidelines[5].

- The growing **technical virtualization** of Document and Publication Services; this includes the use of one technical infrastructure by many Document and Publication Services. Despite this development, modularizing the certificate was refrained from. However, the option exists that more than one person can fill out the application form, should competencies and responsibilities for the operation of a Document and Publication Service be divided among more than one institution.
- A **comprehensive view of the scientific and scholarly research processes**. In addition to the text-oriented publications as the "classical" products of scientific and scholarly communication this brings their bases and production stages into the picture. Especially scientific data and research data resp., their organization and their use are of increasing importance for further research[6]. Science's ever progressing digitization and the virtual research environments it creates give rise to new possibilities in dealing with these data and open up innovative paths in research to scientists and scholars. The collaborative evaluation of research data in international and interdisciplinary projects is but one example for the possibilities of a digital research data management. The DINI Certificate's new edition, like the earlier editions, centers on Document and Publication Services that focus on the provision of text-oriented publications. However, the analog transfer of each individual criterion's requirements and recommendations allows for the certification of data-oriented publication services as well.

In addition to the above, this new edition represents a consolidation and a consistent development of the criteria and the requirements therein. The consequent new phrasing of the individual minimum requirements and the recommendations led to a convergence with the questionnaire that the applicants have to fill out, as they now read like a checklist.

This 2010 edition of the *DINI Certificate for Document and Publication Services*, the third edition, makes allowances for the listed current international developments and describes how a Document and Publication Services can support the Open Access publishing of a scholarly work in a standardized way. Modularizing

---

5 See http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final_2008-11-13.pdf.

6 DINI-Positionspapier Forschungsdaten (DINI Position Paper Research Data), Göttingen 2009, see http://nbn-resolving.de/urn:nbn:de:kobv:11-10098082.

the certificate was deliberately abstained from to emphasize the necessity of common standards and procedures in the scholarly research processes. The certificate's authors and reviewers are aware that discipline-specific factors might oppose a standardization in the narrower sense.
As the criteria for the DINI Certificate are evaluated and updated in accordance with international standards and developments by a working group, the certificate is labeled with the year of the certificate edition.

# 1 Aims

With the World Wide Web (www) science has created a communication tool that to a great extent is being used commercially. Despite the intensive use of the internet by academics in their everyday work the opportunities the web offers to improve scholarly communication are by no means exploited to their full extent. Electronic support of the whole scientific process is focused on more and more strongly: From first idea or draft to funding requests, to preliminary studies and experiments, questionnaires, measurements etc. to the publication of results and their application e. g. in patents or teaching scenarios.

The acceptance of electronic media is steadily growing; especially in the so-called STM areas[7]; the greater part of the research results is published electronically. But it is also in these areas of research where the quasi publishing monopoly of a few publishing houses has led to an exorbitant pricing development that is increasingly exceeding the possibilities of science funding. Setting up institutional or disciplinary publication services within the scope of the green and golden roads of Open Access can act as a regulative measure. Making the majority of academic publications available through non-commercial services would create a communication network that would at least make a profit maximization out of any proportion more difficult.

The DINI Certificate's catalog of criteria and the certification of a Document and Publication Service that is based on it has the following aims:

- Strengthen the service-oriented infrastructures for Open Access publishing.
- Define minimum requirements of Document and Publication Services and offer detailed descriptions.
- Establish a quality seal for Document and Publication Services facilitating the comparison of these services for users, service providers and funding bodies.
- Point out current and future development tendencies in the formation of services and the exchange of information.
- Position Document and Publication Services visibly as high quality services of an institution or discipline.

This third edition of the DINI Certificate is the consequent further development of the work done up to now with a clear description of criteria that ensure a service's adherence to national and international standards and developments. Services

7 STM = Science, Technology, Medicine. These are often contrasted with Social Sciences and the Humanities with regard to the publishing cultures.

that meet the minimum requirements may receive the certificate and show their quality within their own institution or discipline and also worldwide. Certified services are links in a growing chain of non-commercial content providers.

Certification of services as a sign that guarantees their worldwide interoperability in the science's growing communication network for high-quality publications does not have to be limited to universities and research institutions. Open Access publishers, data aggregators and data centers are invited to participate in the science communication of the future and apply for the DINI Certificate.

With the awarding of the certificate DINI facilitates a transparent quality control for Document and Publication Services. To achieve certification minimum requirements of the service and the service provider are listed. Meeting these requirements is a prerequisite for modern scientific and scholarly communication. At the same time, the DINI Certificate formulates recommendations based on foreseeable developments that are already visible today and might become minimum requirements in the future.

# 2 Criteria

The DINI Certificate comprises eight criteria that are described in detail in this section. The criteria are:
Criterion 1 – *Visibility of the Service* (section 2.1)
Criterion 2 – *Policy* (section 2.2)
Criterion 3 – *Support of Authors and Publishers* (section 2.3)
Criterion 4 – *Legal Aspects* (section 2.4)
Criterion 5 – *Information Security* (section 2.5)
Criterion 6 – *Indexing and Interfaces* (section 2.6)
Criterion 7 – *Access Statistics* (section 2.7)
Criterion 8 – *Long-Term Availability* (section 2.8)
The guidelines for the OAI interface provided in Appendix A of this document are also part of the DINI Certificate.
The individual criteria are split into two sections. In the first section minimum requirements (marked with an **M**) are specified, which must be met by the Document and Publication Services and their providers to qualify for certification. In addition to these, recommendations (marked with an **R**) are formulated. They serve as an orientation in the sense of best-practice solutions and hint at future tendencies in the development of Document and Publication Services. To qualify for certification with the current DINI Certificate it is not required to fulfill these recommendations. However, as DINI plans to continuously update the certificate it is likely that in later editions of the DINI Certificate some of these recommendations will be minimum requirements.
Each criterion is introduced by a short paragraph that explains the criterion and the reason(s) for its being a requirement. The requirements in the respective criteria are formulated like a check list to allow answering simply with yes or no. A grey backdrop signifies explanations of termini, interpretations or definitions, rationales or examples.
Complementing the remarks in this chapter, chapter 4 offers additional explanations and examples with regard to the criteria. However, these are not necessary to understand the certificate in the sense of a requirement that must be met.

## 2.1 Visibility of the Service

Greater visibility and a potentially higher recognition are characteristic advantages of electronic publications, especially when published Open Access. To make the most of this potential the entire range of an underlying Document and Publication Service's offers must be widely advertised. It has to be visible not only

to the immediate and individual user–regardless of whether one wants to read a specific publication or use it in another way, or if one wants to publish a document–but also to external services such as search engines or other referencing services. Besides the necessary technical interfaces (as described in criterion 6 – *Indexing and Interfaces* in section 2.6) the registration of a local service with the pertinent agencies is crucial. These agencies serve as facilitator between different, distributed Document and Publication Services and external add-on services.

### Minimum Requirements

| | |
|---|---|
| M.1-1 | The entire range of services must be available via a website. |
| | ⇒ This refers to a Document and Publication Service's main page from which both publication workflow and access to already published documents are possible. |
| M.1-2 | The service's homepage must be referenced in a central location on the institution's homepage. |
| | ⇒ Potential users must be guided mostly intuitively from an institution's, a research facility's or a library's central website to the Document and Publication Service. |
| M.1-3 | The service is registered and listed on the DINI website. |
| | ⇒ This DINI list can be found at http://www.dini.de/wiss-publizieren/repository/. This is also where a service can be registered. |
| M.1-4 | The service is registered with the Directory of Open Access Repositories (OpenDOAR). |
| | ⇒ Find OpenDOAR at http://www.opendoar.org/. This is also where a service can be registered. |

### Recommendations

| | |
|---|---|
| R.1-1 | The service is registered with the Registry of Open Access Repositories (ROAR) |
| | ⇒ Find ROAR at http://roar.eprints.org/. |
| R.1-2 | The service is registered as an OAI Data Provider with the Open Archives Initiative. |

⇒ A list of registered OAI data providers can be found at http://www.openarchives.org/Register/BrowseSites. If a repository's metadata are provided by an aggregating service (e. g. a library consortium) that service's OAI-interface should be registered with the Open Archives Initiative. See also criterion 6 – *Indexing and Interfaces*, section 2.6.

R.1-3 The service is registered with DRIVER.

⇒ DRIVER operates a reference service on a European level for publications available via institutional Document and Publication Services; see http://www.driver-repository.eu/.

R.1-4 All documents published using the Document and Publication Service are available via a hyperlink.

⇒ This facilitates the finding of a document by search robots (spiders). Documents that can only be found through a search request and are not available via a hyperlink will not be found by search engines.

## 2.2 Policy

Reliability and transparency play a major role when providing Document and Publication Services. It is crucial for the respective service provider to describe the offered services clearly and make statements on content related criteria and on the technical operations (e. g. on document types, intended users, sustainability of the service) in a publicly available policy. Such a policy represents the service provider's self-commitment towards the potential users of the services.

### Minimum Requirements

M.2-1 The service provider of a Document and Publication Service publicly provides a policy that describes the services.

⇒ The policy as the service provider's self-commitment is to be linked directly to the service's main page and must be a document of itself.

The policy contains the following items:

M.2-2 A definition of the Document and Publication Service's provider's rights and obligations.

| | |
|---|---|
| | ⇒ This includes a description of the services and statements on for whom and under what conditions it is provided. |
| M.2-3 | A definition of the authors' and publishers' rights and obligations when using the Document and Publication Services to publish their documents. |
| | ⇒ This includes e. g. a statement on what copyrights the user transfers to the service's provider. |
| M.2-4 | A description of the document types published via the Document and Publication Service, and requirements with regard to the documents' content and technical quality. |
| | ⇒ This corresponds to a collection mandate. The additional quality criteria referring to content quality (e. g. peer review) and technical aspects (e. g. file formats) serve primarily as orientation for potential users. |
| M.2-5 | A specification of the minimum timespan that documents published on the Document and Publication Service will be available, plus the respective guarantee. |
| | ⇒ The specified timespans do not have to be identical for all documents but can depend on document or publication type, or on a document's technical or content quality. However, the chosen value must not fall below five years. (See also Criterion 8 – *Long-Term Availability*, section 2.8) |
| M.2-6 | A statement on the long-term archiving of the documents. |
| | ⇒ This includes a description of how the long-term archiving of the publications is ensured, e. g. through the cooperation with another institution. |
| M.2-7 | A statement on the technical operation of the service. |
| | ⇒ This includes information on who is operating the document server technically, and the server's basic performance parameters (especially availability). |
| M.2-8 | A statement on Open Access. |

⇒ This statement must clarify the position of the Document and Publication Service's provider with regard to Open Access as well as point out those parts of the publications that might not be freely available in the sense of Open Access.

⇒ The majority of the publications provided by the Document and Publication Service must be available in the sense of Open Access.

⇒ Should the institution providing the service (e. g. a university) have published an Open Access declaration the Document and Publication Service's policy should refer to it.

**Recommendations**

Additionally, the policy contains statements on the following:

R.2-1 Guidelines and recommendations for authors with regard to Open Access.

⇒ This is especially useful in a policy if the providing institution recommends or intends a certain practice, e. g. the self-archiving of publications (the "green road"), as published in an institutional Open Access declaration. Guidelines may vary according to document or publication type.

E.2-2 Naming and description of the Document and Publication Service's tools.

⇒ This can include e. g. the repository software, upload interfaces, versioning and authentication procedures as well as automated license definitions (for primary publications, the "golden road").

## 2.3 Support of Authors and Publishers

The aim is to support the entire publication process within the Document and Publication Services. For those making use of the services to publish (i. e. authors and where applicable publishers) visible and well-structured information, that answers the most relevant questions on electronic publishing, are important The relevant pages must be accessible via the Document and Publication Service's

website and may additionally be available in other formats (e. g. flyers, brochures). The information may include external resources[8].

### Minimum Requirements

| | |
|---|---|
| M.3-1 | A contact and an advisory service are accessible via the website. |
| | ⇒ The contacts may be email addresses, phone numbers etc. or contact forms on the web pages. It is not required that all the above listed options are available, but at least one is mandatory. |
| M.3-2 | Authors have the option to upload their documents intended for publication directly onto the repository (e. g. a web form) or use other ways to enter the documents into the repository. |
| | ⇒ This requirement is obsolete, where the entire upload process of documents is carried out by a service institution (e. g. a library). |
| M.3-3 | Information on the relevant technical questions on electronic publishing are provided or linked to. |
| | ⇒ This includes especially advice on and practical help for the use of applicable file formats and how to enter electronic documents in the publication server. |
| M.3-4 | Relevant information resources with regard to copyright questions (e. g. about secondary publication as Open Access) are referenced. |
| | ⇒ Most prominent among these resources is the SHERPA/RoMEO list. |

### Recommendations

| | |
|---|---|
| R.3-1 | The SHERPA/RoMEO list API is included in the upload interface. |
| | ⇒ This allows authors to research the usage rights they still hold after a previous (primary) publication of their documents with a publishing house directly during the upload process. For further information see http://www.sherpa.ac.uk/romeo/api.html. |

[8] An example within the German context is the information platform DissOnline, see http://www.dissonline.de/

| | |
|---|---|
| R.3-2 | As an alternative to the independent upload by the authors/publishers a central institution offers an upload service to authors/publishers. |
| | ⇒ This service can be offered by the Document and Publication Service's provider, or the library resp. It can vary depending on the publication type. |
| R.3-3 | Document templates or style sheets for the creation of scientific documents are made available. |
| | ⇒ These should be templates that can be used in word processing or layout software, e. g. OpenOffice, Winword or LaTeX. |
| R.3-4 | Courses on electronic publishing for authors are offered on a regular basis. |
| | ⇒ Possible topics are the use of document templates or style sheets, citing of electronic resources and the conversion of documents into applicable file formats, e. g. PDF/A. |
| R.3-5 | To support publishers of extensive publication projects a workflow system is offered. |
| | ⇒ This encompasses primarily systems facilitating a peer review for electronic journals or scientific conferences. |
| R.3-6 | A helpdesk system is used to answer inquiries. |
| | ⇒ This helpdesk system is an information service that utilizes a so-called trouble-ticket system to answer user requests. |
| R.3-7 | Support is given with regard to adequate usage and citation of electronic documents. |
| | ⇒ This should include e. g. an explanation that electronic publications should best be cited by using a Persistent Identifier, or how to cite selected parts of a publication that do not have page numbers. |
| R.3-8 | The available information or parts thereof are provided in English. |
| | ⇒ This is advised especially when addressing authors and/or publishers whose native language is not German. |

## 2.4 Legal Aspects

The provider of a Document and Publication Service requires certain usage rights to offer documents to the public and to facilitate their long-term archiving. These must be granted by the author(s) or publisher(s). This is done in a formal agreement, the so-called deposit license. In this agreement it must also be regulated that no third party's rights are violated and that the service's provider is exempt from any liability should a third party's rights be violated.

These and other legal aspects that must be observed when operating a Document and Publication Service are subject of this criterion. No statement or remark in this section/criterion is to be understood as legal advice or legally binding information. All providers of a Document and Publication Service are advised to cooperate with their institution's legal department and to seek additional professional advice where legal aspects are concerned.

### Minimum Requirements

M.4-1 The legal relationship between authors and publishers (rights holders) and the provider of a Document and Publication Service is regulated in a formal agreement (grant of rights).

⇒ This agreement includes the granting of rights and is formalized as a deposit license. The rights holder grants those non-exclusive usage rights to the provider of the Document and Publication Service that are necessary to provide the service with regard to the document to be published.

M.4-2 The service provider publishes the deposit license in German on the service's web site.

⇒ The deposit license may vary depending on the type of publication.

By agreeing to the deposit license the rights holder grants the following usage rights on a document and its metadata to the Document and Publication Service's provider.

M.4-3 The right to store the publication electronically, especially in databases, to make the publication available to the public and to disseminate it, as well as to send it, to archive it and to reproduce it for these purposes.

| | |
|---|---|
| | ⇒ Within the boundaries of this agreement users may use these documents free of charge in accordance with the copyright laws, specifically download the document for private purposes, store it and print it in small quantities. (See the relevant paragraphs in the resp. copyright laws; e. g. §53 in the German copyright law.) |
| M.4-4 | The right to notify and transfer the document to third parties e. g. within the framework of national collection mandates, especially for the purpose of long-term archiving. |
| | ⇒ This includes (in Germany) the transfer of documents and metadata to the German National Library. |
| M.4-5 | The right to copy and to convert the document for archiving purposes into additional, different electronic or physical formats while retaining the content's integrity. |
| | ⇒ A conversion may e. g. become necessary should the data/file formats used become obsolete and current presentation/viewing software be unable to present the document correctly. |

The deposit license also regulates questions concerning liability. In detail these are:

| | |
|---|---|
| M.4-6 | The rights holder assures the service provider that no third party's (e. g. co-publishers, publishing houses, sponsoring bodies) copyrights will be violated by publishing a document or parts thereof (e. g. photographs). |
| | ⇒ In the case of a so-called author's copy this is usually regulated in the existing author-publishing house contract (see section 4.4.1). |
| M.4-7 | The rights holder must ensure immediate notification of the service provider in case of doubt or of alleged or actual legal hindrances. |
| | ⇒ This is e. g. the case should the author/publisher at the time of granting a right to the service provider not have been in possession of said rights. |
| M.4-8 | The rights holder exempts the service provider of any third-party claims. |

| | |
|---|---|
| | ⇒ Third party claims can be caused by the use of copyright protected materials (photographs etc.) in a document. |

Additional minimum requirements

| | |
|---|---|
| M.4-9 | An imprint is published on the website that (in Germany) complies with the Telemediengesetz (TMG, Tele-media Law) and other applicable laws. |
| | ⇒ This includes especially state laws. |
| M.4-10 | The service provider documents the legal situation in the published documents' metadata. |
| | ⇒ Information on what rights were granted to the service provider is stored with each document. It is not necessary to make these information publicly available. |

**Recommendations**

| | |
|---|---|
| R.4-1 | The deposit license was written in cooperation with the service provider's legal department or an attorney. |
| | ⇒ This is a very strong recommendation to ensure the agreement's validity in granting the necessary copyrights to the provider. |
| R.4-2 | The service provider offers an English version of the deposit license on the service's website. |
| | ⇒ While the German language version is the legal basis for the agreement, the English version serves as an orientation. |
| R.4-3 | Within the agreement between rights holder and service provider the right is granted to authorize a third party with realizing the public availability of a document. |
| | ⇒ This right is necessary e. g. in the case that a service is (partially) discontinued, and a document's public availability guaranteed through a third party (e. g. an institution specializing in long-term archiving). |
| R.4-4 | During the upload process, the author/publisher has the option of choosing from a list of preselected licenses that grant different rights to the service provider. |

| | |
|---|---|
| | ⇒ This offers the rights holder the opportunity to grant rights in excess of those granted in the standard license–e. g. for commercial use. It is advised to use standard licenses, esp. Creative Commons, DPPL. |
| R.4-5 | The service provider actively supports the upkeep of the SHERPA/RoMEO database. |
| | ⇒ This includes foremost the relaying of information on a publishing house's policy with regard to Open Access (green road). |

## 2.5 Information Security

To guarantee a reliable Document and Publication Service that satisfies the general requirements of scientific publishing[9] the underlying technical system[10] and the organization structure must fulfill basic criteria with regard to information security. These are specified in the *Common Criteria* as published in the international standard ISO/IEC 15408[11]. Main contents are fail safety, operational safety, and trustworthiness of the technical infrastructure, as well as availability, integrity and authenticity of the published documents. The Document and Publication Service must be secure against attacks, misuse, operating errors, and technical malfunctions and failures. To guarantee this, organizational and technical measures must be taken.

### 2.5.1 Technical System

#### Minimum Requirements

| | |
|---|---|
| M.5-1 | The technical system that is the basis for the Document and Publication Service is integrated into the provider's institutional security concept. |
| | ⇒ This concept identifies and qualifies possible risks and describes technical, organizational a personnel-related provisions to adequately counter these risks. |

9 These are especially permanent availability, unchangeability and citeability of the published documents, and the trustworthiness of the entire technical system.

10 Among these are usually at least the storage system, data base system, communication networks, web server, and all applications necessary for the operation of the service.

11 See http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html to download the standard.

| | |
|---|---|
| M.5-2 | A person is named as responsible for the technical systems security. |
| | ⇒ This person is appointed by the service's provider and is the central point of contact for all questions concerning the system's security. |
| M.5-3 | An operational concept exists that includes regulations on the systems maintenance. |
| | ⇒ The operational concept contains descriptions of all tasks, actions and processes necessary to operate the system, as well as the corresponding roles and interfaces. |
| M.5-4 | A written documentation exists on the technical system and all of its components that are necessary for the operation of the system. |
| | ⇒ This documentation does not have to be published (at least not in toto). Security-relevant elements are only for internal use. |
| M.5-5 | Written regulations exist on the access to the technical system and its components. |
| | ⇒ This includes who has access to the facilities and who has extended user/administration rights. |
| M.5-6 | All data and documents are regularly saved in a back-up procedure. |
| | ⇒ At what interval back-ups are run depends to a great extent on how often changes are made in the data, i. e. how often new publications are uploaded. It is advised to run a daily back-up procedure. |
| M.5-7 | A damage-control concept is in place. |
| | ⇒ This concept includes procedures for possible malfunctions of the technical system. |
| M.5-8 | Autonomous software regularly monitors the availability of the servers that are necessary for the service's operation. |
| | ⇒ Applicable documentation procedures for changes applied to hardware or software configurations are e. g. use of autonomous monitoring and alerting software as well as the keeping of change logs. |

### 2.5.2 Document Handling

**Minimum Requirements**

| | |
|---|---|
| M.5-9 | Documents uploaded into the repository will not be changed. |
| | ⇒ Changes on the content of published documents will be considered additional editions that do not overwrite or render obsolete earlier editions. |
| M.5-10 | Every document and every edition/version is assigned a Persistent Identifier (PI). |
| | ⇒ Available PI systems are e. g. URN and DOI. |
| M.5-11 | Persistent Identifiers are indicated on the service's web pages and in the exported metadata as primary identifiers in the form of an operable URL. |
| | ⇒ This requires a resolving service's URL to be added to the Persistent Identifier. As for the metadata export see also criterion 6 – *Indexing and Interfaces*, section 2.6.2, minimum requirement M.6-7. |
| M.5-12 | Deletion of documents is done only as an exception and documented in writing. |
| | ⇒ This could be the case should the publication be a criminal offense. |
| M.5-13 | The data transfer during a document's upload is via SSL and on the basis of a trustworthy certificate. |
| | ⇒ This requirement is obsolete, should the Document and Publication Service not offer the option to upload documents (see criterion 3 – *Support of Authors and Publishers*, section 2.3, minimum requirement M.3-2) |

**Recommendations**

| | |
|---|---|
| R.5-1 | The individual document's integrity is guaranteed and verified through the creation and online publication of a hash value. |
| | ⇒ A secure hashing method must be applied, currently e. g. SHA-1 or RIPEMD 160. |
| R.5-2 | The data and document transfer from web server to user is via SSL and on the basis of a trustworthy certificate. |

| | |
|---|---|
| | ⇒ This refers to read-only access to the Document and Publication Service. |

## 2.6 Indexing and Interfaces

To find a document that is published electronically outside the local system it is crucial that it is indexed with descriptive metadata and that these metadata are available for machine-based processing. At the core of this are reference and other additional services that third parties provide by applying the data and documents provided by the Document and Publication Service. Local search options and additional services are integral parts of a Document and Publication Service. This criterion describes the prerequisites to meeting the requirements.

### 2.6.1 Indexing

**Minimum Requirements**

| | |
|---|---|
| M.6-1 | A written policy exists containing the indexing regulations for documents, which is available online to user of the Document and Publication Service (authors, publishers and readers). |
| | ⇒ It is e. g. of relevance who does the indexing–library personnel or the authors–or if it is done automatically. |
| | ⇒ These regulations may vary depending on the publication type. |
| M.6-2 | Every document is represented in an indexed form that employs the means and methods of the Dublin Core element set. |
| | ⇒ It is not mandatory that these metadata are also stored internally in this format. |
| M.6-3 | A verbal subject indexing with uncontrolled keywords or keywords from a classification system is done for every document. |
| | ⇒ Keywords may be assigned directly by the author. |
| M.6-4 | All documents are classified with the Dewey Decimal Classification (DDC) at least in accordance with the German National Library's subject headings. |
| | ⇒ See http://www.ddc-deutsch.de/anwendung/dnb.htm and section A.2.2. |

| | |
|---|---|
| M.6-5 | All documents are assigned document or publication type descriptions following DINI's recommendations in *Common Vocabulary for Publication and Document Types* (*Gemeinsames Vokabular für Publikations- und Dokumenttypen*) |
| | ⇒ See http://nbn-resolving.de/urn:nbn:de:kobv:11-100109998 and section A.2.3. |

**Recommendations**

| | |
|---|---|
| R.6-1 | Additional standardized systems are used for the verbal or classificatory subject indexing. |
| | ⇒ Examples are SWD, LoC Subject Headings, CCS, MSC and PACS. |
| R.6-2 | In addition, English keywords are assigned. |
| | ⇒ Keywords may be assigned directly by the author. |
| R.6-3 | Additional short summaries or abstracts in English and German are provided. |
| | ⇒ These may be requested from the authors or extracted from the full texts. |

### 2.6.2 Metadata Interfaces

**Minimum Requirements**

| | |
|---|---|
| M.6-6 | A web interface exists that allows all users to access all documents and metadata. |
| | ⇒ Via this website the complete assets provided by the Document and Publication Service can be obtained. |
| M.6-7 | An OAI interface is available that complies with the OAI PMH 2.0 and the DINI OAI Guidelines. |
| | ⇒ The guidelines for the OAI interface can be found in Appendix A of this document. |

**Recommendations**

| | |
|---|---|
| R.6-4 | The metadata (e. g. of parts of the holdings) are provided in additional metadata formats and are available via the OAI interface. |

| | |
|---|---|
| | ⇒ These may be subject or publication-type specific metadata formats for relevant technical or archiving information that facilitate additional services by third parties: One of these is the XMetaDissPlus for the delivery of dissertation metadata to the German National Library. |
| R.6-5 | A direct export of metadata records or search results in adequate data formats is available on the website. |
| | ⇒ These are among others BibTex, EndNote or micro formats such as COinS. This option facilitates the import into reference-management programs. |
| R.6-6 | Link lists are available to access every document. |
| | ⇒ This permits third parties' robots to index the entire data. |
| R.6-7 | Metadata are made publicly available via additional interfaces. |
| | ⇒ E. g. SRU/W or specified APIs. |

## 2.7 Access Statistics

Server-based access statistics can be the qualitative, quantitative or technological basis for the evaluation of a Document and Publication Service. On the level of individual objects (e. g. a document) usage information on electronic documents can reflect a document's impact–be it as an original usage impact that may be taken as complimentary to other impact concepts (e. g. a citation) or as a predictor for citations. In addition to this, object-related usage information may in the future help detect usage cycles of scientific information–even broken down to different disciplines–and enrich scientometric analyses.

### Minimum Requirements

| | |
|---|---|
| M.7-1 | The Document and Publication Service keeps a consistent access log in accordance with the legal regulations. |
| | ⇒ This is usually a web-server log. |
| M.7-2 | Web-server logs are anonymized or pseudonymized for long-term storage. |
| | ⇒ This is mandated in the legal regulations in §15;3 in combination with §13;1 (German) Telemedia law. |

| | |
|---|---|
| M.7-3 | Automatic access is not taken into account for the usage statistics on the individual documents or data on a Document and Publication Service. |
| | ⇒ This can be done e. g. by evaluating the web-server log's user-agent field, by comparing accesses to the robots.txt, by using lists of known robots, or by employing heuristic methods. |
| | ⇒ This is only a requirement, if the statistics are published. |
| M.7-4 | A publicly available documentation exists describing the criteria and standards applied in creating the statistics. |
| | ⇒ Among these standards are COUNTER[12], LogEC[13] and the IFABC guidelines[14]. If access values are published that are not reached by any of these standards, the documentation must contain a paragraph stating that these values are not comparable to those of other Document and Publication Services. This is especially the case, if access values per document are listed. |
| | ⇒ This is only a requirement, if the statistics are published. |

**Recommendations**

| | |
|---|---|
| R.7-1 | Access statistics are added to every document as dynamic metadata and are publicly available. |
| | ⇒ Access values (e. g. in chronological order) could be linked to from a document's start page. |
| R.7-2 | Access to documents is counted according to a standard recommended by DINI. |
| | ⇒ Among these standards are COUNTER, LogEC and the IFABC guidelines. See also the German Science Foundation (DFG) project Open Access Statistics (OA-S) and the DINI publication Usage Statistics of Electronic Publications (http://nbn-resolving.de/urn:nbn:de:kobv:11-100101174). |

[12] COUNTER = Counting Online Usage of Electronic Resources, v. http://www.projectcounter.org/.

[13] See http://logec.repec.org/

[14] See http://www.ifabc.org/

| | |
|---|---|
| R.7-3 | Data transfer to a service provider as developed in the OA-S project are supported. |
| | ⇒ The external service provider harvests the web-server log's data via an OAI interface to calculate the access statistics using a standardized method. See http://www.dini.de/projekte/oa-statistik/. |

## 2.8 Long-Term Availability

This certificate focuses on Document and Publication Services and not on digital long-term archives as dealt with in the *Catalog of Criteria for Trustworthy Digital Long-Term Archives* of nestor[15]. However, certain questions on long-term archiving are also valid for Document and Publication Services, especially since the published documents are often transferred to a long-term archiving institution, which requires adequate pre-conditions be met.

### Minimum Requirements

| | |
|---|---|
| M.8-1 | A minimum time span of no less than five years is defined for the availability of documents and their resp. metadata published through the Document and Publication Service. |
| | ⇒ This definition must be element of the Document and Publication Service's policy (see criterion 2 – *Policy*, section 2.2, minimum requirement M.2-5). The predefined availability times may vary for different publication types. |
| M.8-2 | The original files and possible additional archival copies are free of any technical protection. |
| | ⇒ This includes especially mechanisms in the sense of a Digital Rights Management (DRM), password protection, or limitations regarding the use of the document (copy and paste, printing). Protective measure are barred, as they might interfere with long-term archiving strategies (e. g. migration, emulation). |

[15] For the current version 2 of this document see http://nbn-resolving.de/urn:nbn:de:0008-2008021802.

## Recommendations

| | |
|---|---|
| R.8-1 | Long-term availability of the documents is ensured. |
| | ⇒ This can be done in cooperation with an archiving institution. |
| R.8-2 | For the documents' storage, open file formats are used that facilitate long-term availability. |
| | ⇒ This includes PDF/A, ODF, TXT, TEX. |
| E.8-3 | The deletion of documents is regulated. |
| | ⇒ This regulation includes the conditions and the procedures for the deletion of documents, and on the data that might have to be stored beyond a date of deletion. This definition must be element of the Document and Publication Service's policy (see criterion 2 – *Policy*, section 2.2). |

# 3 Awarding and Evaluation

The German Initiative for Network Information (DINI) or a working group authorized by DINI is responsible for the awarding of the *DINI Certificate for Document and Publication Services*. The certificate's seal shows the year of its version. The certificate acknowledges that the certificated repository meets the minimum requirements for a DINI-certified Document and Publication Services.

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

1. Non-profit organizations
   - DINI members 50,00 €
   - others 100,00 €
2. Profit organizations
   - DINI members 150,00 €
   - others 250,00 €

The provider of the Document and Publication Services applies at DINI for certification by completing an online form on the DINI website[16]. This form has the structure of a checklist and contains the minimum requirements as well as the recommendations laid down in section 2 of this document. By completing the form the provider states that and to what extent the Document and Publication Service fulfills the criteria of the DINI Certificate. Further explanations and clarifications can be added in designated fields in the form, as well as URLs or other options on how or where to receive additional information.

After the online form has been completed and submitted the application and the containing data will be verified; generally two reviewers will be appointed for this. Access to the services to be certified must be permitted to these two. The provider of the Document and Publication Service must be prepared to answer questions from reviewers. An on-site visit will be the exception. Additional costs that may emerge during the certification process must be covered by the provider of the Document and Publication Service. DINI will inform the provider about possible additional costs beforehand.

The certification process should generally be completed within two months. The duration of the certification process depends in part on how quickly the provider answers questions the reviewers might have. The process can take longer should one or more criteria not be fulfilled.

---

[16] See http://www.dini.de/dini-zertifikat/fragebogen/.

The DINI Certificate does not expire for the individual Document and Publication Service. As the certificate shows the year of the version, it will always be clear under what standards a Document and Publication Service is certified, even if a newer certificate version exists. In cases of failing minimum requirements after a certification, DINI is entitled to revoke the certificate.
The provider of the certified Document and Publication Service is entitled to call the service 'DINI-certified Document and Publication Service', and to display 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.
With the publication of this document (in German) the DINI Certificate will only be granted for the 2010 version of the certificate.

# 4 Explanations and Examples

Additional explanations and practical examples of how the DINI Certificate's eight criteria–as formulated and explained in detail in section 2–can be met are listed below. These serve as illustrations but are explicitly not additional criteria or requirements and do not appear in the application form for the DINI Certificate.

## 4.1 Visibility of the Service

On the one hand, a Document and Publication Service must be visible within the institution, organization or learned society to which it is offered. It has to be obvious to potential authors and publishers that it offers a trustworthy service for the publication of scientific documents.

On the other hand, the service–i. e. especially the published documents–must have a high degree of visibility for potential readers and researchers. This requires especially the service's integration in meta-services that function as intermediaries between the users and relevant publications on any number of different servers.

To increase visibility of and knowledge about one's own Document and Publication Service among meta-services it is useful to register with entities such as DRIVER, the Open Archives Initiative or OpenDOAR. At the core of such a registration would be the base URL of the local OAI interface, which permits meta-services standardized access.

It is not only dedicated services such as DRIVER or the Open Access Network that facilitate realizing the global visibility. Commercial service providers, especially search engine operators, play a major role in finding scientific and scholarly publications, but rarely use the OAI protocol to aggregate the metadata. To support the spider technology they use all documents including their resp. start pages should be reachable via a hyperlink.[17] The bi-annual publication *Ranking Web of World Repositories* offers an indication of how well search engines manage to present an individual Document and Publication Service's content.[18]

---

[17] These must be accessible as directly as possible from a Document and Publication Service's start page and should contain short and few HTTP parameters, as some search engine spiders tend to ignore those.

[18] The Cybermetrics Lab in Madrid creates this ranking and besides the number of documents takes the number of external links to the documents into account. Conditions for the inclusion in the ranking and information on how to improve one's own service's ranking can be found at http://repositories.webometrics.info/.

## 4.2 Policy

The Document and Publication Service's policy should be a complete text and have the character of a self-commitment by the service operator addressing the service's potential users. In this sense, an FAQ does not constitute an adequate alternative to a policy, but can be offered as an add-on.

The following hyperlinks lead to examples of policies of existing Document and Publication Services.

- Document and Publication Service of the Humboldt-Universität zu Berlin (edoc server): http://edoc.hu-berlin.de/e_info/leitlinien.php
- Saxonian Document and Publication Service (Qucosa): http://www.qucosa.de/ueber-qucosa/
- Document and Publication Service of the University of Kassel (KOBRA): https://kobra.bibliothek.uni-kassel.de/policy.jsp

## 4.3 Support of Authors and Publishers

An important goal of electronic publishing at scientific and scholarly institutions or in scientific communities is to make the major part of those scientific and scholarly outputs more available where members of the resp. institutions were involved–be they members of the university, of a research institution or of a disciplinary society. The target group of a Document and Publication Service on the productive side, the authors and publishers that is, are to play an important role for the creation and operation of the service. This requires a comprehensive support of this target group.

The kind of support depends on the kind of publication and the aim of the Document and Publication Service.

Information on the internet or in other form that offers at least a description of the service and its policy are mandatory. The same is true for contact data e. g. an email address. The option of online publication via a web form should be available to authors. Where the initiative and actual publication of scientific and scholarly documents is centered at one central service institution–often the library–such a self-service function may be omitted.

A comprehensive Document and Publication Service offers not only the upload process, but can also support the creation of publications or the organization of a publication project. This can range from direct help for authors in writing a structured scientific document (e. g. as an institutional training course), to providing styles or document templates as starting points for high-quality documents (based e. g. on XML-based document models), to offering workflow systems for recur-

ring publication processes (such as the peer-review process of scientific journals), and to the comprehensive assistance for publication projects and the consequent consulting for the publishers. Especially the latter enters into the realm of publishing houses, and in the ideal case is a seamless addition to the Document and Publication Service described in this document.
The support of authors and publishers plays an increasingly important role parallel to the growing importance of the Golden Road in Open Access publishing. In many such cases the publishers are the Document and Publication Service's direct partners and act as intermediaries to their authors.

## 4.4 Legal Aspects

### 4.4.1 Authors' Rights

Regarding the authors' rights a difference exists between original (primary) publications and so-called authors' copies (parallel publications) of already published documents. In case of the former the authors usually still hold all rights; in case of the latter these rights have often been transferred to a third party with the signature of a contract. However, many publishing companies permit within the guidelines of their respective policies or conventions the provision of authors' copies on institutional or disciplinary document repositories. The SHERPA/RoMEO database offers a (legally not binding) overview of these policies and links to the respective publisher contracts.
Document and Publication Service providers should only require the non-exclusive usage right (as laid down in the deposit license) to allow rights holders further dissemination and exploitation of their works (e. g. on disciplinary repositories, personal web sites, or through publishing houses).

### 4.4.2 Third Party Rights

In the contract/agreement between the service's provider and the rights holder it must be foreclosed that any third party's rights are violated, and the service's provider should be excluded from any liability. To ensure this, rights holders must

- ensure the service provider that a document that is to be published or parts thereof (e. g. pictures) do not violate any third parties' (e. g. co-creators, publishing houses, funding agencies) rights,[19]

[19] In the cases of the so-called author's copy this is usually already regulated in the author contracts.

- ensure immediate notification of the service provider in case alleged or actual legal hindrances to the publication of a document or parts thereof emerge,
- exclude the service provider from any legal claims by third parties.

### 4.4.3 Hybrid Publications with an Institutional Publishing House

Should an institution offer an additional print publication to the Open Access online publication (hybrid publication), it should evaluate whether or not to request the exclusive usage rights for the printed edition (for a limited time) from the authors/rights holders to secure its investment in that publication. This can include the obligation for authors/rights holders to abstain for a defined time period from the reproduction, dissemination and the incorporeal transmission of the work (abstention obligation, avoidance of multiple registrations with VG Wort). Creation of the work or parts thereof in a print on demand process should be regulated. In case of a standard print publication (ISBN, sales and distribution through booksellers) authors and publishing institution should sign a contract that clearly regulates all usage and copyright questions.

### 4.4.4 Deposit License Example

The following two examples can be used as basis for the creation of an institution's individual deposit license.

- Deposit License for an institutional repository: Goescholar, University of Göttingen, http://goedoc.uni-goettingen.de/goescholar/help/rights_de.jsp
- Deposit License for a disciplinary repository: ART-Dok – Publikationsplattform Kunstgeschichte, http://archiv.ub.uni-heidelberg.de/artdok/erklaerung.php

### 4.4.5 Copyright Development

The current (German) copyright supports research and teaching in networking communication and information environments only inadequately. The Coalition for Action *Copyright for Education and Research* (http://www.urheberrechtsbuendnis.de/) and the communication platform *IUWIS – Infrastruktur Urheberrecht für Wissenschaft und Bildung* (http://www.iuwis.net/) are lobbying for the copyright's development in the science's and research's interest.

### 4.4.6 Additional Reading

Literature on copyright abounds. Examples focusing on German law are:

- Gesetz über Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetz), see http://www.bundesrecht.juris.de/urhg/BJNR012730965.html.
- Spindler, Gerald (Ed.): Rechtliche Rahmenbedingungen von Open-Access-Publikationen, Göttinger Schriften zur Internetforschung – Band 2,

Universitätsverlag Göttingen 2006, http://www.univerlag.uni-goettingen.de/OA-Leitfaden/oaleitfaden_web.pdf.

- Hoeren, Thomas: Internetrecht, Universität Münster 2010, regularly updated script. http://www.uni-muenster.de/Jura.itm/hoeren/INHALTE/lehre/lehrematerialien.htm.

Additional literature can be found at the Information Platform open-access.net (http://open-access.net/de/allgemeines/rechtsfragen/).

## 4.5 Information Security

### 4.5.1 Technical System

The **documentation** should include the exact names of individual hardware and software components and their respective manufacturers' names. Hardware components should be described listing their relevant data (speed, storage size, etc.), software components with their version numbers, configuration parameters etc. The system documentation is especially important for proprietary developments.

**Regulations on access to the system** should at least include the following:

- Names of personnel with physical access to the server,
- Names of personnel with administrative rights,
- Names of personnel responsible for the technical system or individual components (e. g. database system) incl. substitutes,
- Location of administrator passwords for the system and the relevant servers.

**Back-ups** have to be done on a regular basis in a way that allows **data recovery** with minimal or no losses after a complete system failure.[20] Using a central back-up and automatic back-up service is recommended.

For every server malfunction or dysfunction–from the failure of individual software or hardware components to the loss of an entire server regardless of severity– a **damage-control concept** and action plans should be prepared to ensure resumption of operations in the shortest possible time. Possible system failures include technical problems or damage caused by faulty handling or attacks from outside (hacking).

[20] A regular back-up alone cannot guarantee full recovery of lost data. Full data security (including new data stored on a system between the last backup and an incident) can only be reached by dual data-storage on the basis of a RAID system or a redundant Storage Area Network (SAN).

An **autonomous monitoring and alerting software**[21] continuously monitors the server's operation as a whole, as well as individual services (web pages, database functions, etc.). In case of a malfunction, an alerting service will send out an email or an SMS.

### 4.5.2 Document Handling

A **Persistent Identifier** (PI) must be assigned to every document. This ensures access to the document regardless of changes to the software system used or to an underlying structure. PIs are especially useful for the citation of electronic publications, as they are–other than URLs–stable and persistent.

To use a PI the general schema for the creation of PI's in and the namespace of the respective systems must be known. All PIs within a system must be unambiguous. The PI schema should meet the requirements laid down in RFC 1737[22]. All information about such a trustworthy PI system should be publicly available, documented, and listed in a Persistent Identifier registry (e. g. the *IANA Registry* for URN namespaces[23]). Additionally, a functioning resolving service must exist for the PIs. Finally, every PI, the corresponding schema, and the assigned URL must be listed in a document's metadata, and registered with or accessible via the resolving service.

A possible provider of resolving services within the context of non-commercial Document and Publication Services is the URN-NBN system[24] supported by the Deutsche Nationalbibliothek (DNB, German National Library)[25]. To use URNs the following prerequisites must be fulfilled:

- The service provider must apply for a sub-namespace within the valid national namespace in Germany (urn:nbn:de) at the DNB.
- Within the assigned sub-namespace the service provider may independently create and assign valid[26] URNs to URLs via which the documents will be accessible. The service provider is responsible for keeping the URNs unambiguous,

---

[21] One such software product is Nagios (http://www.nagios.org/).

[22] See http://www.ietf.org/rfc/rfc1737.txt.

[23] See http://www.iana.org/assignments/urn-namespaces/.

[24] The globally used PI system of the National Bibliographic Number (NBN) uses the URN schema to create the PIs and is supported by the National Libraries.

[25] Additional information can be found at http://www.persistent-identifier.de/ and http://www.d-nb.de/netzpub/erschl_lza/np_urn.htm.

[26] A check digit is located at the end of a valid URN.

i. e. avoid multiple assigning of a single URN, and for recording changes to the URLs (especially but not only the deletion of documents) and storing these changes internally with the respective URN. Functions for the automatic creation and assignment of URNs are included in many of the existing software systems for Document and Publication Services.

- Assignments of URNs to URLs–including possible changes–must be relayed to the DNB regularly to permit the DNB to guide users of the central resolving service[27] to the actual documents. To facilitate this, a registration system was developed in the project EPICUR, which today uses an OAI-based procedure. For this the Document and Publication Service must provide a specific OAI interface that offers the required data for a URN in the EPICUR format. URN-URL mapping can also be relayed to the DNB via a web-based form or an email attachment.[28]

However, use of the URN-NBN system is not mandatory. Digital Object Identifier (DOI), Persistent URL (PURL), Handle as well as Archival Resource Key (ARK) are also common systems.[29]

**Hashing procedures** are one answer to the special challenges of data storage and transmission in general and of electronic publishing specifically that lie in the convertibility of electronic documents. Risks to a document's integrity are conscious manipulation as well as unconscious alterations, e. g. through external factors or hardware failures.

So-called hashing procedures serve to appraise a document's integrity. For any kind of file they create small, distinct data packages that can be viewed as digital finger prints, as even minute changes to the original file lead to a different hash value. Comparing a document's current hash value to an earlier value shows whether or not the document has been altered since assigning the earlier value.

Simply creating a hash value does not create legally binding proof of a document's integrity, as a conscious change of a document entails also the hash value's change. As proof of a document's integrity and authenticity additional digital signatures in combination with electronic time-stamps are necessary.

---

[27] In Germany the service can be found at the URL http://nbn-resolving.de/ followed by the URN that is to be resolved, e. g. http://nbn-resolving.de/urn:nbn:de:kobv:11-10079197 for the URN urn:nbn:de:kobv:11-10079197.

[28] See http://persistent-identifier.de/?link=220.

[29] Information on the respective systems can be found at http://www.doi.org/, http://purl.oclc.org/, http://www.handle.net/ and https://confluence.ucop.edu/display/Curation/ARK.

## 4.6 Indexing and Interfaces

### 4.6.1 Indexing

It is crucial that the regulations for the documents' indexing are in a written document and understandable to users of the Document and Publication Service. It is also of importance how the different data were arrived at, e. g. whether the authors or library personnel selected the keywords or wrote the abstracts. In principal it is recommended to use controlled vocabulary for the indexing if the consequent effort is manageable for the Document and Publication Service.

The precise minimal requirements, especially classification according to DDC and the assigning of document and publication types, are essential prerequisites to sort the metadata sets into the standardized sets of the OAI protocol (see sections A.2.2 and A.2.3). This facilitates the selection of certain data by service providers (in the sense of OAI) who cater to a discipline-specific community. Additionally, these standardized sets and nomenclature facilitate e. g. browsing functions.

### 4.6.2 Metadata Interfaces

The OAI interface related requirements contain the most precisely defined aspects of the DINI Certificate. It is of major importance, as all additional, comprehensive services rely on this. To present optimal data to these service providers and to consequently place the local Document and Publication Service in a good position on the global knowledge landscape the mere existence of an OAI interface that is in accordance with the protocol specifications does not suffice. Instead, the OAI guidelines must be adhered to (see Appendix A).

As laid down in the OAI specifications, representation of the metadata in the Dublin Core Simple metadata format is mandatory. It is recommended however, to offer the metadata in other formats as well, e. g. XMetaDissPlus for the German National Library.

## 4.7 Access Statistics

The minimum requirements and recommendations in criterion 7 make statements on the quality of server-wide as well as object-related access information. Especially the visualization for the users requires indications to what extent this information is comparable to other servers' access statistics.

The Open Access Statistics (OA-S) project[30], funded by the Deutsche Forschungsgemeinschaft (DFG, German Science Foundation) created an infra-

[30] For information on this project see http://www.dini.de/projekte/oa-statistik.

structure to collate usage statistics from different servers. In a central service provider the distributed servers' information is processed according to standards (COUNTER, LogEC, or IFABC) and cleared of falsifications, e. g. by spiders or other non-human activities. Document doubles are identified and access to identical documents stored at different locations is joined, and–based on a user de-duplication–double click spans that conform to standards are considered across the distributed servers. The resulting values can be relayed back into the services (repositories) and can be used as document-related metadata. Participation in OA-S as data provider is therefore recommended to Document and Publication Service providers.

Document and Publication Service providers should aim for a long-term availability of the object-related usage information and store it in an open format to facilitate possible future migrations into long-term archiving.

Examples for public, document-related access statistics–displayed on the individual documents' jump-off pages–are the disciplinary repository (http://psydok.sulb.uni-saarland.de/) and the Document and Publication Service of the Technical University Chemnitz MONARCH (http://archiv.tu-chemnitz.de/).

DINI and especially the project partners in OA-S monitor international developments in the area of scientific objects' usage-data analysis, and are trying to reach an international coordination e. g. within the *Knowledge Exchange* program.[31]

## 4.8 Long-Term Availability

### 4.8.1 Minimum Requirements

During a period of no less than five years, regulated in the Document and Publication Service's policy and beginning with a document's publication date, the service provider has to ensure a document's availability independently and in accordance with the minimum requirements described in criterion 5 – *Information Security* in section 2.5. Current recommendations by nestor, the network of competency on digital long-term archiving[32], should be taken into consideration.

---

[31] Knowledge Exchange is a co-operative effort of the funding agencies Danmarks Elektroniske Fag- og Forskningsbibliotek (DEFF, Denmark), the Joint Information Systems Committee (JISC, UK), the SURFfoundation (Netherlands) and the Deutsche Forschungsgemeinschaft (DFG, Germany), that supports the exchange about the use and development of information and communication technologies (ICT) infrastructure for higher education and research, and the coordination of these developments. For information see http://www.knowledge-exchange.info/.

[32] See www.langzeitarchivierung.de for new information on adequate measures to ensure long-term archiving as well as for detailed information and on tools.

Protective measures against copying as employed in Digital Rights Management (DRM) are entirely inadequate for long-term archiving if rendering conversion (migration) and execution in different system environments (emulation) impossible.

### 4.8.2 Recommendations

An institution can offer long-term archiving within the scope of its technical and financial possibilities (depot system according to OAIS). As an alternative administrative and organizational structures can be shaped for an external long-term archiving in cooperation with an archiving institution. The requirements for the transfer of the data and for the necessary metadata for archiving and usage as described in criterion 6 – *Indexing and Interfaces* in section 2.6 must be met.

File formats for the long-term archiving of documents should be open formats. The certificated Document and Publication Service must archive the documents itself or transfer them for this purpose. To ensure long-term availability the original documents or archival copies of it must be stored in an open format. The archived original files delivered by the author do not necessarily correspond to the archival copies. Open formats can be expected to be readable and fully interpretable in the far future, as their source codes are laid open (as DIN/ISO or OASIS) thus permitting everyone for all times to develop software especially for the representation and conversion of these formats. Among these formats are Open Document Format (ODF), ASCII-Text (TXT), Portable Document Format for Long-Term Archiving (PDF/A) and TeX/LaTex (TEX).

Another element of the long-term availability's reliability is the publicly available definition when and under what conditions objects will be deleted.

# Appendix A OAI Interface Guidelines

Appendix A contains the requirements for the OAI interface with regard to the DINI Certificate. Just as the eight main criteria the minimum requirements comprised in this section have to be fulfilled by a Document and Publication Service to be certified (see also criterion 6 – *Indexing and Interfaces*, section 2.6.2, minimum requirement M.6-7.

Since its publication in 2001, the so-called OAI protocol has become the de facto standard for the machine-based and asynchronous exchange of bibliographical metadata between repositories and providers of comprehensive services. In this context, the OAI interface is identified as a functional software component that acts as a data provider in the sense of the protocol, i. e. deliver metadata to requests that are according to protocol. Such an OAI interface is part of the basic components of many repository software solutions[33] and many other systems that administrate metadata[34].

With regard to the requirements that have to be met the OAI protocol offers interoperability at a low level. This has led to a wide dissemination and general acceptance of the protocol in a relatively short time. On the other hand it reduces the service providers' possibilities as the protocol specifications say little about structure and quality of the metadata.

The individual metadata sets must only be made available in the standard format Dublin Core Simple whose specification allows that each of the fifteen metadata elements is optional and may be omitted, but may also be used any number of times. For the elements' inner structures[35] some recommendations exist, but these are not binding. And while the OAI protocol includes a mechanism for the logical separation or structuring of a data provider's data (the so-called *sets*), that permits the selective harvesting, the concrete definition and naming of these sets is up to the data providers' operators.

To build a high-quality service that is based on utilizing data that were harvested using the OAI protocol[36] additional specifications are called for that will fill the gaps (intentionally) left open by the OAI protocol's specifications. The specifica-

---

[33] Examples are OPUS, MyCoRe, ePrints, and DSpace.

[34] Among these are library software, or systems for the realization of electronic journals such as e. g. the Open Journal Systems (OJS).

[35] E. g. the formatting of dates or the coding of languages.

[36] E. g. comprehensive indexing services with search and browsing functions.

tions (see below) refer mostly to a definition of the set structure and the individual metadata element's content in Dublin Core format. Additionally, some requirements are listed that are taken from the protocol's specifications.
Similar to the DINI Certificate's main criteria, the OAI Guidelines list minimum requirements and additional recommendations that the data provider of a Document and Publication Service is not required to fulfill to be DINI-certified. However, these recommendations (marked in each section) mirror current best-practice solutions. They are recommended for application in the OAI interface to optimize the metadata's quality and re-use.
This new attachment to the DINI Certificate is the continuation and replacement of the DINI Recommendations for the Design of the OAI interface[37] that was last updated in 2005. It follows and is compatible to the guidelines[38] developed in the EU project DRIVER[39]. The guidelines like the entire DINI Certificate focus on text-oriented documents and only consider the metadata format Dublin Core Simple (oai_dc).

## A.1 Protocol Conformity

Prerequisite for a functioning data exchange via OAI is a protocol-conform interface, i. e. it complies with the specifications of the OAI Protocol for Metadata Harvesting (OAI PMH) in its current version 2.0[40]. Different ways exist to automatically check existing OAI interfaces' protocol conformity[41]. This verification is done especially if an OAI interface is officially registered as a data provider with the OAI.
The list below emphasizes a few requirements that apply to every OAI interface that meets the protocol specifications, and that require special attention as problems can occur in their implementation.

---

37 See http://nbn-resolving.de/urn:nbn:de:kobv:11-10049220.

38 For version 2 see www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final_2008-11-13.pdf.

39 DRIVER is the acronym for Digital Repository Infrastructure Vision for European Research, see http://www.driver-repository.eu/.

40 For the entire specification see http://www.openarchives.org/OAI/openarchivesprotocol.html.

41 These are among others the Repository Explorer (http://re.cs.uct.ac.za/) or the DRIVER Validator (http://validator.driver.research-infrastructures.eu/). The latter checks not only the conformity with the OAI Guidelines but also with the DRIVER Guidelines.

### Minimum Requirements

| | |
|---|---|
| M.A.1-1 | The OAI interface conforms to the protocol specification version 2.0. |
| | ⇒ All other minimum requirements in this section follow from this. |
| M.A.1-2 | The OAI interface is persistently available under the registered base URL and offers adequate performance. |
| | ⇒ This is a prerequisite for a reliable use of the interface by the service providers, and it ensures the minimization of communication problems, specifically aborted harvesting processes. |
| M.A.1-3 | All replies by the OAI interface are well formed in the XML sense and valid with regard to the XML schema defined in the OAI specification and other XML schemata used for metadata formats. |
| | ⇒ Difficulties arise regularly with the character encoding and special characters within the metadata elements as well as with error messages in the XML stream sent by the database or the application. |
| M.A.1-4 | The OAI interface supports incremental harvesting correctly. |
| | ⇒ Pre-condition for this is that in every record the date of creation or alteration of the metadata is entered in the timestamp element and not e. g. the date of publication of the described document. |
| | ⇒ This allows service providers regular updates of their data without having to harvest all metadata records. For this the data provider must support the parameters *from* and *until* for the OAI requests *ListRecords* and *ListIdentifiers* and deliver the correct subsets of the data with a granularity of at least the day (YYYY-MM-DD). |
| M.A.1-5 | The OAI interface uses set information in a consistent form. |
| | ⇒ This includes especially that all sets that have records assigned to them are delivered upon the *ListSets* request, and that all records that reply to *ListRecords* and *ListIdentifiers* requests qualified by the *set* parameter belong to the respective data set according to their header information. |

**Recommendations**

| | |
|---|---|
| R.A.1-1 | The operator checks the OAI interface in regular intervals with manual tests and validates it with automatic tools. |
| | ⇒ See footnote 41. |
| R.A.1-2 | When making considerable changes to the OAI interface information is given to the registries where the OAI interface or the Document and Publication Service is registered. |
| | ⇒ This allows service providers to react adequately to changes. Relevant alterations in the sense of this recommendation are version changes, change of the base URL, or migrations to new software for the Document and Publication Service. |
| | ⇒ For the relevant registries see criterion 1 – *Visibility of the Service*, section 2.1. |
| R.A.1-3 | The reply to the request *Identify* offers extensive information on the Document and Publication Service. |
| | ⇒ This includes especially an administrator's valid email address in the element *adminEmail* and a short description of the service in the element *description*. |
| R.A.1-4 | The element *provenance* is used in the *about* container for the individual metadata records that are delivered upon the *ListRecords* or the *GetRecord* requests. |
| | ⇒ Additional information on the metadata's sources can be provided in this container. For more information see http://www.openarchives.org/OAI/2.0/guidelines-provenance.htm. |
| R.A.1-5 | The descriptive information in the OAI responses is in English. |
| | ⇒ This includes e. g. the elements in the response to the *Identify* request and the set descriptions with the element *setName* in the response to the *ListSets* request. |

## A.2 OAI PMH: Extended Requirements

The additional requirements described in this section refer mostly to the set structure that the delivered metadata are placed in sections A.2.1 to A.2.4. The structure serves to provide additional standardized information on the documents and to allow selective search queries. This facilitates a better interoperability between Document and Publication Services and the providers of comprehensive services

that are based on them. Further sections contain recommendations on how to deal with deleted documents and records, and on flow control.

### A.2.1 Open Access Document Set

Document and Publication Services not only publish Open Access documents but also documents that are only available e. g. to a user group within an institution. For providers of additional services it is important to discern and select between Open Access and closed access documents and to be able to do so have this identified in the metadata.

**Minimum Requirement**

M.A.2-1 A *setSpec* set exists that states 'open_access' and contains all metadata records of Open Access documents.

⇒ Document and Publication Services that offer only Open Access publications must also meet this requirement. In this case the set contains all metadata records.

### A.2.2 Sets for DDC Groups

To enable a rough disciplinary grouping of metadata sets and the respective documents, in Germany the German National Bibliography's subject groups as used by the German National Library have become the norm. They are based on the Dewey Decimal Classification (DDC) and in principle use its first two items[42]. To allow an external service provider that uses the OAI protocol a pre-selection by subject it is necessary that the subject groups that the Document and Publication Service assigned to the documents are also assigned to the OAI interface's set structure.

**Minimum Requirement**

M.A.2-2 A structure exists in accordance with Table 1, and all metadata records–like the documents–are assigned a *setSpec* according to the table used.

⇒ It is possible to assign each record to more than one DDC class.

[42] See http://www.ddc-deutsch.de/anwendung/dnb.htm.

**Table 1: Name and description of the sets for the subject structure**

| setSpec | setName | German Description |
|---|---|---|
| ddc:000 | Generalities, Science | Allgemeines, Wissenschaft |
| ddc:004 | Data processing Computer science | Informatik |
| ddc:010 | Bibliography | Bibliografien |
| ddc:020 | Library & information sciences | Bibliotheks- und Informationswissenschaft |
| ddc:030 | General encyclopedic works | Enzyklopädien |
| ddc:050 | General serials & their indexes | Zeitschriften, fortlaufende Sammelwerke |
| ddc:060 | General organization & museology | Organisationen, Museumswissenschaft |
| ddc:070 | News media, journalism, publishing | Nachrichtenmedien, Journalismus, Verlagswesen |
| ddc:080 | General collections | Allgemeine Sammelwerke |
| ddc:090 | Manuscripts & rare books | Handschriften, seltene Bücher |
| ddc:100 | Philosophy | Philosophie |
| ddc:130 | Paranormal phenomena | Parapsychologie, Okkultismus |
| ddc:150 | Psychology | Psychologie |
| ddc:200 | Religion | Religion, Religionsphilosophie |
| ddc:220 | Bible | Bibel |
| ddc:230 | Christian theology | Theologie, Christentum |
| ddc:290 | Other & comparative religions | Andere Religionen |
| ddc:300 | Social sciences | Sozialwissenschaften, Soziologie, Anthropologie |
| ddc:310 | General statistics | Statistik |
| ddc:320 | Political science | Politik |
| ddc:330 | Economics | Wirtschaft |
| ddc:333.7 | Natural ressources, energy and environment | Natürliche Ressourcen, Energie, Umwelt |
| ddc:340 | Law | Recht |
| ddc:350 | Public administration | Öffentliche Verwaltung |
| ddc:355 | Military science | Militär |

| setSpec | setName | German Description |
|---|---|---|
| ddc:360 | Social services; association | Soziale Probleme, Sozialdienste, Versicherungen |
| ddc:370 | Education | Erziehung, Schul- und Bildungswesen |
| ddc:380 | Commerce, communications, transport | Handel, Kommunikation, Verkehr |
| ddc:390 | Customs, etiquette, folklore | Bräuche, Etikette, Folklore |
| ddc:400 | Language, Linguistics | Sprache, Linguistik |
| ddc:420 | English | Englisch |
| ddc:430 | Germanic | Deutsch |
| ddc:439 | Other Germanic languages | Andere germanische Sprachen |
| ddc:440 | Romance languages French | Französisch, romanische Sprachen allgemein |
| ddc:450 | Italian, Romanian, Rhaeto-Romantic | Italienisch, Rumänisch, Rätoromanisch |
| ddc:460 | Spanish & Portugese languages | Spanisch, Portugiesisch |
| ddc:470 | Italic Latin | Latein |
| ddc:480 | Hellenic languages Classical Greek | Griechisch |
| ddc:490 | Other languages | Andere Sprachen |
| ddc:500 | Natural sciences & mathematics | Naturwissenschaften |
| ddc:510 | Mathematics | Mathematik |
| ddc:520 | Astronomy & allied sciences | Astronomie, Kartographie |
| ddc:530 | Physics | Physik |
| ddc:540 | Chemistry & allied sciences | Chemie |
| ddc:550 | Earth sciences | Geowissenschaften |
| ddc:560 | Paleontology Paleozoology | Paläontologie |
| ddc:570 | Life sciences | Biowissenschaften, Biologie |
| ddc:580 | Botanical sciences | Pflanzen (Botanik) |
| ddc:590 | Zoological sciences | Tiere (Zoologie) |
| ddc:600 | Technology (Applied sciences) | Technik |
| ddc:610 | Medical sciences Medicine | Medizin, Gesundheit |
| ddc:620 | Engineering & allied operations | Ingenieurwissenschaften und Maschinenbau |

| setSpec | setName | German Description |
|---|---|---|
| ddc:630 | Agriculture | Landwirtschaft, Veterinärmedizin |
| ddc:640 | Home economics & family living | Hauswirtschaft und Familienleben |
| ddc:650 | Management & auxiliary services | Management |
| ddc:660 | Chemical engineering | Technische Chemie |
| ddc:670 | Manufacturing | Industrielle und handwerkliche Fertigung |
| ddc:690 | Buildings | Hausbau, Bauhandwerk |
| ddc:700 | The arts | Künste, Bildende Kunst allgemein |
| ddc:710 | Civic & landscape art | Landschaftsgestaltung, Raumplanung |
| ddc:720 | Architecture | Architektur |
| ddc:730 | Plastic arts Sculpture | Plastik, Numismatik, Keramik, Metallkunst |
| ddc:740 | Drawing & decorative arts | Zeichnung, Kunsthandwerk |
| ddc:741.5 | Comics, Cartoons | Comics, Cartoons, Karikaturen |
| ddc:750 | Painting & paintings | Malerei |
| ddc:760 | Graphic arts Printmaking & prints | Grafische Verfahren, Drucke |
| ddc:770 | Photography & photographs | Fotografie, Computerkunst |
| ddc:780 | Music | Musik |
| ddc:790 | Recreational & performing arts | Freizeitgestaltung, Darstellende Kunst |
| ddc:791 | Public performances | Öffentliche Darbietungen, Film, Rundfunk |
| ddc:792 | Stage presentations | Theater, Tanz |
| ddc:793 | Indoor games & amusements | Spiel |
| ddc:796 | Athletic & outdoor sports & games | Sport |
| ddc:800 | Literature & rhetoric | Literatur, Rhetorik, Literaturwissenschaft |
| ddc:810 | American literature in English | Englische Literatur Amerikas |
| ddc:820 | English & Old English literatures | Englische Literatur |
| ddc:830 | Literatures of Germanic languages | Deutsche Literatur |
| ddc:839 | Other Germanic literatures | Literatur in anderen germanischen Sprachen |

| setSpec | setName | German Description |
|---|---|---|
| ddc:840 | Literatures of Romance languages | Französische Literatur |
| ddc:850 | Italian, Romanian, Rhaeto-Romanic literatures | Italienische, rumänische, rätoromanische Literatur |
| ddc:860 | Spanish & Portuguese literatures | Spanische und portugiesische Literatur |
| ddc:870 | Italic literatures Latin | Lateinische Literatur |
| ddc:880 | Hellenic literatures Classical Greek | Griechische Literatur |
| ddc:890 | Literatures of other languages | Literatur in anderen Sprachen |
| ddc:900 | Geography & history | Geschichte |
| ddc:910 | Geography & travel | Geografie, Reisen |
| ddc:914.3 | Geography & travel Germany | Geografie, Reisen (Deutschland) |
| ddc:920 | Biography, genealogy, insignia | Biografie, Genealogie, Heraldik |
| ddc:930 | History of the ancient world | Alte Geschichte, Archäologie |
| ddc:940 | General history of Europe | Geschichte Europas |
| ddc:943 | General history of Europe Central Europe Germany | Geschichte Deutschlands |
| ddc:950 | General history of Asia Far East | Geschichte Asiens |
| ddc:960 | General history of Africa | Geschichte Afrikas |
| ddc:970 | General history of North America | Geschichte Nordamerikas |
| ddc:980 | General history of South America | Geschichte Südamerikas |
| ddc:990 | General history of other areas | Geschichte der übrigen Welt |

### A.2.3 Document and Publication Type Set

Document type and publication type are a document's important metadata. For a service provider to request certain document types (e. g. dissertations) data providers must provide for a corresponding set structure. Basis of this set structure is the common vocabulary developed for the metadata format XMetaDissPlus and for the DINI Certificate. It is published in the DINI Recommendation *Gemeinsames Vokabular für Publikations- und Dokumenttypen*[43].

[43] See http://nbn-resolving.de/urn:nbn:de:kobv:11-100109998. The heterogeneous use of capital and normal letters in the set names (setSpec) results from the different sources of the vocabulary (among others the Dublin Core Type Vocabulary and Publication Type Vocabulary of the DRIVER Guidelines) and was retained for compatibility reasons.

### Minimum Requirement

M.A.2-3 A structure exists in accordance with Table 2, and all metadata records are assigned a *setSpec* according to the document and publication types.

⇒ As stated in the DINI Recommendation *Gemeinsames Vokabular für Publikations- und Dokumenttypen* assigning a document to more than one document or publication type is recommended (see below Example 1).

**Table 2: Name and description of the sets for the formal structure**

| setSpec | setName | German Description |
|---|---|---|
| doc-type:preprint | Preprint | Preprint |
| doc-type:workingPaper | WorkingPaper | Arbeitspapier |
| doc-type:article | Article | Wissenschaftlicher Artikel |
| doc-type:contributionToPeriodical | ContributionToPeriodical | Beitrag zu einem Periodikum |
| doc-type:PeriodicalPart | PeriodicalPart | Teil eines Periodikums |
| doc-type:Periodical | Periodical | Periodikum |
| doc-type:book | Book | Buch, Monografie |
| doc-type:bookPart | BookPart | Teil eines Buches oder Monografie |
| doc-type:Manuscript | Manuscript | Handschrift oder Manuskript |
| doc-type:StudyThesis | StudyThesis | Studienarbeit |
| doc-type:bachelorThesis | BachelorThesis | Abschlussarbeit (Bachelor) |
| doc-type:masterThesis | MasterThesis | Abschlussarbeit (Master) |
| doc-type:doctoralThesis | DoctoralThesis | Dissertation oder Habilitation |
| doc-type:conferenceObject | ConferenceObject | Konferenzveröffentlichung |
| doc-type:lecture | Lecture | Vorlesung |
| doc-type:review | Review | Rezension |
| doc-type:annotation | Annotation | Entscheidungs- oder Urteilsanmerkung |

| setSpec | setName | German Description |
|---|---|---|
| doc-type:patent | Patent | Patent, Norm, Standard |
| doc-type:report | Report | Verschiedenartige Texte |
| doc-type:MusicalNotation | MusicalNotation | Noten (Musik) |
| doc-type:Sound | Sound | Ton |
| doc-type:Image | Image | Bild |
| doc-type:MovingImage | MovingImage | Bewegte Bilder |
| doc-type:StillImage | StillImage | Einzelbild |
| doc-type:CourseMaterial | CourseMaterial | Lehrmaterial |
| doc-type:Website | Website | Website |
| doc-type:Software | Software | Software, Programme |
| doc-type:CartographicMaterial | CartographicMaterial | Kartographisches Material |
| doc-type:ResearchData | ResearchData | Forschungsdaten |
| doc-type:Other | Other | Verschiedenartige Ressourcen, nicht textgeprägt |
| doc-type:Text | Text | Text |

### A.2.4 Publication Status Set

Document and Publication Services may contain documents at various different stages of a publication process. A correlation may exist between this status and a document's quality. Consequently, a rough identification of a document's status or version is desirable. As in different fields of science different methods of quality evaluation and quality-assurance processes exist, only a very rough structure of evaluation statuses is laid down that includes peer review and other reviewing methods such as the editorial review. The set structure follows the *Version Vocabulary* in the DRIVER Guidelines.

**Recommendation**

R.A.2-1 A set structure exists in accordance with Table 3, and all metadata records are assigned a *setSpec* according to the documents' statuses in the publication process.

**Table 3: Name and description of the sets for the evaluation status**

| setSpec | setName | German Description |
|---|---|---|
| status-type:draft | draft version | Eine frühe Version, die als in Arbeit befindlich in Umlauf gesetzt wurde. |
| status-type:submittedVersion | submitted version | Die Version, die bei einer Zeitschrift eingereicht wurde, um durch Fachleute begutachtet zu werden. |
| status-type:acceptedVersion | accepted version | Die Version, die vom Autor erstellt wurde, in die die Anmerkungen der Gutachter eingeflossen sind und die zur Veröffentlichung angenommen wurde. |
| status-type:publishedVersion | published version | Die Version, die vom Verleger erstellt und veröffentlich wurde. |
| status-type:updatedVersion | updated version | Eine Version, die seit der Veröffentlichung aktualisiert wurde. |

Example 1 shows a possible header of a record provided through the OAI PMH that meets the above listed requirements. The record belonging to this header describes a published Open Access scientific article in mathematics.

```
<header>
 <identifier>oai:MyRepository.de:423569</identifier>
 <datestamp>2010-05-15T12:45:01Z</datestamp>
 <setSpec>open_access</setSpec>
 <setSpec>doc-type:article</setSpec>
 <setSpec>doc-type:Text</setSpec>
 <setSpec>ddc:510</setSpec>
 <setSpec>status-type:publishedVersion</setSpec>
</header>
```

*Example 1: Possible set information in the header as given in response to* ListRecords, GetRecords *or* ListIdentifiers *requests.*

### A.2.5 Deleted Documents

In principle, documents that are published by a Document and Publication Service are not to be deleted. However, reasons may exist that permit a document's deletion in certain cases (see Criterion 5 – *Information Security* in section 2.5). The incremental harvesting by service providers may not reveal the information about deleted documents–and deleted metadata records–to OAI-based service providers. The OAI protocol's specifications do not lay down which information a data provider has to provide for deleted documents, but offer a number of options that every data provider can define as *Deleting Strategy* and must transmit with the replies to OAI *Identify* requests.

**Minimum Requirement**

M.A.2-4 One of the values 'persistent' or 'transient' is selected as *Deleting Strategy* for the data provider.

⇒ The OAI PMH permits the options 'no', 'persistent' and 'transient'. If 'no' is selected, no information on deleted documents is transmitted, which can lead to inconsistent data on the service provider's side.

⇒ If the option 'transient' is used for deleted documents the corresponding metadata records have to be available for at least one month after deletion indicating that the document has been deleted.

### A.2.6 Data-Flow Control

To avoid having to deliver large data amounts as replies to OAI requests the OAI protocol offers a data flow control. The data provider can define a so-called *Harvest Batch Size*, i. e. the maximum number of metadata records to be delivered in one batch to *ListRecords* or *ListIdentifiers* requests. If the number of hits is greater than the number defined, a *Resumption Token* is transmitted with the reply, which permits the continuation of the delivery. The protocol specifications leave it to the data provider what size of packages to deliver, for how long to continue a delivery, or whether to use this option at all.

**Recommendations**

R.A.2-2 The harvest batch size (i. e. the maximum number of data sets in reply to a *ListRecords* OAI request) is no less than 100 and no more than 500.

⇒ Smaller data packages lead to more requests and increase communication duration and the risk of errors.

R.A.2-3 The resumption token's life span is at least 24 hours.

⇒ The attribute *lifeSpan* describes the time in which the data provider guarantees the continuation of incomplete replies. If this time span is too short it can cause the cancellation of the entire harvesting process as it expires before the previous reply has been delivered completely.

R.A.2-4 The attribute *completeListSize* is used.

⇒ This describes the entire result list's size that can be important information for the steering and controlling of the harvesting process. According to the OAI protocol however, it is optional.

## A.3 Metadata Requirements (Dublin Core Simple)

The OAI protocol defines the minimum standard that the metadata be in the Dublin Core Simple format. However, no specifications are given for the precise usage of the individual elements and their inner structures. The following requirements and recommendations on the use of Dublin Core for the OAI interface serve to secure a minimum of interoperability on metadata level.

**Minimum Requirements**

M.A.3-1 The Dublin Core formatted metadata sets (oai_dc) contain at least the elements **creator**, **title**, **date**, **type** and **identifier**.

⇒ The elements are necessary for a minimal description of electronic academic documents.

M.A.3-2 In every used DC element exactly one value is referenced.

⇒ Every DC element can be used multiple times within one metadata set.

⇒ Every author's name should be listed in a single *creator* element, every keyword in one single *subject* element, every URL in a single *identifier* element, etc.

| | |
|---|---|
| M.A.3-3 | Every record contains at least one **identifier** element with an operable URL based on a Persistent Identifier, which links to the document's full text. |
| | ⇒ To transform a Persistent Identifier (e. g. URN or DOI) into a working URL the resolving service's base URL must precede it (see criterion 5 – *Information Security*, section 2.5.2, minimum requirements M.5-10 and M.5-11). |
| | ⇒ Additional *identifier* elements may contain differing URLs to a document's jump-off page or to alternative versions (e. g. in a different file format) or they may contain different identifiers (e. g. ISBN, DOI). |
| M.A.3-4 | The **creator** element has the inner structure: *last name, first name*. |
| | ⇒ The same is true for the *contributor* element when it contains a personal name. |
| M.A.3-5 | Document or publication types according to the DINI Recommendations *Common Vocabulary for Publication and Document Types* (*Gemeinsames Vokabular für Publikations- und Dokumenttypen*) are assigned to all documents. |
| | ⇒ The DINI Recommendation supports the listing of a value from the Dublin Core Type Vocabulary in a *type* element of its own. |
| | ⇒ For the vocabulary see the first column in table 2 (above). |
| M.A.3-6 | Every record contains at least one DNB subject group in a **subject** element, and the document is listed in that group. |
| | ⇒ For the vocabulary see the first column in table 1 (above). |
| M.A.3-7 | The **language** element's content is listed according to ISO 639-3. |
| | ⇒ For German the code is "deu", for English it is "eng". |
| M.A.3-8 | The **date** element's content is listed according to ISO 8601. |
| | ⇒ The corresponding format is YYYY-MM-DD. |

### Recommendations

| | |
|---|---|
| R.A.3-1 | The **identifier** elements' order in a metadata record mirrors their importance. The preferred value is given first. |

| | |
|---|---|
| | ⇒ Many service providers read the position as a marker for the priority given to a URL. From the Document and Publication Service provider's perspective the link to the jump-off page is usually the preferred one. |
| | ⇒ Formally, the elements' order is of no importance in Dublin Core, but adhering to the rule above has proven to be practicable to "recommend" the preferred URL to the service provider. |
| R.A.3-2 | The **contributor** element is used and contains the name of one person or institution that was involved in the creation of the document described. |
| | ⇒ This may be the referee of a dissertation or the editor of a collection. |
| R.A.3-3 | The **source** element follows the *Guidelines for Encoding Bibliographic Citation Information* in Dublin Core metadata. |
| | ⇒ The element is used to name a source of the electronic version (citation); see http://dublincore.org/documents/dc-citation-guidelines/. |
| R.A.3-4 | The **relation** element is used to name objects that are related to the document described. |
| | ⇒ Relations may be hierarchical structures (*isPartOf*) or updates (*isVersionOf*). |
| R.A.3-5 | The **subject** element is used for descriptions of a document's content. |
| | ⇒ In general, the content is described using keywords, or notations from classification schemas. |
| R.A.3-6 | The **date** element is used only once in a metadata record. |
| | ⇒ The publication date is to be preferred over other dates (e. g. upload date or date of creation), as it has the greatest priority for the reader. |

## Appendix B Glossary

In this section the most important terms used in this document are named and defined.

**Data Provider**. Document server in the OAI protocol's understanding. Offers documents' metadata via the OAI interface.

**Deposit License**. Formal agreement in which the rights holder (i. e. the author or the publisher) grants certain usage rights to the provider of a Document and Publication Service in order to allow the provider to make the respective documents publicly available. Moreover, in this agreement the rights holder excludes that any third party's rights may be violated and that the provider could be made liable.

**Document**. Smallest logical entity that is published by a Document and Publication Service, usually a scientific or scholarly work with clearly named creators. Synonyms used in this text: *electronic document*, *publication*, *work*. For data, images, etc. the term *object* should be used.

**Document and Publication Service**. Comprehensive service for the publication and online provision of scientific and scholarly publications. The service caters to producers (authors) as well as to recipients (readers) and contains both the technical infrastructure (i. e. the *document server*) and the organizational and legal frame.

**Document Server**. Document and Publication Service's technical infrastructure, characterized by basic infrastructure components (e. g. network, server, operating system, databases, communication systems) and the document server software (e. g. DSpace, ePrints, MyCore, OPUS). Synonyms used in this text: *publication server*, *repository*.

**Jump-off Page**. Web page containing metadata of and links to a document's full-text files plus additional functions and information (e. g. social network links, export of bibliographical data in machine-readable formats, print on demand services, document-related statistics). Usually the jump-off page is generated dynamically, its content coming from a database. Synonyms: *splash page*, *front page*, *front door*.

**Metadata**. Data for the characterization of an object (in this text mostly documents). Typically, these are divided into descriptive, technical and administrative metadata. Descriptive metadata contain information for the formal and subject classification. Metadata can be coded in different formats and are interchangeable. It is possible that internally stored metadata are not completely made available to the public (example: administrative metadata). Relevant standards for electronic publications are Dublin Core, MARC, MODS as well as especially for the data exchange with the German National Library XMetaDissPlus.

**Open Access**. Worldwide free access to scientific information, especially to scientific and scholarly publications in electronic form and online, as defined e. g. in the 2003 *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*[44]. A worldwide movement with numerous national and international initiatives is dedicated to the dissemination and to the achievement of the goals of the Berlin Declaration. Typically, two forms of Open Access are differentiated: The green and the golden roads. The first describes the additional publication of documents already published elsewhere (usually by a publishing house) or slotted for publication as a parallel publication in a freely available version–usually in a repository. The golden way is the primary publication with Open Access, e. g. in an Open Access journal.

**Persistent Identifier**. Worldwide unambiguous and unchangeable (persistent) name for a digital information object, (for this text) usually an electronic document. Persistent identifiers (PI) are especially useful for the citation of electronic publications, as they are unlike a URL permanent. Different PI systems exist, e. g. URN, DOI, and PURL. A PI's syntactical structure is defined in a formal description of the structure. PIs and related URLs must be registered at a (usual) central point to facilitate the resolving service that reroutes request for a URN to the actual physical addresses.

**Provider**. Institution that is responsible for the provision of (in this text) a Document and Publication Service. It offers the service to various user groups and answers to the users even if responsibilities are divided internally or even sourced out.

**Service Provider**. A service provider offers comprehensive services using distributed data that are aggregated via the OAI protocol. Synonyms used here: *Harvester*.

---

[44] Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003, see http://oa.mpg.de/files/2010/04/berlin_declaration.pdf.

**User**. Natural person who uses services offered by a Document and Publication Service, especially as producer (authors, publishers) or recipient (reader, researcher) of documents.

**Usage Rights**. In this document's context, these are rights that are granted to users of documents or their metadata that are published by Document and Publication Service. Originally these usage rights are held by the creators and consequently must be transferred with applicable processes.

## Appendix C Authors



The following persons are authors of the current version of this document:

**Kim Braun.** Bibliotheks- und Informationssystem der Carl von Ossietzky Universität Oldenburg, kim.braun@uni-oldenburg.de.

**Stefan Buddenbohm.** Niedersächsische Staats- und Universitätsbibliothek Göttingen, buddenbohm@sub.uni-goettingen.de.

**Susanne Dobratz.** Humboldt-Universität zu Berlin, Universitätsbibliothek, Arbeitsgruppe Elektronisches Publizieren, dobratz@cms.hu-berlin.de.

**Ulrich Herb.** Saarländische Universitäts- und Landesbibliothek, Saarbrücken, u.herb@sulb.uni-saarland.de.

**Uwe Müller.** Humboldt-Universität zu Berlin, Institut für Bibliotheks- und Informationswissenschaft, u.mueller@ibi.hu-berlin.de.

**Heinz Pampel.** Helmholtz-Gemeinschaft, Potsdam, pampel@gfz-potsdam.de.

**Birgit Schmidt.** Niedersächsische Staats- und Universitätsbibliothek Göttingen, bschmidt@sub.uni-goettingen.de.

**Frank Scholze.** Karlsruher Institut für Technologie, KIT-Bibliothek, frank.scholze@kit.edu.

**Matthias Schulze.** Universität Stuttgart, Universitätsbibliothek, matthias.schulze@ub.uni-stuttgart.de.

**Thomas Severiens.** Universität Osnabrück, Fachbereich Mathematik/ Informatik, tseverie@mathematik.uni-osnabrueck.de.

**Friedrich Summann.** Universität Bielefeld, Universitätsbibliothek, friedrich.summann@uni-bielefeld.de.

**Tobias Steinke.** Deutsche Nationalbibliothek, Frankfurt, t.steinke@dnb.de.

**Stefan Wolf.** Bibliotheksservice-Zentrum Baden-Württemberg, Konstanz, stefan.wolf@bsz-bw.de.

Additional authors of earlier versions:
Margo Bargheer.
Stefan Gradmann.
Eberhard Hilf.
Wolfram Horstmann.
Elmar Mittler.
Peter Schirmbacher.
Silke Schomburg.
Heinrich Stamerjohanns.
Bert Wendland.
Christoph Ziegler.

Special thanks for support and cooperation goes to the following:
Maren Brodersen. Deutsche Nationalbibliothek, Frankfurt, m.brodersen@dnb.de.
Nina Gerlach. Niedersächsische Staats- und Universitätsbibliothek Göttingen, gerlach@sub.uni-goettingen.de.
Sarah Hartmann. Deutsche Nationalbibliothek, Frankfurt, s.hartmann@dnb.de.
Maxi Kindling. Humboldt-Universität zu Berlin, Computer- und Medienservice, Arbeitsgruppe Elektronisches Publizieren, maxi.kindling@ibi.hu-berlin.de.
Nikola Korb. Universitätsbibliothek Siegen, korb@ub.uni-siegen.de.
Berthold Weiß. Hochschul- und Landesbibliothek Fulda, berthold.weiss@hlb.hs-fulda.de.

**Imprint**
DINI – Deutsche Initiative für Netzwerkinformation e. V.

DINI-Geschäftsstelle
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen
Papendiek 14
37073 Göttingen
Tel.: +49 551 39-3857
Fax: +49 551 39-3856
E-Mail: gs@dini.de
www.dini.de