



## **Implementation of the Data Seal of Approval**

The Data Seal of Approval board hereby confirms that the Trusted Digital repository SLUBArchiv complies with the guidelines version 2014-2017 set by the Data Seal of Approval Board.

The afore-mentioned repository has therefore acquired the Data Seal of Approval of 2013 on June 27, 2015.

The Trusted Digital repository is allowed to place an image of the Data Seal of Approval logo corresponding to the guidelines version date on their website. This image must link to this file which is hosted on the Data Seal of Approval website.

Yours sincerely,

The Data Seal of Approval Board

## Assessment Information

Guidelines Version:	2014-2017   July 19, 2013
Guidelines Information Booklet:	<a href="#">DSA-booklet_2014-2017.pdf</a>
All Guidelines Documentation:	<a href="#">Documentation</a>
Repository:	SLUBArchiv
Seal Acquiry Date:	Jun. 27, 2015
For the latest version of the awarded DSA for this repository please visit our website:	<a href="http://assessment.datasealofapproval.org/seals/">http://assessment.datasealofapproval.org/seals/</a>
Previously Acquired Seals:	None
This repository is owned by:	<b>Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB)</b> •
	T +493514677240 E <a href="mailto:henrike.berthold@slub-dresden.de">henrike.berthold@slub-dresden.de</a> W <a href="http://www.slub-dresden.de/startseite/">http://www.slub-dresden.de/startseite/</a>

# Assessment

## 0. Repository Context

### Applicant Entry

*Self-assessment statement:*

The Saxon State and University Library Dresden (Sächsische Landesbibliothek - Staats- und Universitätsbibliothek Dresden, SLUB, <http://www.slub-dresden.de/en/home/>) is the archive library of the Free State of Saxony/Germany (<http://www.sachsen.de/en/index.html>) and the university library of Technische Universität Dresden (TUD, <http://tu-dresden.de/en>). The long-term preservation of digital research publications of the TUD and digital documents related to Saxony belongs to its statutory mandate.

SLUBArchiv is the digital long-term preservation system of SLUB. It uses the extendable preservation software Rosetta (<http://www.exlibrisgroup.com/category/RosettaOverview>) which has been customized to SLUB's needs. Rosetta is specifically configured for each application-specific preservation workflow, and it is complemented by an application-specific pre-ingest processing and post-access processing to fit the application-specific workflow. SLUBArchiv covers the five OAIS functional entities: ingest, data management, archival storage, administration and access. It is designed as a pure archive with access only for SLUB staff members. External end users are not granted direct access for data security reasons. Instead, they are entitled to use SLUB's presentation systems.

SLUBArchiv currently implements a productive Goobi preservation workflow and interface to the SLUB digitization workflow, which is a process of digitizing print works and other media. SLUB uses the workflow software Goobi.Production (<http://www.goobi.org/en/>) to support this digitization process. Once the SLUB digitization workflow is finished for a document, the resulting data are exported to Goobi.Presentation and the Goobi preservation workflow. External end users/consumers get access to the digitized documents via Goobi.Presentation (see SLUB's digital collections at <http://digital.slub-dresden.de/en/digital-collections/>). The Goobi preservation workflow of the SLUBArchiv is the subject of this certification.

The computing infrastructure, on which SLUBArchiv is based, is integrated in the Center for Information Services and High Performance Computing (Zentrum für Informationsdienste und Hochleistungsrechnen, ZIH) of the Technische Universität Dresden (TUD).TUD/ZIH is the only partner involved in SLUBArchiv. Until the end of 2015, the computing infrastructure will be managed in cooperatively between SLUB and TUD/ZIH. In 2015, a cooperation contract with Service Level Agreements between SLUB and TUD will be drafted and signed. From 2016 on, the computing infrastructure will be managed by the TUD/ZIH as an outsource partner.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

**1. The data producer deposits the data in a data repository with sufficient information for others to assess the quality of the data, and compliance with disciplinary and ethical norms.**

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

**Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

In the SLUB digitization workflow, works which have already been published and for which digitization is legal, are digitized. Quality, legal and ethical standards are constantly being checked in the process of publishing documents. The majority of this material belongs to SLUB print collections and is digitized by SLUB's own digitization center. To complete its digital collections, SLUB also digitizes selected material of other institutions. In these cases, it usually gets the original material and digitizes it. In few cases, when the material has already been digitized by the other institution, it gets digital objects that are then further processed by SLUB staff in the digitization workflow.

In the first step of the SLUB digitization workflow, a record that represents a digital object is created. The scans of the corresponding document and checksums are then added to the record. The descriptive metadata of the original document are taken from the local or remote library catalogue. Further descriptive and structural metadata are added and the whole record is exported to Goobi.Presentation and the Goobi preservation workflow.

Data that belong to a digitized object including checksums are packed into an OAIS Submission Information Package (SIP) in the ingest pre-processing of the Goobi preservation workflow. During ingest processing, checksums are verified multiple times.

In the SLUB digitization workflow, the presentation data of a digital object are inserted into Goobi.Presentation and a subset of its metadata is added to the library's catalogue and in many cases to additional portals such as Europeana and the German Digital Library (Deutsche Digitale Bibliothek, DDB). Examples of catalogue entries and data are here:

- <http://digital.slub-dresden.de/id336230591>

- <http://digital.slub-dresden.de/id351327312>

- <http://digital.slub-dresden.de/id340163720>

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 2. The data producer provides the data in formats recommended by the data repository.

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The target format of the digitization is baseline TIFF. The digitization is compliant with the guidelines specified by the Deutsche Forschungsgemeinschaft (DFG; the German Research Council; Practical Guidelines on Digitization [http://www.dfg.de/formulare/12\\_151/12\\_151\\_en.PDF](http://www.dfg.de/formulare/12_151/12_151_en.PDF)). These guidelines cover technical parameters such as the resolution, color depth, the capturing process and file formats, material-specific parameter and metadata.

SLUB scanners are configured in compliance with the DFG guidelines. Their configurations were checked beforehand to make sure that the DFG rules are met. Parameters were adjusted according to the specifications of Baseline TIFF (<http://partners.adobe.com/public/developer/en/tiff/TIFF6.PDF>). In some rare cases (e.g. collection of maps or on special request), the DGF guideline for scan resolution is exceeded to meet internal customers' needs.

Baseline TIFF is on the list of preferred formats published by the SLUBArchiv (<http://www.slub-dresden.de/ueber-uns/slubarchiv/technische-standards-fuer-die-ablieferung-von-digitalen-dokumenten/>, in German). The SLUB Digitization Center is aware of this list and adheres to it. The list of preferred formats will change over time as new archive workflows with other data formats are added, new formats develop and others become obsolete.

A digitized document undergoes intellectual checks in the SLUB digitization workflow. Beside the image data, metadata are created in the METS/MODS XML format (<http://www.loc.gov/standards/mods/v3/mods-3-5.xsd>, <http://www.loc.gov/standards/mets/version191/mets.xsd>) and OCR data are created in the ALTO XML (<http://www.loc.gov/standards/alto/>) format. The compliance with the official specifications of both formats is verified.

During ingest pre-processing, the so-called submission application validates the METS/MODS XML file against the official METS/MODS schema. During ingest processing, the digital preservation software Rosetta runs a file format identification and validation against all ingested files to check the files' compliance with the format specifications. If Rosetta detects a mismatch or an error, the software rejects the document. Depending on the reported problem, a member of the preservation staff or the digitization staff corrects it. Since the digitization process is an in-house process, all data can be corrected. In the worst case, the document has to be produced anew.

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

### **3. The data producer provides the data together with the metadata requested by the data repository.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

#### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

SLUB itself is the data producer. In the SLUB digitization workflow, the descriptive metadata of the original document are taken from the local or remote library catalogue and, if necessary, supplemented. Administrative metadata such as the intellectual property are added manually. Basic structural metadata are added automatically. For some digital documents, an additional intellectual enrichment takes place in which the document's structure is specified in detail, e.g. table of contents, sections. The workflow software Goobi.Production provides forms to enter the data.

Administrative metadata needed for preservation purposes, e.g. the Pronom-ID of the data format, are automatically derived in the Goobi preservation workflow.

Metadata are recorded on file and object level. The software Goobi.Production stores the metadata in the METS/MODS format (METS: <http://www.loc.gov/standards/mets/>, MODS: <http://www.loc.gov/standards/mods/>). The software Rosetta in its current version (4.1) supports METS/DC (Dublin Core, DC, <http://dublincore.org/documents/dcmi-terms/>). The submission application transforms Goobi METS/MODS data to Rosetta METS/DC data. The original METS/MODS metadata file is added to the SIP because there is no bi-directional, complete mapping between METS/MODS and METS/DC.

Rosetta's ingest processing can be tailored to the needs of a workflow. The ingest processing of Goobi archiving has been configured such that it checks if mandatory DC elements are provided. It furthermore validates METS/DC XML files and METS/MODS XML files against their schemas.

If Rosetta detects incomplete metadata, the software rejects the document. Depending on the reported problem, it is then corrected by a member of the preservation staff or by the digitization staff. Since the digitization process is an in-house process, all data can be corrected.

#### **Reviewer Entry**

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)



*Accept or send back to applicant for modification:*

Accept

*Comments:*

#### **4. The data repository has an explicit mission in the area of digital archiving and promulgates it.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

#### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

SLUB is the archive library of the Free State of Saxony/Germany and the university library of Technische Universität Dresden (see Guideline 0). The long-term preservation of digital research publications of the TUD and digital documents related to Saxony (as formulated in §2 of the law "Gesetz über die Sächsische Landesbibliothek - Staats- und Universitätsbibliothek Dresden",

[http://www.slub-dresden.de/fileadmin/groups/slubsite/Ueber\\_uns/Organisation/Gesetz\\_%C3%BCber\\_die\\_SLUB\\_Dresden-Fassung\\_vom](http://www.slub-dresden.de/fileadmin/groups/slubsite/Ueber_uns/Organisation/Gesetz_%C3%BCber_die_SLUB_Dresden-Fassung_vom) in German) belongs to its statutory mandate.

Furthermore, SLUB has performed and performs digitization projects funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Council) in which it digitizes selected sets of its print material. In these projects, SLUB makes the commitment to digitally preserve the produced digital documents (see [http://www.dfg.de/formulare/12\\_151/12\\_151\\_en.pdf](http://www.dfg.de/formulare/12_151/12_151_en.pdf)).

The mission statement is published (see <http://www.slub-dresden.de/ueber-uns/slubarchiv/>, in German).

SLUBArchiv is based on the software Rosetta. SLUB assumes that the lifetime of the software is shorter than the lifetime of the archival data. There is no published succession plan, but SLUBArchiv staff actively tracks the developments in the area of digital preservation systems and has implemented an exit strategy which allows data migration to a new preservation software system. The exit strategy is based on a software program that has been made publically available (<https://github.com/SLUB-digitalpreservation/rosettaExitStrategy>). It is tested when a new major releases of the software Rosetta is installed and adapted, if the test is not successful.

#### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

## **5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

SLUB itself is the producer in the Goobi preservation workflow because it digitizes mainly print material of its own collections. The majority of printed works that are digitized is no longer subject of copyright regulations. Since 2015, SLUB publishes all digitized documents for which this is responsible under the license CC-BY-SA 4.0 (see Achim Bonte, Simone Georgi: Größtmögliche Offenheit, in BIS – Das Magazin der Bibliotheken in Sachsen, Nr. 1, 2015, in German, [http://www.qucosa.de/recherche/frontdoor/?tx\\_slubopus4frontend\[id\]=16409](http://www.qucosa.de/recherche/frontdoor/?tx_slubopus4frontend[id]=16409)). Works that are subject to copyright regulations or are located in another institution are only digitized and published either if there is a contract between the author/institution and SLUB or if the author/institution has transferred the publications rights to SLUB. In these cases, a more restricted open license might hold.

SLUBArchiv is designed to be a closed archive with no direct access for consumers (see guideline 0). Digitized material is made accessible via Goobi.Presentation. As of now, the license of published digital documents is shown in a box next to the presentation of each collection (a list of SLUB's collections is here <http://digital.slub-dresden.de/en/digital-collections/>) or single documents. The general terms of use are published at <http://digital.slub-dresden.de/en/terms-of-use-for-digital-collections/>.

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 6. The data repository applies documented processes and procedures for managing data storage.

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

Data storage is currently managed cooperatively by SLUB and ZIH/TUD (see guideline 0). It is physically located at ZIH/TUD and integrated in its computing center. The computing center's standard processes for monitoring and handling of tape and disk-based storage systems apply.

SLUBArchiv has a policy for managing data storage and bit-stream preservation (see <http://www.slub-dresden.de/ueber-uns/slubarchiv/erhalt-der-korrektheit/>, in German). Each AIP is stored in three or four physical copies in two different locations, both of which are equipped with disk and tape storage systems. The number of copies depends on the preservation class as specified by the library staff.

Each AIP is stored in two storage pools - a primary and a secondary storage pool - of a clustered file system (IBM General Parallel File System, GPFS). The two storage pools are located at different locations of the computing center. In these storage pools, large files are migrated to tape with the Hierarchical Storage Management (HSM), which is an extension of the IBM Tivoli Storage Manager software (TSM). The third copy of an AIP is a backup copy. TSM is used as backup software. Backup copies of new data in the GPFS-based permanent storage are made regularly (currently three times in 24 hours). They are daily written to tape. Depending on the preservation status, the fourth copy is a second backup copy, which is physically stored at the second location. All tape pools (i.e. HSM and backup tape pools) are protected by Logical Block Protection (LBP, a CRC checksum technology). All actions are taken according to the official TSM Best Practices

([https://www.ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/Tivoli%20Storage%20Manager/page/Best%20Practices%](https://www.ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/Tivoli%20Storage%20Manager/page/Best%20Practices%20)

The recovery policy is as follows. The primary storage pool is the primary access medium. If there are failures in the primary storage pool, the secondary copy pool is used. Between these two storage pools, automatic synchronization and recovery is configured. If there are failures on both storage pools, the data are restored from the backup.

The integrity of archival copies is checked according to the policy described on the SLUB website (<http://www.slub-dresden.de/ueber-uns/slubarchiv/erhalt-der-korrektheit/>, in German). We use two different approaches in parallel.

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

1. Integrity of archival copies is checked yearly for a 1%-sample of AIPs. If an error is detected, corrupt files are replaced by correct copies and AIPs that are produced or updated the same day are checked. Due to the applied storage policy, AIPs that are ingested the same day are located in the same storage area. Depending on the results, the check is extended to a longer time period in which AIPs are stored in permanent storage.
2. A file with a specified bit pattern is daily stored. All those files are quarterly checked. Due to the specified bit pattern, single and multiple bit failures can be detected. If an error is identified, the data that are produced the same day are checked. Depending on the results, the check is extended to a longer time period in which AIPs are stored in permanent storage.

SLUBArchiv can only be accessed by authorized library and administration staff.

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 7. The data repository has a plan for long-term preservation of its digital assets.

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The processes applied to data storage described in guideline 6 above ensure bit-stream preservation. Functionalities in the digital preservation software Rosetta support preservation planning and action. During ingest processing, data formats are identified and validated, technical metadata are extracted and added to the stored metadata. Rosetta's ingest processing is extensible. Plugins with specific functionalities such as validation of a certain data format can be integrated into the processing. Rosetta's Format Library is based on the global format registry PRONOM (<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>). The Format Library Working Group, a working group of the Rosetta User Group, maintains the Format Library i.e. it extends it for example by additional formats that are needed by the Rosetta community and integrates new releases of PRONOM. The Format Library is updated with new Rosetta releases (approximately twice a year). Additionally, local risks can be specified. A risk analysis process runs regularly (once a week) and analyses metadata and risks. The risks are visualized on the dashboard and specific reports.

SLUBArchiv maintains a list of preferred data formats (see <http://www.slub-dresden.de/ueber-uns/slubarchiv/technische-standards-fuer-die-ablieferung-von-digitalen-dokumenten/>, in German and guideline 2). The data format baseline TIFF, which is used in the SLUB digitization and Goobi preservation workflows, is listed there. The list of preferred formats will change over time as new archive workflows with other data formats are added, new formats develop and others become obsolete.

In order to test Rosetta's preservation planning and action functionality and document its use, a preservation action for a format migration from PDF 1.4 (fmt/18) to PDF/A (fmt/354) has been planned and executed.

The process of preservation planning and actions is documented publicly (<http://www.slub-dresden.de/ueber-uns/slubarchiv/erhalt-der-interpretierbarkeit/>, in German).

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

## 8. Archiving takes place according to explicit work flows across the data life cycle.

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The SLUB Digitization Workflow and the Goobi preservation workflow are established and documented workflows (<http://www.slub-dresden.de/ueber-uns/slubarchiv/slub-workflows/goobi-workflow/>; in German).

In this workflow, only material with an archival value is selected by the library staff. Therefore, all data objects produced in this process are digitally long-term preserved. The process has been extended towards different types of material, e.g. newspapers, and special requirements for specific collections, e.g. additional OCR processing.

Library staff that is involved in this process, including the staff working in the area of digital long-term preservation, is trained accordingly (see <http://www.slub-dresden.de/ueber-uns/slubarchiv/organisatorische-und-personelle-einbindung/>, in German).

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## **9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

SLUB assumes permanent legal responsibility for preservation of digitized objects of its own material. For digitization projects funded by the DFG (the German Research Council, see guideline 2) which cover selected print collections and documents, SLUB is contractually obligated to preserve the resulting digital collections and documents. With external partners, contractual agreements about preservation and presentation have been signed.

As mentioned in guideline 0, SLUBArchiv is a pure archive, which can be accessed only by SLUB staff. They access archived digital objects via automated workflows. To ensure availability of the archive, SLUB has implemented a loose coupling between SLUBArchiv and access systems by setting up shared storage between the systems as the only shared resource. This way, even if one of the systems encounters an outage, all other systems can keep producing and processing data unaffectedly.

In case of a catastrophic event, the SLUBArchiv has taken actions to ensure a quick system recovery (as described in guideline 6). Alternatively, an exit strategy has been implemented and tested in preparation for scenarios where all that is left are the file systems and/or backups. The same software can also be used for migration scenarios from Rosetta to another digital preservation software solution.

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## **10. The data repository enables the users to discover and use the data and refer to them in a persistent way.**

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The digital documents that are long-term preserved in the SLUBArchiv are made available to end users via Goobi.Presentation (see SLUB's digital collections at <http://digital.slub-dresden.de/en/digital-collections/>). These documents are presented online in the JPEG format. They can be downloaded in the PDF format. Both data formats are widely used in the designated communities (i.e. the research community and the general public). The digital documents can be searched and browsed via the library catalogue, which has an integrated semantic search. Each digitized document has a unique persistent identifier, a Uniform Resource Name (URN). URNs are requested from and registered by the German National Library (DNB, Deutsche Nationalbibliothek, <http://www.dnb.de/EN/>) whenever a producer generates a new digital object. The URN becomes a part of the digital object's metadata and is ingested into the archival system together with the actual data, where it can be searched. Moreover, each object is assigned a Persistent Uniform Resource Locator (PURL, <https://purl.org/docs/index.html>). All objects can also be harvested via OAI interface ([http://www.slub-dresden.de/sammlungen/digitale-sammlungen/oai/?verb=ListRecords&metadataPrefix=oai\\_dc](http://www.slub-dresden.de/sammlungen/digitale-sammlungen/oai/?verb=ListRecords&metadataPrefix=oai_dc)).

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 11. The data repository ensures the integrity of the digital objects and the metadata.

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

A MD5 checksum (Message-Digest Algorithm; standard document <http://www.ietf.org/rfc/rfc1321.txt>) is calculated automatically for each file in the SLUB digitization workflow. In the pre-ingest phase of the Goobi preservation workflow, the submission application verifies these checksums and adds them to the METS metadata. It additionally generates SHA1 (US Secure Hash Algorithm 1, standard document <http://tools.ietf.org/html/rfc3174>) and CRC32 (Cyclic redundancy check) checksums, and adds them to the metadata. During ingest processing, Rosetta validates these checksums as well (see Rosetta Staff User's Guide, section "Fixity in Rosetta with an External Storage Layer"). The METS/MODS metadata file of a digital object is stored along with all other master files, so checksums are stored in Rosetta's data management system and in the archival storage.

While checksums are held on the file level, the tape storage system maintains its own checksums using a technology called Logical Block Protection (LBP) that adds Cyclic Redundancy Check (CRC) checksums to every block. It runs fixity checks every time data is read from or written to storage tapes.

The software Rosetta manages multiple versions. Each time an AIP is changed, a new version is created. Reasons for changes are (1) preservation actions, (2) corrections initiated by the production, e.g. a single page is scanned anew, or (3) additions, e.g. of an OCR processing result. Older versions of digital objects remain stored and accessible for staff users. Checksums are calculated and validated for new files as described above.

The processes for checking the AIPs' integrity are described in the answer to guideline 6.

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

**Data Seal of Approval Board**

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)

## 12. The data repository ensures the authenticity of the digital objects and the metadata.

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

Basic provenance metadata of the original data are already filed in the catalogue. In the digitization workflow, these data are copied to the metadata of the digitized object. The authenticity of a digitized object is checked by library staff. In most cases, they compare pages of the digitized object with the original. In a few cases, when they only receive digital objects of third-party institutions, they compare the digital data with the metadata contained in the catalogue. The effort needed for these comparisons depends on the project and the complexity of the original. Each digitized object gets a unique identifier (a URN, see guideline 10).

Usually, the depositor of the digitized objects is the SLUB itself. Contact persons to third party institutions that deposit data are well known. No identity check is necessary.

The strategy of data changes and the use of checksums are described in the answer to guideline 11. All changes to an AIP are documented in the metadata (i.e. the audit trail). During preservation planning activities, the significant properties of different versions of the same file will be compared (see list of significant properties at <http://www.slub-dresden.de/ueber-uns/slubarchiv/slub-workflows/goobi-workflow/>; in German).

### Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

### **13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.**

*Minimum Required Statement of Compliance:*

3. In progress: We are in the implementation phase.

#### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

SLUBArchiv is based on the Digital Preservation Software Rosetta, which has been designed to be compliant with OAIS (see <http://www.exlibrisgroup.com/category/RosettaOverview>). Rosetta's data model is based on PREMIS (see <http://www.loc.gov/standards/premis/pif-presentations-2012/RosettaPREMIS.pdf>). METS (see <http://www.loc.gov/standards/mets/>) and Dublin Core (<http://dublincore.org/>) are used to encode the metadata. Rosetta's METS profile is publicly available (see <http://www.loc.gov/standards/mets/profiles/00000042.xml>).

TUD/ZIH and SLUB have a plan for infrastructure development. It is planned to replace the hardware infrastructure in cycles (based on the suggested operating times of the components) considering technology developments. System software and Rosetta is updated regularly. The responsible staff follows specified updating procedures that include a test installation and a functional check. Furthermore, SLUB actively follows the developments in the area of digital preservation systems and has implemented an exit strategy which allows for data migrations to a new preservation software solution.

#### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 14. The data consumer complies with access regulations set by the data repository.

### *Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

## Applicant Entry

### *Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### *Self-assessment statement:*

SLUBArchiv is a pure archive, which can be accessed by SLUB staff members only (see guideline 0).

Digitized material is made accessible via Goobi.Presentation. Goobi.Presentation considers the legal status of the digital objects and will display only objects with the necessary license to the consumers. Since 2015, SLUB publishes all digitized documents for which this is achievable under the license CC-BY-SA 4.0 (see guideline 5). Works that are subject to copyright regulations or that are located in another institution are only published if there is either a contract between the author/institution and SLUB or if the author/institution has transferred the publications rights to SLUB. In these cases, a more restricted open license might hold. As of now, the license of published digital documents is shown in a box next to the presentation of each collection (a list of SLUB's collections is here <http://digital.slub-dresden.de/en/digital-collections/>).

Digitized documents for which SLUB does not yet have presentation rights for is stored separately and not presented to the end user.

## Reviewer Entry

### *Accept or send back to applicant for modification:*

Accept

### *Comments:*

**15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

**Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The terms of use of SLUB's digital collections are published (<http://digital.slub-dresden.de/en/terms-of-use-for-digital-collections/>). SLUB's goal is to provide the data with low barriers.

**Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## **16. The data consumer respects the applicable licences of the data repository regarding the use of the data.**

*Minimum Required Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

The general terms of use are published at <http://digital.slub-dresden.de/en/terms-of-use-for-digital-collections/>. The majority of SLUB's documents are licensed under the Creative Commons License CC-BY-SA 4.0. This license allows to reproduce and to distribute the content, to adapt and build upon the material for different purposes, as long as the users credit SLUB as their source and license their new creations under identical terms. Some digital documents, e.g. those of other institutions, are licensed with "Rights reserved - free access". They are free of charge for academic and private purposes, provided that the source is acknowledged completely. Use of individual images in academic publications with an edition of less than 1,000 copies shall also be free of charge. The commercial use of these documents may be subject to charge and requires the prior written agreement of the rights' holder. When SLUB becomes aware of misuse, it contacts the corresponding person or organization.

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*