Implementation of the Data Seal of Approval

The Data Seal of Approval board hereby confirms that the Trusted Digital repository Repository of Charles University in Prague, Map Collection 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 February 22, 2016.

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:	DSA-booklet_2014-2017.pdf
All Guidelines Documentation:	Documentation
Repository:	Repository of Charles University in Prague, Map Collection
Seal Acquiry Date:	Feb. 22, 2016
For the latest version of the awarded DSA for this repository please visit our website:	http://assessment.datasealofapproval.org/seals/
Previously Acquired Seals:	None
This repository is owned by:	**Charles University in Prague, Central Library**

T +420 222 322 627
E ukn@ruk.cuni.cz
W http://www.cuni.cz/

Assessment

0. Repository Context

Applicant Entry

Self-assessment statement:

The Map Collection at Charles University (https://www.natur.cuni.cz/geography/map-collection?set_language=en) consists of approximately 130,000 map sheets, around 3,000 atlases and 60 globes. Approximately 50% of those assets are made up of early prints and manuscripts dating to about 1850. Its scope and content makes this collection unique not only within Czech Republic, but also throughout Central Europe. Digitization of the collection is funded by a continuing project - TEMAP (http://www.temap.cz/en). The Map Collection is under curation of Faculty of Science. Digitized versions of map sheets and atlases are stored in the Repository (http://repository.cuni.cz/) - a central storage point of digitized materials at the Charles University in Prague (CU - http://www.cuni.cz/UKEN-1.html). The Repository, for the respective Map Collection (as defined in the internal agreement - see Guideline 8) is responsible for SIPs processing, AIPs creation as well as long-term preservation of digitized materials. Access to documents is provided in cooperation with the Central Catalogue of CU (http://ckis.cuni.cz/). Data are also (re-)used by external systems/portals – e.g. OldMapsOnline (http://www.oldmapsonline.org). From the data producer and users point of view, the Map collection in the Repository focuses on an easy-to-use interface, which allows for publishing digitized materials easily, improving research and popularization of maps. Our Map Collection in the Repository follows the standard principles of a high quality digital repository like usage of persistent identifiers and standard metadata (METS - http://www.loc.gov/standards/mets/, Marc21 - http://www.loc.gov/marc/bibliographic/, MIX - http://www.loc.gov/standards/mix/), sharing data and metadata (via standard protocols like Z39.50 and OAI-PMH), authorisation and authentication (LDAP, Shibboleth). The system is based on DigiTool which tries to conform to the OAIS reference model (ISO 14721:2012 - http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284).

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

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:

The original documents of Map Collection are being digitized by an external company - Microna. Faculty of Science at the Charles University in Prague (with its respective Map Collection department) acts as a data producer (as defined in ISO 14721:2013).

During the preparatory stage of the TEMAP project, workflow, guidelines and responsibilities for Microna as well the Faculty were created and modified according the needs and envisioned outcomes of the grant-funded project.

Microna has been contracted not only to digitize the chosen cartographic materials, but also to generate the technical metadata required for object creation (tools, SW used, etc.).

The descriptive metadata (including geographic information) is created by designated experts from the Map Collection department.

User copies of the data and respective SIPs are created by the Repository. Data is imported in the form of bulk ingests, each amounting to ca. 1000 digitized map sheets.

General Guidelines for data producers are published on the website of the Charles University Library.

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:

Each original of a cartographic material (mostly map sheets) has two copies stored in the Repository – a master copy (in TIFF) and user copy (in JP2000). Inges policy (including file formats) is available at http://knihovna.cuni.cz/digitool/import-dat-do-repozitare/mapova-sbirka/.

User copies are generated by the Computer Centre of Charles University according to the rules agreed by all parties in initial stages of the TEMAP project (see Guideline 1).

During the process of image transformation, validity of format is checked. Quality of images is checked manually. In the case of an error (invalid formats or problematic content quality - see Guideline 12 for more details) the files are not ingested into the Repository and are returned to the producer for a future re-submission.

Microna (the digitizing company for the TEMAP project at the Charles University) has been contracted not only to digitize the chosen cartographic materials, but also to generate the technical metadata required for object creation (tools, SW used, etc.).

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:

Descriptive metadata is provided in cooperation with the Central Catalogue of CU (see the Guideline 0 - Repository context). Metadata records are created by the Data Producer (as defined in ISO 14721:2013) at the Faculty of Science. They have the expertise and knowledge required for description of historical cartographic materials. The metadata is stored in the Repository in the MARC XML format.

Structural metadata is automatically generated in the form of a METS record during the initial stages of the ingest process. Structural information itself is created by the Map Collection (in the form of a directory structure). After that an automatic transformation tool creates the METS accordingly.

Access rights metadata is attached to each object (on the level of the METS file). Before the object is displayed, users are required to accept a licence agreement (see Guideline 5 for more details).

Provenance information and MD5 hashes are part of the technical metadata record.

Reviewer Entry

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:

The Map Collection at Charles University in Prague consists of approximately 130,000 map sheets, around 3,000 atlases and 60 globes. Approximately 50% of those assets are made up of early prints and manuscripts dating to about 1850. Its scope and content makes this collection unique not only within Czech Republic, but also throughout Central Europe. Digitalization of the collection is funded by a continuing project - TEMAP. The Map Collection is under curation of Faculty of Science.

The digitized versions of map sheets and atlases are stored in the Repository - the central storage point of digitized materials at the Charles University in Prague (CU).

The Repository is responsible for SIPs processing, AIPs creation as well as long-term preservation of digitized materials. Access to documents is provided via the Repository's user interface and the Central Catalogue of CU search portal. The data is also (re-)used by external systems/portals – e.g. OldMapsOnline.

Given the fact that the owner of both the original documents and the digitized materials is the CU Faculty of Science or, more precisely, The Map Collection as its subsidiary, the succession plan is under the purview of that department. With the upcoming conclusion of the project (2011 – 2015) intensive negotiations are underway concerning significant characteristics and long term preservation of the documents.

The digital repository's Mission Statement is published in Czech language on the website of the founder of the Repository as follows:

“Digital Repository of the Charles University in Prague

Mission Statement

The Digital Repository of the Charles University in Prague serves as the principal storage facility for the digital and digitized documents which the Charles University in Prague (hereinafter CU) owns. Including those documents that have been created by the students, masters, scholars, and other members of staff affiliated to CU. The Repository's mission is a long term storage and preservation of documents in compliance with the national and international standards and "best practice".

The Repository offers access to its content in observance of the copyright law, other applicable legislation and internal policies of the Charles University in Prague.

The research activities of the team affiliated to the Digital Repository focuses on enhancing the awareness of long-term protection of digital documents, in the academic environment in particular.”

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

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:

The Repository itself does not enjoy a separate legal personality but falls under the Charles University in Prague, within which it is administered by the University's Central Library.

An Internal Agreement has been made for the needs of the Map Collection between organisational units of the Charles University in Prague on digitisation and archiving of, and provision of access to, the Map Collection of the CU Faculty of Science. In that Agreement the data producer (Faculty of Science) declared its intention to store the digitized versions of the maps in the Repository and specifies the conditions under which they should be made public.

The Agreement was originally made between the Faculty of Science and the Computer Science Centre under whose purview the Repository had been. With the transition of the Repository's administration under the Central Library, all contractual obligations from the above mentioned Internal Agreement have been assumed.

Public access is only made to the user copies of the full versions of data. Such copies are of lower quality and provided with watermarks. Before viewing a digital object the user must agree to abide by the licence terms. The exact wording of the licence is required by the Faculty of Science (see below):

"Please read carefully the text below before viewing

Warning on same restrictions defined by Copyright Act (Act N. 121/2000 Coll.): Under conditions defined by Copyright Act libraries, collections and archives are authorized make free of charge photocopies or other reproductions of author craft. One of these specified conditions is fact, that this photocopies or other reproductions may not be used for any other purpose than private study, scholarship or research. If a user makes a request of document or later a photocopy or reproduction of document (author craft) in conflict with conditions defined by

Copyright Act for free of charge use, that user will be liable for copyright infringement. This collection reserves the right to refuse acceptation of a copying order, if in its opinion fulfilment of such order would involve copyright infringement. I have read the text above and agree to abide by relevant restrictions."

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:

The technical infrastructure is provided by the Computer Science Centre of the Charles University in Prague. The Repository's principal installation is based on a virtual server (Oracle VM 3.3.) with 4 vCPU and 12GB RAM. OS is RHEL 5 Server x86_64.

The data is subject to regular backups and fixity checks. The data is stored on the disk array DELL Compellent SC8000.

All the data is regularly and completely backed up from the master server to the storage facility at the Charles University campus in Jinonice. Moreover, the increments and changes to the file system are backed up daily to the tape library ADIC Scalar i500, LTO-4 and to a disk space.

The disk space currently available is 25 TB. The expected annual increment for the TEMAP project is ca. 3 TB.

In the previous years we have undergone DRAMBORA self-audit and SPOT analysis (2014) that helped us identified risks and analyze them. recovery of data (moving virtual machines locally) will not have any severe impact to user experience. E.g. in case of a natural disaster (e.g. flods or fire in the server room) a downtime of days (moving data to new server) will occur. We are working on implementation of back-up locality (back-up servers not just data back-ups) which should be finished in Q2 2016.

The checksums (MD5) are conducted on a monthly basis by running a script control. The machine control reports a list of objects where the checksums do not equal the metadata information. The staff will verify the circumstances of such discrepancies and may suggest remedial actions. As the DIP consists of master (JP2) and user copy (TIFF), the staff will identify which copy the MD5 discrepancies is related to. In case of a master copy, users will not be limited, since the remedial actions (back-up receovery) will be done on a master copy. In case it is related to the user copy, a new one from the master copy is generated (by an automatic tool in the Repository),

which is a question of several hours. In case both master or user copy are corrupted, the complete digital object is taken from a back-up. Access to the specific object is a question of hours up to a day.

Automated checks for duplicate objects are also run. In future we plan regular synchronisation of the data with the union catalogue so that it contain the updated versions of descriptive metadata records from the CU Central Catalogue.

In exceptional cases, e.g. in the case of duplicate digitisation, the relevant data is deleted following consultation with the Map Collection. The process of deletion is logged and supported by email correspondence.

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:

Each submission is verified and validated by automatic tools and manually by the Repository's administrator. The Repository performs checks for fixity, image quality and for duplicate objects. If any of these criteria is not met, employees of the Map Collection are notified. Any errors and discrepancies are properly documented in e-mail and following steps are decided. Further processing can be postponed for the whole bulk ingest (if it is a severe error) or only affected files are withdrawn from further processing. Once corrected they are resubmitted in a following bulk ingest. Ingests are numbered, each ingest has a documented list of submitted/rejected/resubmitted documents.

During the process of image transformation, validity of format is checked. Quality of images is checked manually. If an error is detected (invalid formats or problematic content quality - see Guideline 12 for more details), the files are not ingested into the Repository and are returned to the producer for future re-submission.

Each original of a cartographic material (mostly map sheets) has two copies stored in the Repository – a master copy (in TIFF) and a user copy (in JP2000).

User copies are generated by the Computer Science Centre of Charles University according to the rules agreed to by all parties in initial stages of the TEMAP project.

As the formats are specified, we minimize the cases of file formats obsolescence in the near future. Thus, the Repository (for its component – the Map Collection) does not anticipate in foreseeable future a massive migration of the master copies (in TIFF format) of the digitised documents.

With a view to the rapid development of technologies and improved quality of displaying options for pictorial materials on the user side, the Repository currently work on the migration strategy of user copies for 2016 and beyond.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

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 submission workflow is defined on the basis of several years of cooperation and internal agreement with the Map Collection.

The aim of this internal agreement is to set a framework and liabilities of two university bodies for the digitization, archiving and dissemination of digitized cartographic materials of Map Collection. The agreement deals with processing of digitized cartographic materials (texts, maps, plans, atlases and globes) and their respective metadata; incorporating security features into the user copies (DIPs), archiving of digitized documents and metadata as well as accessibility/availability of such digitized documents. The main liabilities of the Map Collection are: consent of archiving, processing and making digitized document accessible to public via the Digital University Repository's interface. The Map Collection provides scans in GeoTIFF format and MIX metadata and a watermark "© Mapová sbírka P?F UK" as one of the security features. The liabilities of the Computer Centre or more precisely the Digital University Repository are: processing of digitized GeoTIFF document, creating user copies in JPG2000 format with the above-mentioned watermark text, archiving digitized materials (applying requirements of OAIS - ISO 14721:2003) and ensuring regular checks for long-term preservation as well as ensuring access to user copies of digitized materials. The Repository makes sure that the materials will not be handed over to a third-party without a written approval of the Map Collection representative. Both parties agree that the processed digitized copies are kept in Digital University Repository and on back-up servers of the university. This agreement is valid until 31st of December 2015, except for the archiving, long-term preservation and dissemination of digitized cartographic materials. For these purposes the agreement is valid for an indefinite period (as stated in the Article IV, Section 1).

The workflow is documented in the form of a flow chart and comprises the following steps:

1. Delivery of data on a portable disk (or, as the case may be, sFTP) – the data includes both the files themselves and the technical metadata containing the MD5 checksums.

2. Running the data checks against the MD5 supplied

3. Generation of user copies (migration to JPEG2000 and generation of watermarks)

4. Generation of SIP – container format METS – contains structural information and technical metadata

5. Ingest to the Repository

6. Automated assignment of access rights

7. Exporting the list of digital entities identifiers

8. Complementing the descriptive metadata on the basis of the list of identifiers (manual check of archived objects)

9. Regular checks for duplicate objects and for objects without descriptive metadata

10. Regular checks of the objects stored in the Repository against the MD5 supplied during the scanning process

The data is subject to regular backups and fixity checks.

All the data is regularly and completely backed up from the master server to the storage facility at the Charles University campus in Jinonice. Moreover, the increments and changes to the file system are backed up daily.

The checksums (MD5) are conducted on a monthly basis by running a script control. The machine control reports a list of objects where the checksums do not equal the metadata information. The staff will verify the circumstances of such discrepancies and may suggest remedial actions.

Automated checks for duplicate objects are also run. In future we plan regular synchronisation of the data with the union catalogue which may contain more up-to-date versions of descriptive metadata records.

In exceptional cases, e.g. in the case of duplicate digitisation, the relevant data is deleted following consultation with the Map Collection. The process of deletion is logged and supported by email correspondence.

Documentation of these processes is in the implementation phase. The documentation is shared by means of a shared and backed up storage facility.

The data is made accessible on the Repository's website in compliance with the producer's requirements as provided for in the "Internal agreement between organisational units of the Charles University in Prague on digitisation and archiving of, and provision of access to, the Map Collection of the CU Faculty of Science" and formulated at regular meetings and in email correspondence.

The metadata records are further provided to the systems Google and Georeferencer. As the latter works with a proprietary metadata standard, conversion is required. The process is automated via scripts.

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:

Repository, or, as the case may be, CU Central Library assumes responsibility for the long term protection for the digitized materials received on the basis of the established workflow and the provisions of the "Internal agreement between organisational units of the Charles University in Prague on digitisation and archiving of, and provision of access to, the Map Collection of the CU Faculty of Science" signed in 2013. The Agreement specifies obligations of the parties thereof, including the form of presentation of user copies in the Repository's interface and long-term preservation of digital documents.

Subsection 3 of Article II of the above mentioned Agreement stipulates: Contractor agrees to ensure the secure long-term preservation of data in the Digital University Repository. While preserving data, he obliges himself to account for the impacts of changing technologies (e.g. by migrating the data to more advanced software) and to conduct periodical checks of data preservation: for both of these purposes, he shall implement the requirements of the OAIS model (ISO 14721:2003) and the requirements for a trustworthy digital repository.

Descriptions of the technical infrastructure, provision of availability of digital objects and implementation of the OAIS standard in the Repository are provided in Guidelines 6 and 10.

In the previous years we have undergone DRAMBORA self-audit and SPOT analysis (2014) that helped us identified risks and analyze them. For the time being, our crisis management concerning the availability of the digital objects (digitized cartographic materials) is addressed on a technical level. In this case (moving virtual machines locally), the procedure will not have any severe impact to user experience. E.g. in case of a natural disaster (e.g. flods or fire in the server room) a downtime of days (moving data to new server) will occur. We are working on implementation of back-up locality (back-up servers not just data back-ups) which should be finished in Q2 2016.

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 Map Collection preserved in the Repository is available to the target audience through the user interface at http://digitool.is.cuni.cz/R/?func=search&CON_LNG=ENG. Because of better remembering of the Repository URL name (for our Czech users), we decided to create aliases to the original URL, at http://repozitar.cuni.cz and http://repositar.cuni.cz (Czech version of the UI). For the general public from abroad and our Erasmus students we created easy-to-remember alias with English interface: http://repository.cuni.cz.

The digital documents which are part of digital objects include, in addition to the relevant metadata (in Marc21), the master copy in TIFF and the user copy in JPG2000. The user copies are made available via the Repository's native viewer JP2000, so the users do not need to download any additional software for viewing them.

The master copy is not available to users, not only due to the sheer size of the files but also for the sake of long-term preservation. Each digital object is provided with an internal persistent identifier (PID) generated by the system DigiTool. As a result, the Repository provides for the direct and persistent linking to the files of user copies. Such links are also used in the project Georeferencer and available in the OldMapsOnline interface. Repository also offers the function of Handle identifiers but it was not employed for the needs of the TEMAP project.

The outputs from the TEMAP project can be searched for in the CU Central Catalogue, from where the user is redirected to the Repository. Searching functions are also offered by the Repository's own interface, including both simple and advanced options.

The metadata TEMAP records in the Repository are harvested by Google.

The metadata can be shared via OAI-PMH protocol.

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:

Availability of the files, web and applications servers is regularly monitored. The access to them is subject to a stringent policy.

All the data is regularly and completely backed up from the master server to the storage facility at the Charles University campus in Jinonice. The increments and changes to the file system are backed up daily.

The checksums (MD5) are conducted on a monthly basis by running a script control. The machine control reports a list of objects where the checksums do not equal the metadata information. The staff will verify the circumstances of such discrepancies and may suggest remedial actions.

Automated checks for duplicate objects are also run. In future we plan regular synchronisation of the data with the union catalogue which may contain more up-to-date versions of descriptive metadata records.

The Repository does not currently work with versions of the digital object (on the capacity grounds). In exceptional cases, e.g. in the case of duplicate digitisation, the relevant data is deleted following consultation with the Map Collection. The process of deletion is logged and supported by email correspondence. So the new versions of files replace the original versions which are no more available for discovering or linking to.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

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:

The Repository (for its component - The Map Collection) does not anticipate in foreseeable future a massive migration of the master copies (in TIFF format) of the digitised documents.

With a view to the rapid development of technologies and improved quality of displaying options for pictorial materials on the user side, the Repository currently work on the migration strategy of user copies for 2016 and beyond.

Manual spot checks of master copies for the quality of scanning are conducted at the submission to the Repository. The checks include, inter alia, improperly cleaned glass of the scanner, calibration of colours, illumination of the document being scanned, scanned images with mechanic distortion of the document, legibility of information on maps, etc.

The checks are also run for the user copies and just twice – during submission and when assigning the descriptive metadata.

When a corrupt file is detected, the master copy is checked for a similar corruption. If the master copy is OK, a new user copy is generated to replace the corrupt file. If the master copy shows the same corruption, it is replaced with a backup copy.

The Repository stores information on the creation of a digital object (the type and setup of the scanner, the image processing software) within the technical metadata. All operations carried out in the Repository are logged; such logs are available for future inspection if needed. History of changes is generated for each object.

The objects submitted to the Repository are provided with an identification number of the descriptive record with which they are subsequently enriched. The identification number, however, is further stored independently of the descriptive metadata.

In the case of the Map Collection, the producer is the Repository's staff. Their identities are controlled on the basis of User ID and Password. The authorization to work with the descriptive metadata is granted to a limited group of the Map Collection staff. These members of staff, however, are not authorized to delete or change the digital objects.

The Repository currently work on establishing a precise list of significant features of the images from the Map Collection.

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:

The Repository is being built within the system DigiTool, which covers nearly all functional units of the Reference Model of an Open Archival Information System (ISO 14721:2013) – ingestion, archival storage, data management, access, and administration. Preservation planning is provided outside the technical infrastructure of the system DigiTool.

1) Pre-Ingest: Data are delivered to the Computer Centre on a portable disk (or, as the case may be, sFTP) – the data includes both the files themselves and the technical metadata containing the MD5 checksums. The staff at the Computer Centre runs checks against the MD5 supplied. If MD5 are valid, user copies (migration to JP2000 and generation of watermarks) are generated. This way prepared data are ready for the Repository to be processed. First step included creating SIPs (with the use of custom made application). METS is used as a container format METS.
2) Ingest: The Submission Information Packages (SIPs) are prepared for bulk-ingesting. The default way is that the ingestion process is done through our Unix based interface, so that ingest process can be followed in the real time and in case of problems stopped or re-run again.
3) Data Management: Data Management tools assign to the digital objects respective access rights. A list of digital entities identifiers is exported and sent to the staff of Map Collection for descriptive metadata addition.
4) Archival Storage: After the completion of Ingest and Data management's jobs, the Archival Information Packages (AIPs) is created, metadata is stored in Oracle database, files are submitted to the file system (respective storage group for Map Collection data).
5) Administration: We have developed a specific robust administration interface including specific detailed reports on the contents of our repository.
6) Access: The available Dissemination Information Package query responses and reports are delivered to consumers. DIPs are publicly available, however, prior to access to digital objects, users need to agree with our copyright agreement (on screen). Repository allows for searching, locating and description of the Map Collection stored.
7) Preservation Planning: As described in 6., we monitor and backup our system. More preservation details are described in 9). In repository context, each submission bitstream has MD5 checksums which are regularly checked. There is a list of supported and known formats (for user, master copy and thumbnails) whose consistency are regularly checked using existing tools.

For descriptive metadata, the Repository employs the standards METS (see http://www.loc.gov/standards/mets/), MODS (http://www.loc.gov/standards/mods/), Marc21 (http://www.loc.gov/marc/) and the qualified Dublin Core (http://dublincore.org/). The technical metadata support the standard schemas:

Images - NISO Z39.87 Technical Metadata for Digital Still images (MIX - http://www.loc.gov/standards/mix/)

Text – based on the schema METS (textmd.xsd - http://www.loc.gov/standards/textMD/)

Audio – based on the schema METS Audio Technical Metadata Extension Schema (AMD - http://www.loc.gov/standards/amdvmd/)

Video – based on the schema METS Video Technical Metadata Extension Schema (VMD - http://www.loc.gov/standards/amdvmd/)

The schema for long-term preservation concerns the metadata for long-term preservation of digital documents in particular. The system DigiTool employs the schema based on the metadata elements which were published by the PREMIS working group as part of their data dictionary. The metadata for long-term preservation contain the elements described in the dictionary in the subsection "Object entity".

The schema History is concerned with some metadata elements related to long-term preservation and focuses on the activities/processes/events performed with a digital object. As in the case of the preservation metadata, this schema is also based on the data dictionary of the PREMIS working group (subsection "Event entity").

The access rights are defined by the standard METS Rights Declaration (http://www.loc.gov/standards/rights/METSRights.xsd).

The monitoring activities of the Preservation Planning module are provided by a script tailored to our needs and checking at regular intervals integrity of the files.

The monitoring of the target audience is covered by questionnaires. Based on the results of the end-2014 poll we will re-design (thanks to awarded grants) the user interface presenting the user copies of the digital documents. The goal is to present a user-friendly interface.

We currently work on formulation of a university-wide policy for long-term preservation.

Other elements of the Preservation Planning module are gradually added to the entire Repository, not just for the digitized outputs of the Map Collection.

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:

Only the user copies of the full versions of data are published. These copies are of lower quality and provided with a watermark. Therefore, full authentication is not required. The data consumer, however, is required to agree to abide by the licence terms (see Guidelines 15 and 16).

If the licence had not been adhered to, it is possible to retrieve, using the IP address, the materials which were made available to a given data consumer.

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:

Only the user copies of the full versions of data are published. These copies are of lower quality and provided with a watermark. Therefore, full authentication is not required. The data consumer, however, is required to agree to abide by the licence terms (see Guideline 16).

If the licence had not been adhered to, it is possible to retrieve, using the IP address, the materials which were made available to a given data consumer.

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 user copies made accessible by the Repository are, unlike the master files, of lower quality and provided watermarks. Therefore, the Map Collection opted for a free access without authentication. Before viewing a digital object the data consumer must agree to abide by the licence terms. The exact wording of the licence is required by the Faculty of Science (see below):

“Please read carefully the text below before viewing

Warning on same restrictions defined by Copyright Act (Act N. 121/2000 Coll.): Under conditions defined by Copyright Act libraries, collections and archives are authorized make free of charge photocopies or other reproductions of author craft. One of these specified conditions is fact, that this photocopies or other reproductions may not be used for any other purpose than private study, scholarship or research. If a user makes a request of document or later a photocopy or reproduction of document (author craft) in conflict with conditions defined by Copyright Act for free of charge use, that user will be liable for copyright infringement. This collection reserves the right to refuse acceptation of a copying order, if in its opinion fulfilment of such order would involve copyright infringement. I have read the text above and agree to abide by relevant restrictions.”

If the licence had not been adhered to, it is possible to retrieve, using the IP address, the materials which were made available to a given data consumer.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments: