Implementation of the Data Seal of Approval

The Data Seal of Approval board hereby confirms that the Trusted Digital repository GAMS (Geisteswissenschaftliches Asset Management System) 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 December 19, 2014.

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: GAMS (Geisteswissenschaftliches Asset Management System)
Seal Acquiry Date: Dec. 19, 2014

For the latest version of the awarded DSA   http://assessment.datasealofapproval.org/seals/
for this repository please visit our website:

Previously Acquired Seals: None

This repository is owned by: **Austrian Centre for Digital Humanities -- Zentrum für Informationsmodellierung**

Austria

T 0043 316 380 8009
E zim@uni-graz.at
W http://informationsmodellierung.uni-graz.at/

Assessment

0. Repository Context

Applicant Entry

Self-assessment statement:

The Centre for Information Modelling – Austrian Centre for Digital Humanities was founded in 2008 with the intention of focussing methodological skills at the intersection of ICT and Humanities and establishing a sustainable infrastructure for ICT-assisted humanities' research.

According to the original intent stated in our founding declaration, the Centre's focus during its formative phase was on:

a)   Building a research-related IT infrastructure (virtualized server pool with dedicated servers for certain services and web-services as well as a digital repository),

b)   Applied research in the fields of humanities, information modelling and -processing,

c)   Research support and participation in a variety of cooperative projects and

d)   Establishment of a pertinent academic lecture program.

The Centre took over a technologically proprietary pool of research-supporting software projects from its predecessor. These were transferred to a structural project for long-term archiving and provision of scientific data and content in a standardized environment, which goes by the acronym GAMS (Geisteswissenschaftliches Asset Management System). GAMS was conceived and developed on the basis of the Open-Source project FEDORA (Flexible Extensible Digital Object Repository Architecture) (http://fedora-commons.org/) and has been continuously improved in the course of cooperative projects, addressing the specific needs of university research. FEDORA is fully OAIS (Open Archival Information System)-compliant and the GAMS repository covers the full life cycle of digital objects from receiving the SIP (submission information package), archiving the AIP (archival information package) and delivering the DIP (dissemination information package) to the public. Consequently, a Java application for object management and data curation was developed. This Cirilo Client offers applications which are particularly suited to being used as tools for mass operations on Fedora repository objects, such as ingest or replacement processes. It also fulfills a lot of functions with regards to metadata enrichment and quality control

of the resources. It is available as an open source software package as a contribution to DARIAH-EU.

Centre Website (2014-08-26):

http://informationsmodellierung.uni-graz.at/en/

Repository (2014-08-26):

http://gams.uni-graz.at/

Documentation (2014-08-26):

http://gams.uni-graz.at/doku

GitHub Site of the Cirilo client (2014-08-26):

https://github.com/acdh/cirilo

Please note: The Centre's new founding declaration will be published until the end of this year; the hyperlink will stay the same (http://static.uni-graz.at/fileadmin/gewi-zentren/Informationsmodellierung/PDF/gruendungserklaerung-zim.pdf). The new declaration will state more explicitly the Centre's mission in long-term preservation and archiving.

At the end of 2012, the Centre underwent an external assessment and was evaluated very positively. The new founding declaration ensures the existence of the Centre for another 5 years starting in 2013. Thus, sustainability and funding can be guaranteed at least until 2018. Since 2012 the number of employees has doubled and the university will set up a professorship for Digital Humanities next year. We are therefore confident, that also after

2018 the Centre and its infrastructure will not cease to exist.

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 Centre conducts projects as a cooperative process of negotiation and consulting with researchers from various disciplines. This process is also characterized by raising the awareness for IT methods and standards among "traditional" humanities scholars in their respective disciplines. The projects considered suitable for inclusion in the repository are usually externally funded and peer-reviewed research projects, which ensures that the methods applied and data generated conform to domain-specific standards and have adequate scholarly quality.

The name and affiliated organization of the data producer can be traced in the source document of all resources, for instance by annotating the relevant information in the header of a TEI (Text Encoding Initiative) document (cf. an example here: http://gams.uni-graz.at/o:vase.1469/TEI_SOURCE). Creator, annotation rules and revision information are therefore directly available for assessment by the user of the resource. The legal status of the data and objects in question are clarified with the project partner; usually all resources are available under a non-commercial Creative Commons license (cf. http://gams.uni-graz.at/, http://creativecommons.org/licenses/by-nc-nd/3.0/at/).

There are no standardised guidelines or checklists for project partners. These questions are treated individually upon advent of a new project and respective data. Each project begins with the formulation of a "Cooperation Agreement" covering the workflows, responsibilities and rights of the partners in the research project and a separate "Deposition Agreement" which lists the rights and responsibilities of the Depositor and the Repository with regards to the research data placed in the repository. A reference deposition agreement is available at http://static.uni-graz.at/fileadmin/gewi-zentren/Informationsmodellierung/PDF/Repository-Depositors-Agreement_GAMS_V3.pdf .

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:

GAMS is based on XML (eXtensible Markup Language) -based standards and technologies for data storage and representation. If the data in question does not conform to any XML-based international standard, the Centre will implement suitable workflows for the conversion of the content in agreement with the project partners. For text and metadata, the Centre uses (among others) the following de facto standards: TEI (Text Encoding Initiative), DC (Dublin Core), METS/MODS (Metadata Encoding and Transmission Standard/Metadata Object Description Scheme), RDF (Resource Description Framework), SKOS (Simple Knowledge Organization System). This list of preferred formats is reflected in the use of dedicated content models for the respective standards (cf. http://gams.uni-graz.at/doku#cirilomodels). The Cirilo Client then checks the well-formedness of the XML and validates the document against the given schema to ensure conformity. Data producers must deliver images in the recognized standards JPEG/JPEG2000 (Joint Photographic Experts Group) or TIFF (Tagged Image File Format); if necessary the Centre assists in a conversion process.

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:

Data without metadata is not considered suitable for deposit in the repository: The presence of a DC metadata datastream is the minimum requirement for descriptive metadata. Structural metadata is available in the METS/MODS format and is also recorded together with administrative metadata for every object in its compulsory RELS-EXT (relationships-external) datastream. Usually, metadata will already be included in the primary source document (for instance TEI and METS/MODS) and will be mapped during the ingest process into the repository to the minimal DC datastream (cf. http://gams.uni-graz.at/doku#cirilomodels). This provides an efficient and user-friendly tool for the capture of metadata. The DC metadata record is the basis for the application of OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting) and is also represented in Europeana (www.europeana.eu) to support resource discovery.

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:

In the published mission statement of the Centre (German, http://static.uni-graz.at/fileadmin/gewi-zentren/Informationsmodellierung/PDF/gruendungserklaerung-zim.pdf) long-term archiving and preservation of research data in a digital repository are stated explicitly as main tasks of the Centre. This is acknowledged by the University, and the Centre's infrastructure partly acts as an institutional repository for the Faculty of Arts and Humanities in Graz (cf. the wide variety of projects and collections at http://gams.uni-graz.at/). The Centre also contributes its expertise in these fields to the Austrian DARIAH (Digital Research Infrastructure for the Arts and Humanities) and CLARIN (Common Language Resources and Technology Infrastructure) activities (cf. http://acdh.oeaw.ac.at/dha/node/30), with the Cirilo Client being released as a free software package on GitHub (https://github.com/acdh/cirilo).

The new founding declaration states the following as one task of the Centre: „Betrieb eines Digitalen Repositoriums (Geisteswissenschaftliches Asset Management System GAMS) und einer forschungsbezogenen IT-Infrastruktur zur Langzeitarchivierung von (geisteswissenschaftlichen) Forschungsdaten." Translation: "operation of a digital repository (GAMS) and a research-related IT infrastructure for long-term preservation of (Humanities') research data". It also states that with regard to the covered research topics "questions of sustainability and long-term preservation of Humanities' research data are central".

The Centre actively contributes to the new formed DARIAH-EU working group on preservation.

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 GAMS repository is run as a project of the Centre for Information Modelling – Austrian Centre for Digital Humanities. The Centre is an organizational unit of the Faculty of Arts and Humanities of the University of Graz, which is in turn (like all Austrian Universities since 2002) a "juristische Person öffentlichen Rechts" – a legal entity organized under public law. The Centre and the repository are not separate legal entities (cf. http://static.uni-graz.at/fileadmin/gewi-zentren/Informationsmodellierung/PDF/gruendungserklaerung-zim.pdf).

Data can only be deposited as part of a cooperation project. Therefore, issues like IPR (Intellectual Property Rights) and licensing (possible limitation of access) are discussed and determined already in the course of project planning. The Centre also offers consulting and expertise in these fields (cf. our research focus on IPR: http://informationsmodellierung.uni-graz.at/en/research/other-projects/current-projects/). Project partners are responsible for respecting national and international laws. Depending on the project, suitable agreements on these legal terms are signed, but – due to the diversity of the data and the project partners – not in a standardized way.

A reference deposition agreement is available at http://static.uni-graz.at/fileadmin/gewi-zentren/Informationsmodellierung/PDF/Repository-Depositors-Agreement_GAMS_V3.pdf, cooperation agreements are set up with the partners at the beginning of a project. Both are project specific and customizable with regard to access and use of data.

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 provided via SAN by our university's IT department (UNI IT). Data is stored redundantly in two data centers in different campus buildings. A formal service level agreement between the Centre for Information Modelling and the IT department as hardware and data provider is currently in preparation.

Data backup in GAMS is part of the central backup processes of the University. Backups run daily and are stored on a disk array and later moved to tape. There is an additional offsite backup managed by the Centre which is also run every night. The combination of both backups ensures their accessibility over a period of seven years.

Backup consistency is guaranteed because every FEDORA object is entirely stored in FOXML format containing all binary data streams in base64 encoding. Additionally, all datastreams are preserved in the original format as distinct files. As each object provides MD5 checksums for the datastreams, corrupted data can be identified easily.

No data recovery has been necessary during the last 8 years. Nevertheless, data recovery is regularly exercised on a spare machine for training purposes of the administrators.

All services are monitored 24/7 with a reaction time of typically only a few minutes during working hours to some hours during weekends and nights.

Documentation on data security is available at http://gams.uni-graz.at/doku#d5e18. The whole infrastructure documentation is partly a preservation policy.

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 GAMS repository accepts data only in formats which are considered to be suitable for long-term preservation. For text-based research, this focuses on XML-based file formats, for images JPEG(2000) and/or TIFF. XML is a platform-independent, non-proprietary, machine and human readable encoding format, which reduces the risk of obsolescence. To ensure availability and usability of the content, recognized international standards for annotation are used in the repository. These include for instance DC, TEI or METS/MODS (cf. http://gams.uni-graz.at/doku#cirilomodels). This policy notably reduces the need for migration of formats and resources. Nevertheless, the Centre additionally implements continuous monitoring of the whole infrastructure with respect to new advances in long-term preservation and technical challenges by skilled staff.

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:

Archiving takes place following specific procedures. In cooperation with the project partners, the Centre takes over the data according to the specifications of the repository. This SIP is then ingested using the Cirilo Client. During ingest, metadata is extracted from the source and mapped to a DC record. Additionally, semantic enrichment (like resolution of place names and ontology concepts) can take place. Images and other related materials are bundled within the resource automatically. This also includes validation and quality control of the data as well as assignment of a PID (persistent identifier) in the system (cf. http://gams.uni-graz.at/doku).
The GAMS repository is not only used for long-term preservation but also for the web-representation of the resources. In that respect, it takes advantage of FEDORA's object model, which assumes that all digital assets are completely self-descriptive (cf. https://wiki.duraspace.org/display/FEDORA36/Fedora+Digital+Object+Model).
The production of DIPs (for instance via stylesheets and transformations) is directly encapsulated in every resource. Thus, the GAMS repository and the Cirilo Client cover the whole lifecycle of objects from SIP to DIP in the OAIS model.

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:

Usually, the project partners acting as data providers keep all rights of their data and appear as providers in the imprint (cf. an example here http://gams.uni-graz.at/archive/objects/context:vase/methods/sdef:Context/get?mode=imprint). The Centre consults partners on an access policy, for instance the selection of a suitable Creative Commons license or specific terms of use (like in this example: http://gams.uni-graz.at/archive/objects/context:vase/methods/sdef:Context/get?mode=use). The Centre promotes open access and public availability of research data as stated in the mission statement (German, http://static.uni-graz.at/fileadmin/gewi-zentren/Informationsmodellierung/PDF/gruendungserklaerung-zim.pdf), but specific access models can be negotiated with the partners, if a justifiable need arises.

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:

Electronic resources are accessible as a range of web representations and in the plain source format (see http://gams.uni-graz.at/doku). Search facilities are given for each project website and for the whole repository (http://gams.uni-graz.at/archive/search). Furthermore, resources of the repository can be harvested via OAI-PMH (http://gams.uni-graz.at/oaiprovider?verb=Identify and are represented in Europeana (www.europeana.eu).

All resources in the repository have a PID and are addressable with the permalink http//:gams.uni-graz.at/PID. Datastreams can be accessed in the same way with http//: gams.uni-graz.at/PID/DATASTREAM. This assures direct access, quotability and persistent identification for scientific contexts (cf. http://gams.uni-graz.at/doku#pid).

The Centre is member of the handle network and is running its own handle server, the prefix is 11471. A unique handle for each object can be generated in the Client. This persistent identifier is stored as part of the objects metadata and is published in the handle infrastructure.

Our repository provides a full text search. This can be used to search for all objects in the repository or for objects in specific collections. Beside this full text index we maintain additional Lucene indices for defined fields. For example all DC core elements are extracted by default when an object is ingested and stored in particular indices. This means that we can restrict searches to all or single DC elements, but also to some important content related fields like dates.

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:

FEDORA supports versioning of every aspect of the digital resource from primary source to metadata and associated materials. All changes and previous versions of the material can be retrieved.

The Cirilo client checks well-formedness of XML-formats and validates against the referenced schema (if applicable). This applies also to metadata, as it is stored as an XML-based datastream within the digital object (cf. http://gams.uni-graz.at/doku).

FEDORA uses MD5 Checksums to guarantee the integrity of the resources in the digital archive. This operation is carried out each time new material is ingested or a resource is modified.

All changes to an object are logged. This information becomes automatically part of the objects metadata. The decision if a datastream change should generate a new version or replace the existing data stream is up to the data curator. This means there is no general versioning policy for the repository, but individual policies depending on the project and material in question.

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:

Every project is supervised by a metadata coordinator. Authenticity and quality of the data is checked and maintained upon ingest of new data objects. New data is created and transmitted to the coordinator in standardized and documented workflows. In the course of this process, sufficient metadata and relations to other datasets are established. FEDORA monitors and records all changes and the complete version history of the resources automatically. Editorial changes of the content are usually also recorded in the metadata (typically the TEI Header).

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 GAMS repository and its underlying infrastructure project FEDORA are fully OAIS-compliant. The repository covers all functions of the OAIS Reference Model. Before accepting the SIP, the Centre already implements its guidelines for file formats and quality control. During ingest of the SIP, the Cirilo Client turns the package into a complete digital object in the FEDORA infrastructure. Many processes from encapsulation of metadata and images and semantic enrichment take place at this point.
The creation of the DIP for the user is also integrated in the digital object itself. Thus, the repository covers the whole lifecycle of an information package in the digital archive following a standardized workflow. Persistent identification, quotability and sufficient metadata are also monitored and managed automatically by the infrastructure.
Preservation planning and data management are continuously performed by skilled staff. The use of standards and open file formats supports long-term accessibility of the resources in the repository, while interfaces (e.g. OAI-PMH) enhance visibility. The archival storage follows state of the art principles with regard to data integrity, backup and recovery strategies. Constant monitoring of technological advances and data formats is in place and migration is performed if necessary (http://gams.uni-graz.at/doku).

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:

As indicated in its mission statement, the Centre promotes open access and free availability of research data. If not otherwise indicated, all resources of the GAMS repository are licensed under the Creative Commons license CC BY-NC-ND 3.0 AT (http://creativecommons.org/licenses/by-nc-nd/3.0/at/). This is also indicated on the repository website (cf. http://gams.uni-graz.at). Upon request of the cooperation partners who act as data providers, access regulations to their digital objects can be put in effect. FEDORA supports management of access right via XACML (eXtensible Access Control Markup Language). Another possibility is to use a password for the web representation, which can be requested by the data consumers directly from the data providers.

FEDORA offers the technical means to implement finely grained access policies down to the individual datastreams inside the objects.
In practice, we recommend the use of open Creative Commons licenses (CC-BY) to all publicly-funded projects, but offer the implementation of the full range of license types covered in the Europeana Rights Statement [http://pro.europeana.eu/available-rights-statements]. Since the licences for accessing the data are part of the "Deposition Agreement", both parties can be held responsible for breaches of contract regarding these regulations (and any other regulations stated in the agreement).

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:

Cooperation partners are responsible for ensuring the conformance to any codes of conduct, IPR or other legal regulations relevant to their research. With regards to confidential data like personal or human subjects' data, the repository only accepts appropriately anonymized research data (though this has occurred rarely so far). Specific terms of use for consumers can be negotiated between the cooperation partners and the Centre, though the Centre encourages the use of a non-commercial Creative Commons license. Where needed, authentication by passwords or specific XACML regulations is established (cf. http://gams.uni-graz.at).

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 GAMS repository is aware of and complies with national and international legislation with regards to data (re-)use, IPR or privacy rights. The Centre has a research focus on the legal implications of digital research and also offers consulting in this field
(http://informationsmodellierung.uni-graz.at/en/research/other-projects/current-projects/).
The resources in the repository are licensed under a Creative Commons standard, usually the license CC BY-NC-ND 3.0 AT (http://creativecommons.org/licenses/by-nc-nd/3.0/at/).
Users are informed of this policy on the web representation of the repository (http://gams.uni-graz.at).

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments: