Implementation of the CoreTrustSeal

The CoreTrustSeal board hereby confirms that the Trusted Digital repository Meertens Institute complies with the guidelines version 2017-2019 set by the CoreTrustSeal Board.
The afore-mentioned repository has therefore acquired the CoreTrustSeal of 2016 on March 5, 2018.

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

Yours sincerely,

The CoreTrustSeal Board

Assessment Information

Guidelines Version: 2017-2019 | November 10, 2016
Guidelines Information Booklet: [CTS Requirements 2017-2019](#)
All Guidelines Documentation: [Documentation](#)

Repository: Meertens Institute
Seal Acquiry Date: Mar. 05, 2018

For the latest version of the awarded   https://www.coretrustseal.org/why-certification/certified-repositories/
CoreTrustSeal for this repository:

Previously Acquired Seals: None

This repository is owned by: **Meertens Instituut - Royal Netherlands Academy of Arts and Sciences**
Joan Muyskenweg 25
1096 CJ Amsterdam
The Netherlands

T +31 20 462 85 00
E douwe.zeldenrust@meertens.knaw.nl
W http://www.meertens.knaw.nl/cms/

Assessment

0. Context

Applicant Entry

Self-assessment statement:

Established in 1926, the Meertens Institute (www.meertens.knaw.nl) is an institute of the Royal Academy of Arts and Sciences, KNAW (www.knaw.nl) and has a recognized reputation for its unique collections and data in the area of Dutch language and culture. The information contained in these are a valuable source for research inside as well as outside the institute and are the foundation of an (inter)national expertise centre for Dutch language variation and everyday culture. It is part of our core mission to manage and disseminate these world renowned collection through state of the art infrastructure facilities. Our Collectieplan Meertens Institute (2013-2018)[1] provides a full background on our collection management strategy and is recommended reading in addition to the information supplied in this assessment. The collection management strategy plan should be read in conjunction with "Crossing boundaries - research plan 2013-2018 Meertens Institute"[2] as the collection management and research objectives of the Meertens Institute are strongly connected. Next to that the Meertens Institute is supporting the necessities for open science, i.e., the persistence and sustainable storage of research data for research reproducibility and data sharing.

(1) Repository Type:

The repository at the Meertens Institute is a domain and subject-based repository focussing on Dutch language variation and cultural phenomena. It houses unique data collections and research data gathered as a result of our past and ongoing research projects, which are part of national and international research infrastructures. In addition, our repository provides the opportunity for individuals and organizations to deposit collections that are considered worthwhile preserving for future generations and research projects. Beyond the domain and subject-based repository it thus also serves as an institutional, national, research project repository and as an archive.

(2) Repository's Designated Community.

Our collections and data have proven to be relevant to several groups within our society. Our designated community can be classified as the (international) research community, higher education students, politics and government, business, professionals in education and government and the interested general public. When one defines valorisation as 'creating impact out of knowledge, regardless of its economic value', then valorisation is considered an important aspect of the Meertens Institute's objectives. In addition to a scientific mission the Meertens Institute also carries a social responsibility. So different groups within our society are encouraged to make use of documented and scientific materials maintained by the Meertens Institute. These different user communities are reflected in our organizational policies.

(3) Level of Curation.

The Meertens Institute offers two levels of curation:

- Basic curation: collections may be deposited by third parties deciding to retain full ownership of the data. In this case they also assume responsibility for data curation activities, such as future migrations to new data formats. The Meertens Institute will perform basic checks on the metadata and data (e.g. completeness, validity and checksums) and may request additional information if deemed relevant for maintenance, discoverability and usability of the collection.

- Enhanced curation: collections for which the Meertens Institute accepted ownership will be considered as part of the regular curation management tasks

(4) Outsource Partners.

All of our data services are delivered by data centres located in the Netherlands. The Royals Academy's I&A department is our main service provider and all of their services are covered by Service Level Agreements (SLA). These services, in turn, are obtained from Vancis which is a provider of high quality ICT services to educational and healthcare organizations and to businesses. Vancis maintains the following certifications: ISO9001, ISO27001, ZSP/GZN, NEN7510, ISAE3402 Type II report. For more information you are kindly referred to: https://vancis.nl/certificeringen. Several copies are stored and backups are maintained for all data. Backups are made (a complete backup is made once and incremental backups are made every day - during the night -) and access from outside to the servers for maintenance is restricted.

All Archival Information Packages in our repository are also replicated to SURFSara through their B2SAFE services for additional safe keeping.

(5) Other Relevant Information.

The Meertens Institute is a certified CLARIN B centre. See http://hdl.handle.net/11372/DOC-106

The archives, research collections and data of the Meertens Institute can be classified into the following categories:

- **Paper collection**: The Meertens Institute maintains over 450 paper collections. Substantial and important collections are, amongst others, questionnaires from the dialectology, ethnology and onomastic departments, the collection 'Brieven aan de Toekomst' ('Letters to the Future'), the pilgrimage and processions collection, Dutch songs and folk tales. These collections are largely housed at the International Institute for Social Science and History, with which the Meertens Institute forms the KNAW Humanities Cluster.

- **Audio collection**: The audio collections at the Meertens Institute maintain a special position within our collections due to the nature, size, information carriers and digitization expertise. They consist largely of fieldwork recordings in the area of language (dialect) variation and ethnology. Digitized materials are made accessible via the Meertens Institute's infrastructure for digital collections (see below). The majority of the material is also archived onto CD's and stored in specialized climate controlled facilities.

- **Digital collections**: The Meertens Institute has been publishing digital research collections for a long time, both for research purposes and for the general public. Publically available databases are accessible via the Meertens Institute's website and repository interface. Also, metadata and data are made available via research infrastructures such as CLARIN. The digital data pertaining to the collections is stored in the Meertens Institute digital archive.

- **Research data**: Research projects need to keep project results available for research reproducibility and sharing. For this reason research data sets are stored and curated in the Meertens Institute digital archive.

[1] http://www.meertens.knaw.nl/cms/images/stories/pdf/algemeen/Collectieplan.pdf

[2] http://depot.knaw.nl/13180/1/OZP2012.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

1. Mission/Scope

Minimum Required Statement of Compliance:

    0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

Documentation and research of Dutch diversity in language and culture has been a central theme for the Meertens Institute from its foundation in 1926. The importance of these aspects are emphasized in our "*Collectieplan Meertens Instituut (2013-2018)*" and "*Crossing boundaries - research plan 2013-2018 Meertens Institute*" respectively describing the collections and research strategy.

The importance of our collections is elaborated in the mission statement of the "*Collectieplan Meertens Instituut (2013-2018)*":

Leading principle for this collection plan is the mission of the Meertens Institute, which can be described as follows: 'The Meertens Institute provides a scientific insight into the nature and organization of Dutch society with regard to language variation and the culture of daily life'. It is the ambition of the Meertens Institute, as one of research institutes of the KNAW, to play a leading role in this field, both in the Netherlands and in the European context. Our collections play a crucial role in the research process and sustainable storage and processing of data will be actively supported. [from Collectieplan Meertens Instituut (2013—2018) p 6]

"… These facilities include the management and dissemination of unique infrastructural facilities and world-renowned collections." It is precisely this objective that the Meertens Institute considers important in the coming period. [from Collectieplan Meertens Instituut (2013—2018) p 5]

The role of our collections in relation to our research agenda is further elaborated in our research plan. As the objectives in "Crossing boundaries - research plan 2013-2018 Meertens Institute" state:

The Meertens Institute has a long tradition in the documentation and research of Dutch language variation, in particular dialects, and in the traditions and rituals we encounter in everyday culture. In the previous decades the Institute's focus has shifted from documentation to research. At present, research is the major component of the Institute's activities, and documentation and other activities are to a large extent made dependent on the scientific efforts. [p.10]
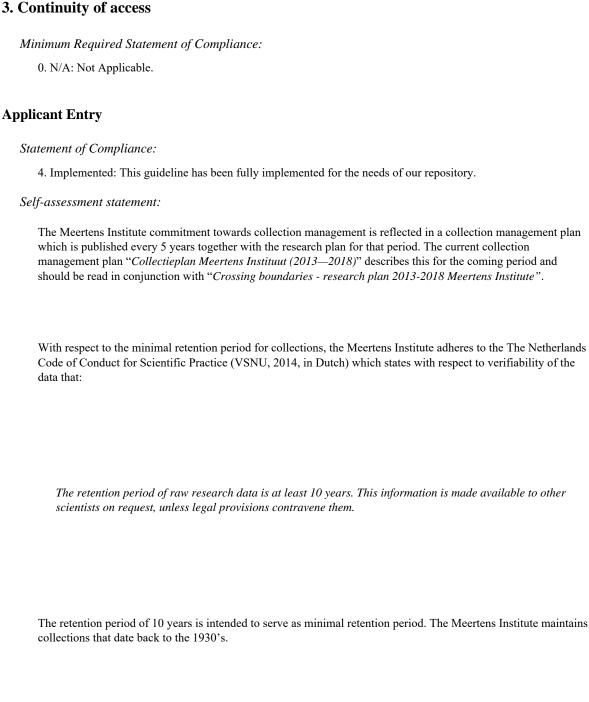
And

The Meertens Institute has systematically built and acquired a large number of collections which provide unique perspectives on our language and culture. These collections and, whenever relevant, new data to be collected, form the starting point for most research projects. [p 10]

The importance of these collections is furthermore recognized in the Royal Academy's strategic agenda "Science and scholarship connect Strategic Agenda for 2016-2020":

A number of these institutes manage sizeable and unique collections that serve as national and international research infrastructures. As national expertise centres, the Academy institutes complement Dutch universities and other knowledge-based organisations. [p 6]

This is further elaborated in the ambitions and goals for 2020:

Aim 5: improve and position the Academy's research institutes as dynamic national organisations that complement Dutch universities [p 8]

- All the institutes will actively pursue valorisation in the broadest sense of the word, i.e. creating value for society;

- access to the institutes' collections will be optimised.

Aim 6: help reinforce national and international networks and research infrastructures [p 9]

- the Academy will function as a catalyst for new networks and infrastructures;

- the Academy's institutes will act as the guardians of national and international infrastructures;

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

2. Licenses

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The Meertens Institute maintains a standard Creative Commons Attribution-NonCommercial 2.0 Netherlands (CC BY-NC 2.0 NL)[1] recommendation policy for its repository holdings. Researchers and third party depositors are allowed to deviate for this policy but are required to submit relevant license information upon deposition. Our deposition agreement[2] further states that that material may only be deposited with the consent of the (co-) beneficiary (s) as intended by applicable law.

[1] https://creativecommons.org/licenses/by-nc/2.0/nl/deed.en

[2] http://www.meertens.knaw.nl/cms/images/documents/LicentieMI.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

3. Continuity of access

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The Meertens Institute commitment towards collection management is reflected in a collection management plan which is published every 5 years together with the research plan for that period. The current collection management plan "*Collectieplan Meertens Instituut (2013—2018)*" describes this for the coming period and should be read in conjunction with "*Crossing boundaries - research plan 2013-2018 Meertens Institute*".

With respect to the minimal retention period for collections, the Meertens Institute adheres to the The Netherlands Code of Conduct for Scientific Practice (VSNU, 2014, in Dutch) which states with respect to verifiability of the data that:

The retention period of raw research data is at least 10 years. This information is made available to other scientists on request, unless legal provisions contravene them.

The retention period of 10 years is intended to serve as minimal retention period. The Meertens Institute maintains collections that date back to the 1930's.

Retention policies regarding collections maintained at the Meertens Institute are dependent upon the research agenda of the Meertens Institute. As stated in the "*Crossing boundaries - research plan 2013-2018 Meertens Institute*":

The collections acquisition [and retention] policy is closely linked to the research agenda of the Meertens Institute. The material on language variation and everyday culture to be collected by the Institute is selected on the basis of its relevance to research purposes, its quality, its relationship to existing collections, and its (digital) state. This policy implies that we generally do not strive for completeness in the fields of Dutch language variation and everyday culture. We aim instead at a high-quality digital set of collections for research purposes, as much as is legally possible available in open access (for the library collection, the open-access criterion is not a leading principle). For the coming period we intend to remain active in developing and acquiring collections directly related to our research topics.[p 68]

Next to archiving digital collections, the Meertens Institute also archives where relevant any resulting data sets of its research projects for research reproducibility and sharing.

To clarify the status of archived data-sets retention time in general, the Meertens Institute management declares that it will take all reasonable measures necessary for storing and providing access to the data sets in the digital archive of the Meertens Institute for which it has accepted responsibility. The data sets in the Meertens Institute archive consists of:

(1) data pertaining to Meertens digital collections

(2) research data sets resulting from Meertens research projects or collaborations that have to remain available for reproducible science or sharing and reuse by others.

(3) data sets deposited by external researchers and organisations, that are accepted as being valuable for the research community and aligned with the Meertens Institute research agenda or collection policy.

If Meertens Institute research interests cease to be aligned with the care for a specific data-set, it will look for another suitable organisation with an interest in maintaining that data-set. If none can be found, Meertens Institute will remain responsible for curating the data.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

4. Confidentiality/Ethics

Minimum Required Statement of Compliance:

    0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The Meertens Institute feels a great responsibility regarding disciplinary norms, ethical and legal issues. All of our data holdings are exclusively stored in Dutch data centers and are subject to Dutch national regulations and laws. For our disciplinary field protection of personal data, for example gathered through questionnaires or language related studies, is of primary importance. Our current procedures concerning confidentiality and ethical issues largely focus on proper management of personal data protection. Cases that are beyond the scope of current procedures will be separately evaluated by the management team of the Meertens Institute.

Processing of personal data in the Netherlands is bound by legal frameworks, more specifically the Wet bescherming persoonsgegevens (Wbp).

The Wbp, which dates back to 2001, is the Dutch version of the 1995 European Personnel Directive (No. 95/46 / EC). Each European Member State has drawn up its own privacy act based on this directive. This means that privacy laws in the different European countries are not completely equal. To align the differences, a European privacy regulation has been developed. This General Data Protection Regulation (AVG) was adopted in May 2016 and will apply from 25 May 2018. From that moment on, a single privacy regime applies throughout the European Union.

The Netherlands already knows the Datalekken Meldplicht Act, which generally knows the same provisions as the AVG [1].

The Wbp contains rules for the processing of personal data, with the emphasis on the automated processing of personal data. The most important provisions of the Wbp on the lawful handling of personal data are summarized as follows [2]:

- Personal data may only be processed in accordance with the law and in a proper and careful manner.

- Personal data may only be collected for specific, pre-explicit and justified purposes. And then only processed further for purposes that are compatible with it.

- The person whose personal data is processed (the person concerned) must be aware of the identity of the organization or person who processes such personal data (the so-called responsible) and the purpose of data processing.

- The data processing must be adequately protected. For special data, such as race, health and belief, additional strict rules apply.

The Wbp has a number of exemptions from reporting personal data. One of these exemptions concerns the processing of personal data by scientific research organizations which are solely for the benefit of research or research carried out by them. These organizations do not need to report these processing if certain conditions are met. To check whether the processing of personal data for a particular investigation must be reported the Code of Conduct for Use of Personal Data in Scientific Research, contains a checklist of five questions which should be answered with an unequivocal "yes" or "no". Every researcher associated with the Meertens Institute and third parties wishing to deposit data to our repository are expected to comply with these rules of conduct, and therefore collection management can check and take the appropriate steps to follow the Code of Conduct. This checklist is listed below:

1. Does the processing take place by an organization for scientific research which is solely for the benefit of research or research carried out by this organization (hereinafter referred to as "organization")?

2. Is processing only necessary for collecting, processing and checking the data for a particular research project?

3. Does the processing only apply to the following personal data:

1. Name, first names, initials, title, gender, date of birth, address, postal code, place of residence, telephone number and similar information required for communication (for example, the e-mail address) and the bank's and bank account number of the person concerned;

2. A non traceable administrative number

3. Other than the above summarized data obtained for a particular investigation?

4. Does the organization provide the personal data to:

    1. Those including third parties who:

        i.   - are in charge of the activities listed above in question 2; Or

        ii. - administer the work listed above in question 2; Or

        iii. - are necessarily involved in the activities listed above in question 2?

2. Others, if:

      i.   - the person has granted unambiguous consent for data processing;

      ii.   - data processing is necessary for compliance with a statutory duty by the scientific research organization;

      iii.   - data processing is necessary because of the vital importance of the data subject (for example an urgent medical need) or

      iv.   - data is processed further for historical, statistical or scientific purposes?

      AND

      v.   Does the organization ensure that data is processed for these specific purposes only?

5. Does the organization remove the personal data referred to in point a of question 3 (excluding sex, place of residence and date of birth) no later than six months after the other data on the person concerned referred to in point c of question 3 has been obtained?

If one or more of these 5 questions is answered with 'yes' then processing of personal data is to be reported to the College Bescherming Persoonsgegevens. While these considerations apply to the use of personal data in any research project, this check list is also applied to all of data depositions at the Meertens Institute. The collection manager at the Meertens Institute will evaluate each submission and may advise depositors on any data protection matters. In cases where data protection is to be considered the management of the Meertens Institute will decide upon the necessity to create specialized repository facilities and additional procedures to safe guard this data, or request the depositor to deposit their data in a manner which is compliant with above regulations. Any data containing personal data that cannot be anonymized and for which no specialized handling procedures are considered necessary will be placed under embargo restrictions in our repository. Anyone wishing to access this material must make a well motivated request to our collections manager. Approval of such a request is subject to approval by the management of the Meertens Institute.

In case of a dispute the Meertens Institute maintains a complaint procedure. If digital provision of certain data infringes any rights or damage to (privacy) interests, these may be expressed directly to the Meertens Institute. The procedure for submitting a complaint are explained on our website (http://www.meertens.knaw.nl/ndb/disclaimer.php). In case of a substantiated complaint, the Meertens Institute will make the material inaccessible and / or remove it from the website or will strive to come to an agreeable solution. In addition, a person may contact the College Bescherming Persoonsgegevens directly with a request for mediation in case personal data protections interests are infringed.

If access to restricted data is requested, the data owner (if not the Meertens Institute) is contacted and has to decide about granting access to the requestor and under what conditions. When in doubt, the management team of the Meertens Institute decides all requests for access to closed data. It currently consists from the institute's scientific director, the administration director, three senior scientists and the IT department leader and is considered to have authority and access to sufficient expertise to handle any disputes. The management team will naturally respect all legal requirements pertaining to handling personal data and IPR restrictions and abide by rulings from relevant legal institutions e.g. College Bescherming Persoonsgegevens.

[1] http://www.justitia.nl/privacy/

[2] https://autoriteitpersoonsgegevens.nl/nl/over-privacy/wetten/wet-bescherming-persoonsgegevens

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

5. Organizational infrastructure

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

As stated in the mission statement, the Meertens Institute has a core missiontowards actively supporting sustainable storage and processing of data of our collections as a crucial role in the research process.

The Meertens Institute maintains a highly qualified Research Collection department and a Technical Development department as part of its standing operations. Each department has been assigned it own funding and is represented in various research national and international projects. Members from these departments have a long history in domain driven research data management initiatives and are, as internationally recognized experts, represented in national and international standing committees, such as the National Coordination Point Research Data Management, the Standing Committee for CLARIN Technical Centres (SCCTC) and Centre Assessment Committee (CAC). Active participation in ongoing national in international research projects is actively pursued thus guaranteeing development of all research data management aspects, such as policymaking and technical infrastructure development. While repository management is considered to be a primary focus point of the research data management life cycle our efforts are directed towards embedding the institutional holdings and procedures in the full landscape of the information management process. This includes Virtual Research Environments making use of repository holdings, (pre) processing and enrichment facilities and direct participation in dedicated research projects.

Currently the Institute employs nine allround ICT experts and one specialised collection manager. The data management and repository tasks are distributed over different experts. The IT staff collaborates with the institute's researchers and collection manager for creating Data Management Plans in project proposals and setting up an efficient research data workflow.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

6. Expert guidance

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The Science Committee of the Meertens Institute (WeCo) is a committee nominated by the KNAW, which advises the Director of the Meertens Institute and the Executive Board of the KNAW on the Institute's work program, in particular regarding the nature, direction and quality of scientific research, and also research collections and data. Members of the WeCo are well respected experts from various research fields and technical disciplines represented at the Meertens Institute. The "*Crossing boundaries - research plan 2013-2018 Meertens Institute*" and "*Collectieplan Meertens Instituut (2013—2018)*" plans have been formulated in close consultation with members of this committee.

The Meertens Institute also maintains an active representation in several CLARIN committees and taskforce initiatives to communicate regularly on new insights and progress of participating CLARIN centres. These centres represent national CLARIN initiatives in various European countries and collaborate towards making (language) resources available for scholars, researchers, students and citizen-scientists from all disciplines.

All digital metadata in our repository is regularly harvested by the CLARIN ERIC and subjected to curation activities on behalf of CLARIN's Virtual Language Observatory (VLO). As part of this curation process we communicate regularly on the issues that relate both to structuring of the metadata content and of the quality of the content itself in an effort to harmonize metadata best practices within our domain. Progress on these efforts is regularly reported as part of CLARIN's Metadata Curation Taskforce.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

7. Data integrity and authenticity

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The Meertens Institute digital repository software (called FLAT) maintains a generic ingest procedure (called Doorkeeper) controlling the ingest procedures of Submission Information Packages (SIPs). This process is implemented as a web service and checks authentication and authorization of each depositor.

The Doorkeeper contains an (extensible) check procedure validating basic requirements of each SIP, such as completeness, schema and file type validity. Each metadata record is expected to comply with the CLARIN's Component Metadata Infrastructure (CMDI) specification (ISO 24622-1). Admissible file types are checked against the list of admissible formats [1]. As a final step of this process the collection manager performs a final, manual quality check before authorizing archiving of the SIP. In case of SIP validation failure, the original depositor is contacted by the collection manager to resolve any outstanding issues before resubmission of the SIP.

FLAT maintains digital objects and CMDI metadata as a single Archival Information Package (AIP) and produces MD5 checksums for each of these. The data repository ensures the authenticity of the digital objects and the metadata by performing automated validity checks at regular intervals. FLAT also versions both digital objects and metadata, which means that it is always possible to retrieve originally deposited objects, and registers an audit trail to reflect which changes have been made by which user. FLAT supports the notion of relations between collections which can be defined by the collection manager. This may be used, for example, to define a relation between paper collection holdings and their digitized counter parts. One concrete example of this are (digitized) questionnaires which are represented in our archive by describing both the paper versions and the results of digitization procedures. These are not considered to be versions of each other but represent separate, however related, data sets.

The base implementation of FLAT is delivered by the Fedora Commons repository system, which meets some requirements out of the box, e.g., ensuring integrity and authenticity of digital objects. The access layer is provided by Islandora modules, thus providing the front end. Fedora Commons has a very flexible digital object model that can be configured to meet specific needs. For FLAT two content models, taken from Islandora, provide the base models: the compound and collection content models. Each CMDI record corresponds to one compound. Members of the compound consist of objects for all the resources described by the record. Compounds can be parts of collections, which themselves can be part of collections again, thereby allowing for nested collection hierarchies of arbitrary depth. Every object, be it a CMDI record or a resource, has a PID in the form of a Handle. Such a handle resolves to the Fedora Commons API call to retrieve the object's main data stream. FLAT stores resource-related data streams outside Fedora Commons in so-called external data streams. This makes it possible to differentiate between different resource usage scenarios by employing different storage facilities for different types of resources

e.g., videos can be available on faster media so they can be streamed out quickly. OAI-PMH is supported by configuring a well-known extension for Fedora Commons based on Proai which is configured to provide CMDI records upon request.

[1] http://www.meertens.knaw.nl/cms/images/stories/data/PreferredFormatsMI.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

8. Appraisal

Minimum Required Statement of Compliance:

> 0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

> Policy considerations concerning appraisal and selection of collections are extensively outlined in the "*Crossing boundaries - research plan 2013-2018 Meertens Institute*" and reflected in the "*Collectieplan Meertens Instituut (2013—2018)*". Primary considerations here are:

> • Relevance for research purposes

> The collections acquisition policy is closely linked to the research agenda of the Meertens Institute. The material on language variation and everyday culture to be collected by the Institute is selected on the basis of its relevance to research purposes, its quality, its relationship to existing collections, and its (digital) state.

> • Related collecting institutions

The Institute has excellent connections with related collecting institutions in The Netherlands and Flanders. Dutch and Flemish Universities, professional organizations such as the Dutch Centre for Intangible Heritage VIE, and museums such as the Netherlands Open Air Museum quite often offer collections of papers, photographs, recordings and/or books to the Meertens Institute. Individual collectors and researchers consider it an honour when the Meertens Institute is prepared to accept their collections or archive as part of the Meertens collection.

- National heritage

In particular domains we intend to strengthen our position as an institution taking responsibility for parts of the Dutch national heritage.

- Research and preservation

Similar domains in which research and the preservation of cultural heritage go hand in hand are, among others, the collections on pilgrimages, state inventories and the audio collections of dialects dating from the middle of the last century, and more recent dialect databases such as GTR and SAND. Acquisition in these cases is not based on research considerations only, but is also determined by the wish to keep these collections relevant, up-to-date and as complete as possible.

Our collection management and technical development departments will provide stewardship for each collection and research data-set which are accepted according to these considerations and deposited into our repository. They will assist in metadata creation and/or evaluation, ensure that proper metadata standards (CMDI) [1] are followed, that appropriate (sustainable) file types [2] are used and provide consultation on license and ownership issues. For digital data this process is supported by automated checks, e.g. file type checking, upon ingest into the FLAT repository. While collection management will strive to make metadata and data suitable for long-term preservation it is recognized that the nature of certain collections will not meet these standards. This may apply to the minimal amount of required metadata but also to deviations form the preferred formats. It is left to collection management and the management of the Meertens Institute to make an informed decision in these cases to either accept, reject or defer the collection to the original depositor. Due to its extensible nature the CMDI specification provides ample possibilities to implement any decisions made in the acceptance process.

[1] https://www.clarin.eu/content/component-metadata

[2] http://www.meertens.knaw.nl/cms/images/stories/data/PreferredFormatsMI.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

9. Documented storage procedures

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The *Datanotitie Meertens Institute* [1] describes the policies and procedures that are to be followed for processes involving digital data deposition, of collections or research data-sets from either internal or external sources.

Data access is governed by our Open Access policy which, in line with KNAW Open Access policies [2], state that research data pending publication may be placed under embargo for a maximum of twelve months. An extension of this period can only be obtained with the permission of the institute's director. Data containing personal information, as indicated by the Wet bescherming persoonsgegevens (Wbp), may be subject to access control policies indefinitely. Access to digital resources is controlled by our access control procedures that are implemented as part of our repository system. The following access control levels are in place and may be applied to individual data files:

- Public open access:

Data files are open for download to the general public, no authentication required.

- Research community open access

Data files are open for download to members of the research community. Authentication is required, either through login via the CLARIN Service Provider Federation (SPF) access or by applying for a personal login account.

- Restricted access

Data files are available for download to specific users. Access is restricted to authenticated and authorized users only. Authorization is determined by the owner of the data file.

The Meertens Institute exclusively employs storage facilities located in Dutch territories, placing all data under Dutch applicable law. Data depositors have the option to maintain ownership of the data files or transfer these to the Meertens Institute. In case a data depositor retains ownership they also assume responsibility for necessary format updates. Data files for which the Meertens Institute has assumed ownership are periodically reviewed with respect to current applicable file formats. All data files are subject to periodic consistency and validation checks.

The Meertens Institute maintains several replicas of its archived data in addition to maintaining daily backups of its servers' content. Replicas are distributed across multiple organizations, the KNAW I&A and the EUDAT B2SAFE service, for additional safety. Full data recovery can thus be established from a number of sources, depending upon the impact level of disaster recovery. The Meertens Institute maintains full AIP replication packages containing metadata, data and access restrictions information and is thus capable of recreating the full repository content should this be applicable. Consistency checks between replicas typically rely on checksums which are calculated on individual metadata and data files or replicated AIPs.

Full risk management strategies are formulated at the institutional level, which includes data disaster recovery procedures. These are formulated in a separate risk assessment plan available from the Meertens Institute.

[1] https://www.meertens.knaw.nl/cms/images/stories/data/DatanotitieMI.pdf

[2] https://www.knaw.nl/en/topics/open-access-and-digital-preservation/open-access/policy?set_language=en

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

10. Preservation plan

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The Meertens Institute maintains a separate collection management plan describing a five-year strategy with respect to collection management. The current version of the collection management plan, *Collectieplan Meertens Instituut 2013-2018*, contains describes the tentative work plans for the two coming years at the time of writing (see p.29). The focal areas of these work plans are dynamic in nature and individual work plans are updated each year. Separate parts of a longer term preservation plan are at different stages of implementation. This includes currently:

1. recommendations for researchers how to archive and share research data that need long-term preservation, using the Meertens Digital Repository

2. a project to move current digital collections into the Meertens Digital Repository (project is in progress)

3. a plan for safe storage of the Meertens digital archive's content, currently using archiving services from ICTS (the current ICT service provider) and the B2SAFE service from the EUDAT CDI provided by SURFsara

4. a proces of regular review of the status of the archived data to determine (new) curation requirements

Activities 1 and 2 are supported by a research data manager reponsible for research data archiving, who liaises with the Meertens collection coordinator and the Meertens ICT support group.

The *Collectieplan Meertens Instituut 2013-2018* also describes acquisitions made by the collection management department. A more passive form of acquisition is whether or not to accept donations and loan collections from third parties. See Appendix 9.2 for the format of a donation agreement. (See: Voorbeeld Schenkingsovereenkomst p.53). Deposition of digital data by any third party will furthermore be recorded in a separate agreement between the depositor and the Meertens Institute [1]. These agreements cover relevant aspects of data deposition such as transfer of custody, the rights of depositor and repository (including the right to disseminate, copy and store). Documentation with respect to accepted formats for the repository are recorded in a separated Accepted Formats document [2]. More information on research data management practices at the Meertens Institute can be found on our website [3].

[1] http://www.meertens.knaw.nl/cms/images/documents/LicentieMI.pdf

[2] http://www.meertens.knaw.nl/cms/images/stories/data/PreferredFormatsMI.pdf

[3] http://www.meertens.knaw.nl/cms/nl/collecties/research-data-management

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

11. Data quality

Minimum Required Statement of Compliance:

> 0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

> The Meertens Institute provides long term archival facilities to researchers working at the institute and to third parties.

> Our researchers are expected to include a data management paragraph to each research proposal:

> *For each new research project, a section should be included in the research plan indicating how to handle data collected by the project. Coordination of data management procedures takes place between the interested researchers, the coordinator of research collections, the Technical Development Department (TO) [1].*

> Deposition of data by third parties is to be coordinated directly with the coordinator of research collections [2].

> For digital resources the repository is capable of handing heterogeneous metadata structures, provided they adhere the CMDI specifications. It thus becomes possible to specify customized metadata profiles that are tailored to the resource or research project at hand. For interoperability reasons, such as common search, it may be mandatory to include a minimal set of fields. The coordinator of research collections will advise on these during the deposition preparation phase.

> Upon ingest validity of the metadata schemas and other relevant properties are automatically checked by our ingest sub system (*Doorkeeper*). The *Doorkeeper* consists of several configurable modules that may be extended to include additional automated check procedures. Before final ingest all AIP's are subject to final approval by the

coordinator of research collections.

After ingest all of our metadata is published through the OAI-PMH end point of the repository and is periodically harvested by the CLARIN community for integration into the Virtual Language Observatory [3]. As part of this process all metadata is subjected to additional metadata quality checks reflecting the expected quality standards of the community. Feedback from these quality checks is regularly communicated back to the Meertens Institute and allows us to align our internal quality standards with the ones formulated by the CLARIN community. We currently do no provide direct feedback options by our end users. Although this would be technically feasible given the flexible and extensible nature of CMDI this has currently not been formulated as part of the requirements for our repository.

[1] https://www.meertens.knaw.nl/cms/images/stories/data/DatanotitieMI.pdf

[2] http://www.meertens.knaw.nl/cms/nl/collecties/research-data-management

[3] https://vlo.clarin.eu

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

12. Workflows

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The standard data deposition workflow is described in the *Collectieplan Meertens Instituut 2013-2018* (Appendix 9.4, pp 58-61) and in the *Datanotitie Meertens Instituut* [1] and the consists of the following steps:

1. **Create metadata**. *Metadata profiles are selected or created matching the type of research data to be archived. Metadata is filled out for each resource. The results are CMDI metadata profiles and descriptions for the research data eligible for archiving. **Stakeholders**: researcher, coordinator of research collections, technical development department **Responsible**: researcher*

2. **Review metadata**. *In this process step, metadata provided is assessed in accordance with the guidelines set by best practices criteria and from the infrastructure. **Stakeholders**: coordinator of research collections **Responsible**: coordinator of research collections*

3. **Specify Authorization**. *In general, the Meertens Institute endorses an open access policy. Where appropriate, this may be waived, for example if IPR or copyright restrictions play a role. The researcher has the possibility to indicate whether he wants to use this facility. Before access to the research data is granted, approval must always be obtained from the collection owner. **Stakeholders**: researcher, coordinator of research collections, technical development department **Responsible**: coordinator of research collections*

4. **Specify licenses**. *If specific licenses applicable to a resource, they will be linked to the resource. Before access is granted, the user licenses must be accepted by the user. **Stakeholders**: researcher, coordinator of research collections, technical development department **Responsible**: coordinator of research collections*

5. **Move data to persistent storage**. *The research data is stored at a location suitable for persistent storage. In addition, the research data is made available in a web accessible manner in accordance with the specified*

access rights. **Stakeholders**: *coordinator of research collections, technical development department* **Responsible**: *technical development department*

6. **Move metadata to persistent storage**. *The metadata provided will be moved to a persistent storage in a manner that will also be accessible via the web.* **Stakeholders**: *coordinator of research collections, technical development department* **Responsible**: *technical development department*

7. **Assign persistent identifiers**. *Persistent identifiers (pids) provide a unique identification of the research data and metadata in a location-independent manner. This means that even data migration or metadata will continue to use the same identifier. Within the Meertens, the Handle system is used and pids are granted in the form of URIs. These can be used, for example, in a browser.* **Stakeholders**: *technical development department* **Responsible**: *technical development department*

8. **Publish metadata via OAI-PMH**.*The descriptive metadata are made available to third parties by a standard protocol (OAI-PMH). In this way, the availability of the data sets available on the Meertens can be disclosed in environments such as the CLARIN-EU Virtual Language Observatory (VLO).* **Stakeholders**: *technical development department* **Responsible**: *technical development department*

The Meertens Institute maintains dedicated contact persons, the coordinator of research collections, for communications researchers and third parties and a research data manager responsible for the research data archiving process. Any data offerings to the repository will be subject to approval in accordance with the policies, guidelines and mission of data holding at the Meertens Institute. While significant portions of the workflow may be automated all information concerning the ingest process is routed via the coordinator of research collections. All data must be reviewed by the coordinator of research collections before being accepted. The depositor decides which data is to be archived and who has access to it, the coordinator research collections decides on the metadata quality and assesses whether the data is in line with the mission of the Meertens Institute. Necessary data format conversion may be carried out by the technical staff present at the institute. This includes anonymizing data records where appropriate.

[1] https://www.meertens.knaw.nl/cms/images/stories/data/DatanotitieMI.pdf

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

13. Data discovery and identification

Minimum Required Statement of Compliance:

> 0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

> The repository maintains metadata records in compliance with the CLARIN B center requirements and guidelines. The repository is thus capable of handling flexible metadata and concept definitions following the CMDI specifications (ISO 24622-1)

> To facilitate resource discovery and search processes our repository is equipped with state-of-the-art search facilities, including facetted search, autocomplete and several display options. Each metadata record and associated resource(s) are identified through persistent identifiers using handles (handle prefix: 10744) which are obtained via the ePIC consortium[1]. Current version of the API is the ePIC API V2.

> The repository is equipped with a fully functional OAI-PMH endpoint capable of serving DCMI and CMDI records. Metadata records are regularly harvested by CLARIN and represented in the CLARIN VLO[2]. The OAI-PMH service is actively monitored using Nagios, providing notifications in case the service becomes unavailable.

> The repository currently does not offer citations in a standardized format, such as MLA or APA. This has not yet been stated as one of the requirements from our user community, but can be added upon request. All necessary technical expertise to make such an extension is available in-house.

[1] http://www.pidconsortium.eu

[2] https://vlo.clarin.eu

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

14. Data reuse

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

In collaboration with our collection manager metadata and concept definitions following the CMDI specifications (ISO 24622-1) are constructed capable of providing context specific metadata descriptions. These metadata profiles are constructed in line with common practice and take common practices, such as DCMI, into account. Partial transformation to DCMI is mandatory due to OAI-PMH requirements for delivering DCMI as well as CMDI metadata formats.

Data formats are preferably delivered or converted to those described in the Preferred Formats document [1]. Format evolution is regularly evaluated by our collections manager in collaboration with our technical staff. Format conversions are carried out through designated projects and are under version control, i.e. converted data files are added to the repository as new versions, maintaining references to previous versions and audit trails.

The repository is furthermore extended with a separate indexing component capable of indexing arbitrary CMDI metadata files and (annotated) content files via the MTAS component [2]. This provides the opportunity to retrieve metadata and data files directly from the repository and make them fully searchable for reuse, e.g. via Virtual Research Environments.

Resubmission of reused or enriched data or metadata is supported by the repository and are subject to the same process as newly deposited data. Our collection manager will evaluate any outstanding legal and rights issues prior to final submission into the repository. Whenever appropriate, version controls mechanisms will be used. In addition, the repository provides mechanisms for creating additional relationships between data sets. As an example, our repository currently maintains records for paper versions of data sets and for their digitized counter parts. Repository records such as these are maintained as separately controlled data sets, each with their own version history. A separate relationship is however maintained to describe the link between the paper versions and the digitized versions.

[1] http://www.meertens.knaw.nl/cms/images/stories/data/PreferredFormatsMI.pdf

[2] https://github.com/meertensinstituut/mtas

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

15. Technical infrastructure

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The repository's architecture is based on the OAIS reference model and distinguishes between the architectural concepts identified in the model.

The repository is capable if accepting SIP information from data producers through a SWORD interface [1] where metadata, resources and authorization information are packaged in the BagIt format [2].

Metadata in the SIP is expected to be compliant to the CMDI metadata specification (ISO 24622-1).

Offered SIPs are checked through a Doorkeeper process. The Doorkeeper provides a pluggable architecture capable of performing all necessary checks and transformations that are relevant prior to final deposition as an AIP into the repository system. These include assignment of persistent identifiers and preparation of access rights. Checks can be extended or replaced, depending upon the application domain in which the repository system is deployed.

Archival storage is implemented using Fedora Commons, current version is 3.8.1, with custom extensions for administration and data management. Front end access layer is provided by Islandora. Version control is in place to manage subsequent updates of metadata and/or resources. Several copies of all archived data are replicated to off site locations, including EUDAT's B2SAFE [3].

Data management is mainly performed through standard Fedora Commons tools supplemented with some customized tooling. Metadata is distributed via OAI-PMH, supporting incrementing and selective harvesting.

Preservation planning. An active technology development plan is pursued in collaboration with participating partners to ensure that the repository software remains up to date and is adjusted to the progressive requirements of

our user community. The archive's software is maintained at a public Github repository [4]. Source code and documentation are thus made available to the community. To facilitate the deployment and testing process dockerized containers are delivered.

All digital objects may be accessed through their persistent identifier. Public information associated with these handles may be accessed through the handle registry. Access to digital objects stored in our repository is subject to access control policies which may be specified for each individual digital object. Access control mechanisms may require the user to authenticate. This may be done via via a separate user registration process or via the CLARIN Service Provider Federation access or via a separate user registration. To facilitate this the repository is part of the CLARIN Service Provider Federation.

The repository's architecture allows for deeper integration with infrastructure development programs. The repository's SWORD interface makes it possible to process real time data deposition from 'live' research data collection systems. Through this set up is is possible to connect current and future data collection systems directly to the archive.

The repository architecture also allows for further extension into the Virtual Research Environment domain, either through the DIP delivery processes or by specifying additional triggers in the Doorkeeper process. This makes it possible to connect the repository directly to more advanced analysis methods, such as linguistic content analysis using our MTAS annotation content search extension [5].

[1] https://github.com/DANS-KNAW/easy-sword2

[2] https://tools.ietf.org/id/draft-kunze-bagit-14.html

[3] https://www.eudat.eu/b2safe

[4] https://github.com/TLA-FLAT

[5] https://github.com/meertensinstituut/mtas

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

16. Security

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The repository system is actively monitored and subject to standard security policies and procedures maintained at the Royal Dutch Academy of Arts and Sciences (KNAW).

Organizational systems are regularly subjected to vulnerability scans and to a yearly audit process (SURFaudit). In terms of SURFaudit recommendations all KNAW organizations are expected to operate at level 3 ('embedded in the organization'). Progress or deviations from this expected level of outcome are monitored on a yearly basis and results and improvement points are communicated in a yearly Security Workplan.

All security incidents are registered, coordinated and handled by the Computer Security Incident Response Team of the KNAW in accordance with the process for handling information security incidents ('Proces voor afhandeling informatiebeveiligingsincident CSIRT (nov 2015)') Technical administrators in collaboration with functional administrators resolve incidents. Each institute within the KNAW maintains an Information Security Officer acting as an intermediary between the central CSIRT group and institute. If an incident is reported the following standard procedures are followed:

1. Identification

2. Damage assessment and control

3. Repair activities

4. Evaluation and reporting

Each step is further subdivided into a specific set of actions related to the incident level.

As of January 1st 2016 each organization is obliged to report data leaks with the Autoriteit Persoonsgegevens [1] if a serious data leak has been discovered.

All of our servers are obtained from I&A, the KNAW's internal supplier of ICT services, and covered via Service Level Agreements. These also cover recuperation procedures in case of an organization wide outage. Risk management procedures are described at the organizational level.

To safeguard data deposited into the repository system multiple copies are maintained at off site locations. All data is furthermore replicated using EUDAT's B2SAFE service in AIP packages that include metadata, data and authorization information. In case of a system outage all data can thus be retrieved from several locations.

Software can be restored from our source code repositories using several options, including a quick recovery using a Docker setup.

[1] https://autoriteitpersoonsgegevens.nl/en

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

17. Comments/feedback

Minimum Required Statement of Compliance:

0. N/A: Not Applicable.

Applicant Entry

Statement of Compliance:

0. N/A: Not Applicable.

Self-assessment statement:

not at this moment

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments: