Implementation of the Data Seal of Approval

The Data Seal of Approval board hereby confirms that the Trusted Digital repository CLARIN.SI Repository 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 21, 2015.

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

Yours sincerely,

The Data Seal of Approval Board

Assessment Information

Guidelines Version:	2014-2017 | July 19, 2013
Guidelines Information Booklet:	DSA-booklet_2014-2017.pdf
All Guidelines Documentation:	Documentation
Repository:	CLARIN.SI Repository
Seal Acquiry Date:	Dec. 21, 2015
For the latest version of the awarded DSA for this repository please visit our website:	http://assessment.datasealofapproval.org/seals/
Previously Acquired Seals:	None
This repository is owned by:	**CLARIN.SI consortium**

T +386 1 477 31 75
E info@clarin.si
W https://www.clarin.si/

Assessment

0. Repository Context

Applicant Entry

Self-assessment statement:

The Jožef Stefan Institute (JSI) is the home of the CLARIN.SI research infrastructure (http://www.clarin.si/), which provides, as one of its services, the CLARIN.SI repository of language resources and tools, https://www.clarin.si/repository/xmlui/.

For the digital repository we use the LINDAT platform, which has been developed by the Institute of Formal and Applied Linguistics, Charles University in Prague, where it is used to host LINDAT-Clarin (Centre for Language Research Infrastructure in the Czech Republic) repository, http://ufal-point.mff.cuni.cz/. The LINDAT-Clarin centre received their Data Seal of Approval on Jan. 10, 2014. The LINDAT platform is purpose-built for archiving and distributing language resources and is available on GitHub at https://github.com/ufal/lindat-dspace.

The CLARIN.SI LINDAT digital repository platform is maintained at the JSI by the Department of Knowledge Technologies and the Laboratory for Artificial Intelligence, with the support of the JSI Networking Infrastructure Centre. We do not outsource the repository nor any connected service.

For the data producer, the repository offers an easy-to-use user interface for data publishing.

For the data consumer, the repository offers faceted searching and browsing of the deposited resources. The submissions are regularly harvested by several other projects using OAI-PMH protocol in order to offer additional ways to find the resources in the repository (cf. Sec. 1).

After the data is submitted to the repository, a curation platform is employed to ensure quality and consistence of the data. It offers the possibility to return the data to the submitter for additional changes. The data and metadata are regularly backed up ensuring robustness and sustainability (cf. Sec. 6).

We follow the standard principles of a high quality digital repository, such as the usage of PID (Persistent Identifiers), authorisation and authentication, and sharing of metadata and data. The system is based on DSpace which follows the OAIS (Open Archival Information System) reference model.

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:

In the curation framework each submission is verified and validated using automatic tools and manually by one of the repository editors.

A submitter is authenticated either through Shibboleth (where we manage the list of IdPs - Identity Providers) or using a local account which is created only after validation. We currently support eduGAIN IdPs, Clarin IdP and the Slovenian national AAI federation.

This ensures a basic level of trust which can be further increased if the submitter is from a list of “well known” submitters. In cases where the submitter is outside of the repository’s well known submitters, special care is taken to validate the input.

We encourage submitters to use open licences, in particular Creative Commons 4.0, but for legacy and other exceptional reasons, we allow data to be associated with other types of public or private licences. It is possible to add new licences to the repository. This policy of maximal openness allows for any party to assess the scientific and scholarly quality of data as much as possible, which is common practice in the area of language resources.

We require a set of metadata attributes providing information about submitted data and the authorship to be filled-in. The submission cannot be completed unless all the required metadata is filled out. The required metadata depend on the type of submitted data, currently language corpus, lexical conceptual resource, language description, or tool. Explanations, examples and suggestions are provided to the submitters in order to obtain high quality metadata.

During the submission process, the submitter must agree to accept our policy that leaves him/her the responsibility for the correctness of the submission, its legal status, accessibility and any related ethical issues. A basic validation

is done by our automatic tools and the submission is carefully checked by the responsible editor. The editor checks the quality of the content and if there are errors or unclarities he/she returns the data to the submitter asking for corrections or additional information. The editor can also ask the research community connected with the repository (in particular the members of the CLARIN.SI consortium) for help.

Each submission is given a PID and we strongly encourage people to use it when citing the repository item (cf. https://www.clarin.si/repository/xmlui/page/cite). We support OAI-PMH, OAI-ORE and several other protocols of metadata and data sharing. We offer different formats, from Dublin Core to CMDI. We are regularly harvested by several institutions which reuse the metadata provided by our repository (e.g., the CLARIN Virtual Language Observatory, VLO: http://www.clarin.eu/vlo/, and by Google Scholar) and are registered to different archive initiatives (e.g., http://www.openarchives.org/Register/BrowseSites). We support browsing and searching in the submission content using our internal search platform.

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:

We show a recommendation to use standard formats when uploading files during the submission process, e.g., for language resources we currently show http://www.clarin.eu/node/2320. Use of standardised formats is encouraged but not enforced. The validity is checked manually by an editor.

If the format is unknown, it must be well documented, and the documentation must be either a part of the submission or the metadata must contain a link to it. If there is a new emerging commonly used format, it can be added to the recommendation.

The description of the submission process for the data producer is given at https://www.clarin.si/repository/xmlui/page/deposit.

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 are submitted to the repository using a graphical user interface (c.f. the description of the submission process at https://www.clarin.si/repository/xmlui/page/deposit). The submission workflow consists of several steps where the data producer must enter mandatory metadata otherwise he/she is not able to proceed to the next step. There are exceptions when submissions can make sense without attached data, in which case the submitter must clearly state the reasons and link to the place where the data can be acquired. In the graphical user interface the metadata format is hidden from the user.

We also support automatic import of metadata (and data) from repositories which support standard sharing protocols, i.e. OAI-PMH, OAI-ORE, and DSpace Archival Information Package.

We employ a set of automatic curation tools which report the quality of metadata and in case of invalid fields (in particular, no longer existing URLs), the author is asked to improve the quality of metadata.

During the submission we require that the user provides at least the following information:

- resource type; currently we allow only 4 types (corpora, lexical conceptual resources, language descriptions, tools)
- title

- list of authors
- issue date
- description
- publisher
- (if applicable) resource language(s) code(s)
- contact person (the responsible person for the submission information) - at least surname and email
- distribution information: access rights, license information, license restrictions, distribution media
- content information: type of media (e.g. text/audio/...), (if applicable) further classification of the resource (e.g. ontology/thesaurus for lexical conceptual resources)
- size information: size in bytes, words, n-grams, etc. (if applicable)

We are able to provide the description of the resource as required by the (minimal) META-SHARE schema (http://www.meta-net.eu/meta-share/metadata-schema) and/or the CMDI (Component MetaData Infrastructure) schema (http://catalog.clarin.eu/ds/ComponentRegistry/rest/registry/profiles/clarin.eu:cr1:p_1349361150622/xsd).

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The national research infrastructure CLARIN.SI has an explicit mission to archive language resources and tools, focusing on the Slovene language, but not excluding other languages. The resources can be deposited by associated researchers as well as researchers who are not affiliated with us. We promote this mission on the repository web page and, as much as possible, in (inter)national conferences. Moreover, the CLARIN.SI consortium comprises all the main institutions that are involved in the production and use of language resources in Slovenia (cf. http://www.clarin.si/info/partners/), and we are providing them with information and guidance about the repository.

The CLARIN.SI mission is supported by integration of the repository into the EU CLARIN infrastructure (CLARIN ERIC, cf. http://www.clarin.eu/ and http://www.clarin.eu/files/centres-CLARIN-ShortGuide.pdf). As part of the CLARIN infrastructure, the repository is included in promotional activities carried out at the national level (CLARIN.SI) as well as at the European level (CLARIN ERIC). We also promote our repository at suitable national conferences e.g., http://nl.ijs.si/isjt14/proceedings-en.html.

The repository implements standard protocols for sharing metadata and data. Public submissions can be easily mirrored. Protected submissions can be mirrored after legal requirements are met.

Link: https://www.clarin.si/repository/xmlui/page/about

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The repository is not a legal entity on its own but is a part of the Jožef Stefan Institute. The repository requires submitters to electronically sign the right to archive the data and that the responsibility of the content lies with them.

After submitting an item, the editors validate the submission before making it public. The repository enables the submitters to restrict the access to their resources at various levels. This include assigning licences to the submissions which must be electronically signed by authenticated users. The signature information is archived.

We currently distinguish three types of contracts:

1) For every deposit, we enter into a standard contract with the submitter, the so-called “Deposition License Agreement”, in which we describe our rights and duties and the submitter acknowledges that they have the right to submit the data and gives us (the repository centre) the right to distribute the data on their behalf.

2) Everyone who downloads data is bound by the licence assigned to the item – in order to download protected data, one has to be authenticated and needs to electronically sign the licence.

3) For submitters, there is a possibility for setting custom licences to items during the submission workflow.

We also offer an option to put an embargo on submissions, meaning that the submission will be completed and its metadata made public, but the data associated with it will become available only after the specified date.

The contracts are available at https://www.clarin.si/repository/xmlui/page/about.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The data storage at CLARIN.SI consists of four different components: 1) submitted datasets, 2) metadata for repository and the datasets, 3) digital repository software and its configuration and 4) the underlying operating system instance with its configuration and logs. Each component has its own data security and backup policy and implementation.

The repository is running in a virtualized OS instance on a small application cluster consisting of two identical servers. Both servers implement their data storage with hardware RAID controllers and RAID-6 volumes.

Submitted datasets are stored in a DSpace repository bitstream store on a network-attached volume on one of the application servers and replicated on the other server.

Repository and dataset metadata is stored in a virtualized PostgreSQL instance inside the virtual machine instance, but regular daily database exports are backed up and replicated.

Repository software and configuration is tracked with GIT version control system and exported from the virtual machine instance.

The active virtual machine instance is cloned and backed up regularly and before each software configuration change or update.

Additional application servers are available through Network infrastructure centre support at JSI in case of failure of both servers or technical issues in our server room.

Complete application server setup, virtual machine instances, database files and exported datasets and DSPace repository bitstreams are backed up with Bacula on a different backup-server residing in a different building. A daily incremental backup is kept for 3 months and a monthly full backup is kept for 1 year.

The policy described above applies to the digital repository, the data and the metadata.

The digital repository software source code is publicly available and is stored in multiple places on multiple machines. Our modifications and configuration changes are stored in a local GIT repository on a different machine in a different building, which follows a similar back-up policy to the one describe above.

All backups follow standardised ways of using MD5 checksums for determining the consistency and we use automatic monitoring tools at various levels.

The preservation policy relates to backup policies above and to the fact that our digital repository uses DSpace software, which defines its preservation as specified at https://wiki.duraspace.org/display/DSPACE/User+FAQ#UserFAQ-HowdoesDSpacepreservedigitalmaterial?

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:

Our repository runs the LINDAT platform, which is, in turn, based on the DSpace repository system. DSpace supports state-of-the-art preservation tools in various forms, from simple replication to standard backup formats and easily manageable collections. The metadata can be exported into various formats suited for long time preservation including self-describing ones like XML. Multilingual support is secured by using Unicode at every level. XML format is used by LINDAT in several instances e.g., when exporting to specific CMDI (Component MetaData Infrastructure) profiles or when archiving AIP (Archival Information Packages). The format validation is done regularly using an external harvesting service (http://validator.oaipmh.com/); moreover, there are several institutions that regularly harvest our repository, which also perform metadata validation.

As mentioned above, we support standard metadata/data sharing protocols (e.g., OAI-PMH, OAI-ORE, DSpace AIP) which allows for duplicating our repository.

We try to minimize the cases in which obsolescence of file formats occur in the near future by encouraging data submitters to use standardised formats. In case proprietary or custom formats are used, we insist on a detailed and exhaustive documentation. Thus it is at least possible to specify and implement data converters, ensuring in this way the long-term preservation of the data.

After submission, the editors validate each submission and check the uploaded files. The editors also allow binary formats if the submitter provides good reasons. Automatic summary of the file formats present in our repository gives us a good overview of what file formats are really used.

Editors have several tools available which help them to validate the submission. Firstly, the submission metadata are listed and can be edited. Then, the standard DSpace curation framework was made available (https://wiki.duraspace.org/display/DSDOC18/Curation+System#CurationSystem-StarterTasks) which includes checks for known/supported file formats, required metadata, link checkers and our internal checks.

Files are checked three times. The file extensions (file format) is checked and marked whether it is supported, known or unknown. The file integrity is checked for several supported and known types. Finally, MD5 checksums are checked to ensure the consistency if submission.

The item lifecycle is described at https://www.clarin.si/repository/xmlui/page/item-lifecycle.

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:

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

Self-assessment statement:

The submission workflow is internally configured in our repository and the submitter goes through each of the required steps. His/her work is finished when he/she submits the item.

The submitted item is then added to the pool of submissions awaiting editors' attention. One of the editors (possibly in consultation with the others) takes the submission and evaluates it in terms of metadata and data. We distinguish between known submitters and the rest, where special care is taken to validate and verify submissions. We have automatic tools helping the editors to verify and validate metadata and the integrity of the submitted data which are performed during the curation step and also automatically at regular time intervals. These tasks involve e.g. checking that all URL mentioned in the metadata do in fact (still) exist, that the required metadata is present and internally consistent, and that the bitstreams (data items) have correct checksums. The editor's job is also to manually check the (meta)data. Special care is devoted to the downloadable files (i.e. data), which is inspected to see if it is consistent with the metadata, i.e. its content description and size and with the published format (e.g. validating XML files).

If the submission is inappropriate for the repository or if it needs to be corrected, the item is rejected, with the reasons and / or requests for modifications emailed to the authors, who are then free withdraw the item or, more typically, to edit and resubmit it.

When the submission is correct in all respects, it is accepted into the repository, at which stage the PID is also assigned to it, and its meta made public, as well as the data, modulo licencing conditions and items under embargo date.

The item can be withdrawn from the repository on the request of the depositor or the author(s), but valid reasons for the withdrawal have to be given. Typically, the PID and metadata of the item will still persist (with a note as to its status as withdrawn) but the item will no longer appear in repository searches or statistics and its metadata will no longer be made available for harvesting.

If a new version of an item is deposited, the old one will still persist in the repository but instead of links to its data, a notice will be displayed that a new version is available, and a link to it is given. Technically, the superseded item will be given the metadata attribute »isreplacedby« which points to the new version. The data of the older version will typically still be reachable, e.g. for cases where experiment reproduction needs access to an exact version of the data.

More details for the depositors are available at: https://www.clarin.si/repository/xmlui/page/deposit

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:

The author of the work will always remain the proprietor. The repository takes care of its copy according to the terms of the licence contract and the terms and conditions for use. The licences with its full text must be signed at the end of each submission.

JSI makes copies for backup purposes which are not publicly available. For the management plan including the technical details cf. Sec. 6.

Our licensing policy is based on the licence selected by the submitter. Each licence can be either free, or a data consumer must digitally sign it, meaning that only authenticated users can access it after submitting a form where they agree to adhere to the licence. We keep track of those signatures and because the authenticated users are real (and traceable) people this process is well defined.

In case of serious (low probability) technical difficulties with main IT infrastructure, the archived data could be imported into another instance of the repository available anywhere.

Crisis management concerning the availability of the digital objects is addressed on a technical level. Since a PID system is used in CLARIN, moving resources from one CLARIN resource centre to another one is possible without affecting the validity of references (e.g. the PID of a resources used in a paper). Our setup consists of virtual machines which may be moved to other CLARIN partners.

Our framework already uses duplication and migration of virtual machines in case of failure on the HW level locally at the JSI. In case of moving to another virtual machine locally the procedure will not have any severe impact on the user experience: a single reboot is needed for the migration, and live migration is planned. In case the machines need to be moved to another location a limited downtime will occur.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The repository enables downloading the data in formats commonly used by the research community. An advanced metadata search utility is provided, as well as a deep search tool for metadata textual content. The public submissions in the repository are being indexed by the CLARIN VLO and Google Scholar. All metadata can be harvested e.g., via the OAI-PMH protocol and free data using the OAI-ORE protocol (unless copyright issues are resolved, we can export all of the data).

Unique persistent identifiers according to the Handle system are provided for each archived object.

Additional links:

- https://www.clarin.si/repository/oai/request?verb=Identify
- https://www.clarin.si/repository/xmlui/

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:

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

Self-assessment statement:

MD5 checksums are calculated for all objects and checked periodically.

Once deposited, files in data sets cannot be changed by submitter but only by administrators for e.g., typos in metadata (however, the importance and feasibility is evaluated on a case-by-case basis). This is also important for the assigned persistent identifiers; they always refer to the same content.

As explained in Sec. 8, if a new version of an item is deposited, the old one will still persist in the repository, but instead of links to its data, a notice is displayed that a new version is available, and a link to it is given.

We are in the implementation phase of a Nagios installation at the Dept. of Knowledge Technologies, which, when set-up, will continuously monitor the availability of files, web and application servers.

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:

After submitting, the only way the data producers can change the (meta)data is to directly contact the editors with their request. As described in Sec. 11, for nontrivial changes a new version of the submission is suggested.

For each change by the editors, the provenance metadata are stored including appropriate log messages.

As described in Sec. 1, the submitters can be only people authorised by well-defined authorities e.g., eduGAIN users using Shibboleth.

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:

For metadata we rely on the group of emerging standards around CMDI (ISO-CD 24622-1). Since our repository is based on DSpace (cf. http://registry.duraspace.org/registry/repository/4697) and uses the workflow supported by the repository's interface, the CLARIN.SI repository meets the requirements of OAIS as described below.

1) Ingest: The Submission Information Packages (SIPs) are received for curating and are assigned to a task pool where our editors can process them. There is a number of pre-configured supported SIP formats (see https://wiki.duraspace.org/display/DSDOC18/Importing+and+Exporting+Content+via+Packages#ImportingandExportingContentviaPack However, the default way is that the ingestion process is done through our web-based interface which hides the implementation details.

2) Archival Storage: After the Ingest step, one of our editors takes charge. Using the web interface, the metadata is checked and can also be updated (added, deleted, or modified). The submitted bitstreams are validated. In short, the editors ensure consistency and quality of each submission. If the editor approves an item, the Archival Information Packages (AIPs) is available.

3) Data Management: This function is executed during the creation of the metadata (descriptive, administrative and structural), as seen in the prior step.

4) Preservation Planning: As described in Sec. 6, we monitor and backup our system. More preservation details are described in Sec. 9. In the repository context, each submission bitstream has MD5 checksums which are regularly checked. There is a list of supported and known formats whose consistency are regularly checked using existing tools (e.g., integrity testing of bzip format is done using bzip -t).

5) Administration: In general, there is no administration with the data producer prior to submitting an item. We are open to all submissions which meet our standards; the data producers must be authenticated which means they

must have academic background or have verified local accounts. A digital contract is signed during the ingestion process. We have developed a specific and robust administration interface including specific detailed reports on the contents of our repository.

6) Access: The available Dissemination Information Package types (https://wiki.duraspace.org/display/DSDOC18/Importing+and+Exporting+Content+via+Packages#ImportingandExportingContentviaPacl query responses and reports are delivered to data consumers. Few submissions require authenticated access which is granted to academic users (through Shibboleth) and locally registered users. Some submissions have their bitstreams available only after a specified date. DSpace allows for searching, locating and description of the information stored. All metadata are publicly available.

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:

An account is necessary to access protected data as described in Sec. 1. When an item is protected, the data consumer must sign the appropriate licence in order to be able to download the data. The metadata themselves are always public. Note that we strongly encourage data providers to use CC licences.

Each submission is clearly marked with its licence and if the licence requires signature, only authenticated users can sign. We rely on the standard academic network which must assure that each authenticated user is a person. We offer local accounts too and in this case we perform the verification manually.

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:

Data providers need to make sure that IPR and personal rights (e.g. mentioning of people in context with personal information or events in texts) are respected in their deposited data. Access to restricted resources is protected via authentication. The licence of each item is clearly visible.

If the licence is not adhered to, we can retrieve the exact dates and specific IDs of people which have accessed the resources.

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 data consumer is made aware of usage restrictions using clear visual indicators. If the dataset is under a licence that requires signing, the data consumer is asked to electronically sign the licence before it is possible to download the dataset (compare e.g. http://hdl.handle.net/11356/1041 to http://hdl.handle.net/11356/1042).

In case of misuse, the only thing that can be practically done is to deny the user further access to the repository and to make the research community and esp. the authors of the resource aware of the misuse. Each signing is stored.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments: