



Implementation of the Data Seal of Approval

The Data Seal of Approval board hereby confirms that the Trusted Digital repository DIGITAL.CSIC 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 14, 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:	DIGITAL.CSIC
Seal Acquiry Date:	Dec. 14, 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:	<ul style="list-style-type: none">Spanish National Research Council CSIC Unidad re Recursos de Informacion Cientifica para la Investigacion - Unit of Information Resources for Research Consejo Superior de Investigaciones Cientificas-Spanish National Research Council C/Joaquin Costa 22 28002 Madrid Spain <p>T +34-91-5681667 E isabel.bernal@bib.csic.es W https://digital.csic.es/</p>

Assessment

0. Repository Context

Applicant Entry

Self-assessment statement:

DIGITAL.CSIC (<https://digital.csic.es/>) was launched in January 2008 with the aim to organize, preserve and provide open access to research outputs by the scientific community of the Spanish National Research Council (CSIC), the organization in which the repository belongs. Its ultimate goal is to maintain a records management system and maximize the impact of CSIC science on the web, while preserving it for future generations. The repository develops, through a dedicated team ("technical office" from now on), under the coordination of the CSIC's Unit of Information Resources for Research (URICI), a Central Services Department that provides knowledge management and research libraries related services to all scientific institutes that belong to CSIC. In addition, DIGITAL.CSIC receives support from the CSIC's Central IT Services Department in terms of systems applications and processes, storage and preservation tasks.

CSIC was founded in 1939 as a public research performing institution composed of a network of research institutes across the country. Although open access cannot be guaranteed for all its research outputs, given CSIC creation year, the repository seeks to enable open access to as much content (journal articles, conference proceedings, books and book chapters, presentations, audiovisual material, data, working papers and so on) as possible and in all instances provide a free bibliographic database. Increasingly over the last few years, DIGITAL.CSIC has paid growing attention to support researchers to comply with funders' open access mandates and to promote open access to data. Before DIGITAL.CSIC the Spanish National Research Council did not have any open access repository and thus the infrastructure was not first fed with data from any other previous institutional platforms. The growth of the repository's content has thus been mostly a result of the so-called Mediated Archiving Service, that is to say, a service put in place by the repository's team in collaboration with the network of CSIC research libraries, to manually upload contents in the platform.

All services offered by DIGITAL.CSIC to CSIC research community are available at its web site, in particular:

-DIGITAL.CSIC policies page <https://digital.csic.es/dc/politicas/> (consulted on July 30, 2015)

Within DIGITAL.CSIC policies there are specific sections devoted to data policy, metadata policy, content policy, deposit policy, preservation policy, privacy policy, formats policy, edition/replacement/elimination, copyright policy and statistics policy. In addition, there is a full page devoted to management and open access to research datasets, <https://digital.csic.es/dc/politicas/politicaDatos.jsp> (consulted on July 30, 2015)

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

-Resources for researchers page: <https://digital.csic.es/dc/recurso/investigadores.jsp> (consulted on July 30, 2015)

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:

Since 2013 DIGITAL.CSIC holds a dedicated space in its web site to address a number of issues around the management and open access dissemination of research data. The page is <https://digital.csic.es/dc/politicas/politicaDatos.jsp> (consulted on July 30, 2015) and it is divided into two main sections: the first section addresses main considerations that researchers should think about before, during and after embarking on data management and dissemination (<https://digital.csic.es/dc/politicas/politicaDatos.jsp#politica1> , consulted on July 30, 2015) whereas the second section covers requirements and recommendations by DIGITAL.CSIC Technical Office for researchers who wish to use the repository as their data recipient (<https://digital.csic.es/dc/politicas/politicaDatos.jsp#> , consulted on July 30, 2015).

In addition, we have produced a template to help researchers in the description of datasets in DIGITAL.CSIC: http://digital.csic.es/bitstream/10261/81323/11/Datasets_DC_plantilla.pdf (consulted on February 3, 2015). The template offers a model that encapsulates information about the data following the Qualified Dublin Core Schema and has taken into account general recommendations by FORCE11 as to how to describe and cite data. The template includes metadata for all pieces of information relevant to data (title, authorship, copyright and license terms, spatial and chronological coverage, date of creation and dissemination, version of dataset, keywords, methodology used to generate data and tools and software requirements to interpret/reuse them, supporting documents and data sources, and other pieces of information). The template is general enough as to be usable to describe data from all 8 scientific research areas within CSIC: Physics, Chemistry, Biology and Biomedicine, Natural Resources, Humanities and Social Sciences, Agrarian Sciences, Materials Science and Food Science. Thus far DIGITAL.CSIC has not dealt with sensitive data that may rise ethical issues and therefore a specific policy for such cases is not included in the repository's data policy. This said, the repository's team provides help and advice for researchers who may want to describe their data by using a specific disciplinary vocabulary or point them to techniques to anonymize their data, for instance.

Data quality is a main activity in the repository and therefore the Mediated Archiving Service offered by the repository's team and the network of CSIC libraries is very much promoted amongst researcher community for them to delegate the upload, description and copyright checks. Through this service, 90% of new uploads take place.

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:

In our data policy we have published our recommendations as regards preferred formats and conversion tools (consulted on July 30, 2015, <https://digital.csic.es/dc/politicas/politicaDatos.jsp#politica3>). On the one hand, we advocate for open formats rather than proprietary ones (so, explicitly in this policy we recommend open formats like ODF, ASCII, XML, MPEG and CSV although we also acknowledge that some proprietary formats such as Microsoft.doc, xls. and ppt. and SPSS are widely used by the researcher community and are thus accepted in the repository) and on the other hand we recall our researchers about the formats supported by DSpace software (<https://digital.csic.es/dc/politicas/#politica9> , consulted on July 30, 2015) so that they are well aware of what services DSpace offers for every single type of format. Equally, we give hints at preferred formats by some research domains (explicitly, for geospatial and audiovisual data) and provide with links to formats recommendations and/or format conversion by other initiatives (explicitly, <http://data-archive.ac.uk/create-manage/format/formats-table> , *Open Refine* <http://openrefine.org/>, Data Exchange Tools and Conversion Utilities <http://data-archive.ac.uk/create-manage/projects/qudex?index=1> and <http://www.w3.org/wiki/ConverterToRdf>). Plus, the repository's team usually uses the free software *Format factory* to change formats of different types of resources, mostly videos, and in particular for web accessibility and visualization purposes.

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:

We have produced a template to assist researchers in the description of their data. This template is available at http://digital.csic.es/bitstream/10261/81323/11/Datasets_DC_plantilla.pdf (accessed on February 9, 2015) and serves as a guideline for them to provide descriptive, structural and administrative metadata. Every time a researcher contacts us (the repository's central office) or its institute's library with the aim to upload a dataset the template is given to her to make sure that the most complete item is uploaded, also metadata-wise. We promote the mediated archiving service to guarantee a complete record of the dataset.

This template has been available from the repository's web site since the summer of 2013 and has been revised a couple of times to add or clarify some metadata. At the same time, at the repository's technical office we have been editing all past datasets whose bibliographic record does not comply with the minimum recommendations of the repository's model template. Equally, for the last 2 years we have delivered at least one annual internal training workshop on data management issues (for instance, <http://digital.csic.es/handle/10261/95802>, accessed on February 9, 2015) and provide online support to researchers and librarians on copyright issues (most often, about licenses options available and how to license datasets) and on journal data policies. As a result of these activities, a growing amount of CSIC librarians and researchers are knowledgeable about best practices in data description, and anyways the repository's main office guide them in the process to describe data for their first time. Equally, if the repository's team identifies items with an incomplete description, such records are edited and improved following the repository's recommendations and the original depositor is informed about such edits. If the edits are minor (i.e., adding of a recommended metadata, no mention is made in the metadata registry as it is not considered a new version of the item as such, simply a correction in its description).

These are typical datasets at DIGITAL.CSIC: <http://digital.csic.es/handle/10261/108758>, <http://digital.csic.es/handle/10261/72264> (both accessed on February 9, 2015).

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

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

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

In the repository's policies there is an explicit section dedicated to digital archiving (<https://digital.csic.es/dc/politicas/#politica8>, consulted on July 30, 2015) and such activity is also part of the mission of the repository as stated in its HOME page: **DIGITAL**.CSIC organizes, preserves and provides open access to CSIC research outputs (accessed on February 9, 2015).

Digital preservation tasks are the responsibility of the CSIC's Central IT Services Department. This is a central department based in main CSIC headquarters in Madrid which provides support to all CSIC centers and institutes. Amongst their preservation activities we can mention the following ones:

- servers monitoring

- servers enhancement

- backups policies

- Mirroring of copies located in geographically distant servers

This department is not an external organization alien to CSIC, rather, it is an own department which is responsible to provide IT support to all centers and institutes that are part of the CSIC.

At a secondary level, the repository's technical office also carries out other activities to guarantee a long term archiving including quality metadata control, format conversion, identification and elimination of corrupt files and broken links, recommended files labelling and software upgrades. However, the repository is dependant on the Central IT Services Department of the CSIC to undertake more complex and systems-related tasks.

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:

- ***Is the repository a legal entity? Please describe its legal/organisational status.*** The repository belongs to the Spanish National Research Council (CSIC), which is the legal entity, and is a not for profit research information platform (that is to say, it does not charge to access the platform and its resources to any user and content available is at the disposal of any user for research/study purposes). The repository develops under the umbrella of the CSIC's Unit of Information Resources for Research, the centralized institutional department that is responsible, amongst other things, for licensing to e-resources, maintaining the institutional union catalog, advancing digitization projects, modernizing the network of CSIC research libraries and promoting open access to science.
- ***Does the repository use standard/template/model contract(s) with data producers? Yes, and such model contract is available publicly under its deposit licence at <https://digital.csic.es/dc/politicas/#politica5> [accessed on July 30, 2015] and also in the deposit template in the intranet of the repository***
- ***Does the repository use standard/template/model contract(s) with data consumers? Right, its data policy at <https://digital.csic.es/dc/politicas/#politica4> [accessed on July 30, 2015] states that free educational/research not for profit usage of resources available in the repository is permitted. It is explicitly forbidden to use DIGITAL.CSIC content for a profit purpose, unless more open user licences in individual items allow so (these user licenses go beyond the repository's general policy and are explicitly chosen by the works authors) .***
- ***Are the repository's conditions or terms of use published? Yes, they are. Please see question above, plus all records carry a copyright notice at the bottom of their page. Such notice reads: "Items in Digital.CSIC are protected by copyright, with all rights reserved, unless otherwise indicated".***

- **Are there measures in place if the conditions are not complied with?** Right. The repository policy includes a section about withdrawal or replacement of items should a copyright infringement be identified or denounced, at <https://digital.csic.es/dc/politicas/#politica6> [accessed on July 30, 2015]. In addition, the repository team often undertakes activities to check whether there are items that breach copyright and use terms. These activities entails, for instance, the review of articles in open access in the repository that have been published in journals whose self-archiving policies are not particularly OA friendly to make sure that their OA availability in the repository is not breaching any copyright transfer agreement or publisher policy.
- **Does the repository ensure knowledge of and compliance with national and international laws? How?** Through annual training workshops on international and national intellectual property rights targeting institutional librarians and authors, for instance <http://digital.csic.es/handle/10261/95802> , <http://digital.csic.es/handle/10261/73242> and <http://digital.csic.es/handle/10261/49947> [accessed on February 12, 2015]. In addition, there is a dedicated space in the repository about copyright issues and compliance terms <http://digital.csic.es/copyright/> [accessed on February 12, 2015] and its technical office provides online and in person consultancy.
- **Are special procedures applied to manage data with disclosure risk?** We havent had any case yet. If content is deemed confidential, our advice is not to disclose it. This means that we do not have in place yet any mechanism whereby the contents of a specific collection or by a specific research group are accessible to a limited community of users. Contents in the repository are either open access for all, metadata-only for all or embargoed for all during a limited period of time.
- **Are data with disclosure risk stored appropriately to limit access?** Not yet, we havent had any case thus far. A different issue pertains items with an embargo access, in which case the repository has a functionality to keep the item invisible for the required period of time, for instance <https://digital.csic.es/handle/10261/116273> [accessed on July 30, 2015]
- **Are data with disclosure risk distributed under appropriate conditions?** Please see answer above.
- **Are procedures in place to review disclosure risk in data and to take the necessary steps to either anonymise files or to provide access in a secure way?** We would like to start tackling these cases in the coming months. We are working on a project to upload immigration related data in the repository using anonimization techniques and the data wont be available until 2016 in DIGITAL.CSIC.
- **Are staff trained in the management of data with disclosure risk?** In progress. We ran a workshop last March to address this issue but there are no technical implementations in the repository yet. Slides of the workshop are available at <http://digital.csic.es/handle/10261/112797> [accessed on June 6, 2015]

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

We understand that data with disclosure risk are not currently held but that you are actively addressing this for the future. Given the intention to address data with disclosure risk and to hold data of that nature in the future, we would expect your next DSA renewal to include details of your approach.

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:

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

Self-assessment statement:

- **Does the repository have a preservation policy? It does at** <https://digital.csic.es/dc/politicas/#politica8> [accessed on July 30, 2015].
- **Does the repository have a strategy for backup / multiple copies? If so, please describe.** The CSIC's Central IT Services Department makes a backup of the contents in the following ways: in situ snapshots for a quick recovery, on-site back up and back up outside of the Central 's CSIC CPD.
- **Are data recovery provisions in place? What are they?** Yes, they are managed through the CSIC's storage equipment services following the Backup policies. These services are managed by the CSIC's Central IT Services Department.
- **Are risk management techniques used to inform the strategy?** At DIGITAL.CSIC we plan to run an audit in the first semester of 2016 by using tools like JHOVE and DRAMBORA. We regularly check corruption of files, broken links, format conversion and use preservation metadata at every record. In addition, the CSIC Central IT Services Department applies the usual techniques inherent to CPD systems.
- **What checks are in place to ensure consistency across archival copies.** CSIC's Central IT Services Department carries out restoration activities in the systems of all institutional databases in a systematic way (at least once a year but in general more often as at every stop made in institutional databases a recovery process is completed and its consistency is verified).

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

- *What levels of security are required and how are these supported?*

- *How is deterioration of storage media handled and monitored?* Every copy is made at the disk thus its functioning is guaranteed (at maintenance level and RAID)

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:

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

Self-assessment statement:

- ***Are there provisions in place to take into account the future obsolescence of file formats? Please describe.***
*We follow the changing landscape of technological development and try to migrate every 2 years to the newest version of the repository's software (DSpace). In fact, on July 22, 2015 we migrated to a recent DSpace version which has entailed performance improvement, more services for end-users and a new look and feel. DSpace is moving growingly into the domain of digital preservation, by providing tools of its own <https://wiki.duraspace.org/display/DSPACE/DigitalPreservationToolsAndStrategies> and more recently, by reaching agreements with external partners such as Archivematica, <https://www.archivematica.org/en/>. From our side, and regarding file format conversion, we often use the free software **Format Factory** to convert files into more stable and/or accessible formats (mostly with audiovisual material)*
- ***Are there provisions in place to ensure long-term data usability? Please describe.*** *We regularly revise records from a bibliographic point of view to make sure that they carry all required metadata for future access and reuse. In addition, we strongly recommend labelling files in an appropriate way (an FAQ at <https://digital.csic.es/dc/faqs/#faq26> [accessed on July 30, 2015] and convert files to open formats, for instance <http://digital.csic.es/handle/10261/108758> [accessed on February 12, 2015]. Equally, we strongly recommend specific preservation metadata (dc.description.provenance) for digitized material (a dedicated guide is available at <http://digital.csic.es/handle/10261/77430> [accessed on February 12, 2015] as well as specific metadata for format information (dc.format) and reuse rights (dc.rights.license).*

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

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

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

Does the repository have procedural documentation for archiving data? Provide references to:

- ***Workflows:*** We have such in the *DIGITAL.CSIC handbook to make deposits*, <http://digital.csic.es/handle/10261/20101> [accessed on February 12, 2015]. This is a step by step guide which explains how to upload data in the repository. It gets updated as required.
- ***Decision-making process for archival data transformations.*** By conducting internal meetings with the Director of the CSIC Department of Information Resources for Research and CSIC Central IT Services Department. We do not have public links to supporting documents that explains this decision-making process. Such is made via emails and/or meetings to consider the feasibility and benefits of such transformations.
- ***Skills of employees.*** Not yet in a public form.
- ***Describe the types of data within the repository and any relevant impact on workflow.*** All accepted categories are listed under the Contents and collections policy, <https://digital.csic.es/dc/politicas/#politica1> [accessed on July 30, 2015]. This policy first stresses that only research-related digital outputs by CSIC community or funded by CSIC may be uploaded in the repository. The second part of the policy focuses on the research outputs types that may be included in the platform, including journal articles, conference proceedings, reports, books and book chapters, patents, datasets, musical compositions, teaching and outreach material, software, audiovisual material, digitised material, maps, and reviews.
- ***Selection process.*** Please see above. Plus we explicitly indicate that data showing personal, religious, political, sport-related and antiscientific opinions are not accepted and strongly recommend open formats over proprietary ones. These remarks are at the bottom of the Contents and collections policy.

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

- **Approach towards data that do not fall within the mission.** An internal meeting with the Director of the Department of Information Resources for Research to decide how best to act. Normally, we ask the authors of the work to send us a written request via email stating the reasons why they would like to upload such works in the institutional repository. In case of doubt, the data are not uploaded.
- **Guarding privacy of subjects, etc.** There is not any reference to guarding of subjects in the Contents policy. There is such a reference under the Privacy policy at <https://digital.csic.es/dc/politicas/#politica10> (accessed on July 30, 2015) which focuses on the commitment of the repository not to publicly disclose personal data about users registered in the platform in order to upload contents and/or activate collections alerts. The repository does not disclose either personal data in its usage statistics module (<https://digital.csic.es/dc/estadisticas/estadisticasUso.jsp>, accessed on July 30, 2015).
- **Clarity to data producers about handling of the data.** Handling of data is indicated in the DIGITAL.CSIC Handbook to make deposits <http://digital.csic.es/handle/10261/20101>, and data management and open access policy at <https://digital.csic.es/dc/politicas/politicaDatos.jsp#> [accessed on July 30, 2015]. Such resources will be updated with information about available and recommended visualization and analysis tools, and anonymization techniques shortly.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

We understand that you are in the implementation phase. We would hope that a future renewal of the DSA would include references to documentation to support internal processes including ingest, data management and archival storage. Maintaining such documentation is valuable to a trusted digital repository, simplifies DSA renewal and, if made publically accessible supports the wider community and increases user confidence in the organisation.

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:

DIGITAL.CSIC repository offers a non exclusive distribution, storage and preservation licence to data producers in order to complete a deposit. The licence clearly states that it requires permission from the data producers to store, distribute and make copies for preservation purposes and also makes an explicit reference to the fact that the data depositors are either the legal copyright holders of the work or have got the required permission to make such upload. The text of the licence is publicly available at <https://digital.csic.es/dc/politicas/#politica5> (accessed July 30, 2015). The licence also grants CSIC the permission to convert the data into other formats should it be necessary for preservation purposes. Such data conversion does not give CSIC the permission to change the content of the works, though.

At the same time, DIGITAL.CSIC strongly recommends data producers to choose an open data license to make explicit references as to the way data may be reused. DIGITAL.CSIC recommends either database compatible **Creative Commons** licences as well as **Open Data** licences and thus these licences are explicitly explained in the policy at <https://digital.csic.es/dc/politicas/politicaDatos.jsp#politica7> (accessed on July 30, 2015), provides training and support to those data producers in need for clarification. This said, the repository recognises that some data may include sensitive information or for their nature they require special or restrictive licences and thus recommends standard protocols and licences under these circumstances. A few links are included in this section for data producers that may want to know more about model restrictive licenses, such as <http://ukdataservice.ac.uk/get-data/how-to-access/conditions.aspx#/tab-end-user-licence> and <http://www.icpsr.umich.edu/icpsrweb/DSDR>. Further, related information is publicly available at recent workshop slides, <http://digital.csic.es/handle/10261/112797> (please see slides 37-48).

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:

DIGITAL.CSIC offers persistent identifiers in the shape of handles and provides support and training concerning recommended formats, conversion tools and data citation standards. The repository web site also links to external resources about recommended formats by research areas and other materials of interest. This information is publicly available at <https://digital.csic.es/dc/politicas/politicaDatos.jsp#politica3> (accessed July 30, 2015) and recent workshop slides, <http://digital.csic.es/handle/10261/112797>. We have several examples whereby the same datasets have been uploaded in open standards and in discipline-specific formats for reuse and analysis, too: <http://digital.csic.es/handle/10261/23139>, <http://digital.csic.es/handle/10261/23051>, and <http://digital.csic.es/handle/10261/22449>. The 3 formats were chosen by the data producers themselves and the repository had no objection as it includes an open format (Text) and also two other formats (NetCDF and binary raster) widely used by the scientific community working with weather data, which promotes data study, replication and reuse.

The repository supports **OAI-PMH** and **SWORD v.2**.

Due to the growing interest of CSIC researchers to upload data in DIGITAL.CSIC and specific DOI requirements in some publishers' open data sharing policies we are considering the functionality to assign DOIs to datasets in the repository briefly.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

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

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

The repository uses checksums as offered by DSpace. Checksums are automatically generated by DSpace when a data submitter is in the process of uploading a new item and the information is stored both in the item record and in the repository internal database permanently. Further, regularly we conduct quality tests to identify attachments in obsolete formats by throwing queries against the repository internal database (for instance, identifying items with no extension termination in the name of their files) or manually checking whether the items that support audiovisual material can be open and visualized. We also raise awareness about the known, supported and unknown formats by Dspace at <https://digital.csic.es/dc/politicas/#politica9> (consulted July 30, 2015).

We have developed a tool to monitor and identify items in the repository that do not comply with metadata standards or are lacking compulsory information like dc.contributor.author, dc.date.issued or dc.title. This home-grown tool validates against the DC terms used in the repository, is in the intranet of the repository and can be only used by DIGITAL.CSIC administrators in order to edit massively wrongly or poorly described items, to enrich items with controlled vocabularies and to identify duplicates, for instance. This validation task takes a relevant share of work by the repository's technical office, as quality data ranks as a priority in the work agenda.

The repository recommends data producers to create a new item for every single new version of a dataset and keep it properly described and documented. New versions do not substitute older versions and we recommend creating new items with the newer versions precisely in order to allow permanent access and citation to older versions of the work. Version information is recommended to be present in the title of the work and authors can resort to a model template prepared by the repository's administrators, http://digital.csic.es/bitstream/10261/81323/11/Datasets_DC_plantilla.pdf. An example of the same work with different versions in the repository: <https://digital.csic.es/handle/10261/48169>, <https://digital.csic.es/handle/10261/72264> and <https://digital.csic.es/handle/10261/104742>.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

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

Minimum Required Statement of Compliance:

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

Applicant Entry

Statement of Compliance:

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

Self-assessment statement:

DIGITAL.CSIC has a version policy whereby it is accepted to upload different versions of the same work and as far as data are concerned, the recommendation is to create a new item for each of them and clearly identify the version in the title metadata or as a second best in the description metadata. The creation of new items with newer versions does not imply the deletion of older versions of the same work, on the contrary all versions are recommended to stay so that users are able to access and cite each of them as required. Such policy and recommendation are explicitly stated in a dedicated section of DIGITAL.CSIC data policy at <https://digital.csic.es/dc/politicas/politicaDatos.jsp#politica6> (visited on July 30, 2015), including a good practice example, <http://digital.csic.es/handle/10261/72264>.

Provenance data are captured and preserved through the usage of dc.description.provenance, a metadata which is visible to repository's users as they log in and which shows the email and name of the depositor, date of upload, and checksum and bytes details of files archived. Equally, DIGITAL.CSIC uses the dc.relation.isreferencedby and dc.relation.isbasedon metadata to link and interrelate items so that it is easy for the end-user to see connections amongst works. For instance, <https://digital.csic.es/handle/10261/113294> (visited on May 6, 2015). We have also given recommendations as to how works in these metadata fields ought to be cited in our deposit handbook (page 18, point 30 at http://digital.csic.es/bitstream/10261/20101/3/DC_manual_archivo.pdf, visited on May 6, 2015).

More than 90% of items in DIGITAL.CSIC have been uploaded through the so-called Mediated Archiving Service whereby authors send basic metadata and files to their CSIC's institutes libraries or the repository's Technical Office. This system does not only assure reaching acceptable metadata quality and authenticity, format and copyright checks as regards files but also serves as an incentive for authors to participate in the open access and open data movement through help from knowledge management professionals. In addition, this system implies the existence of around 60 active depositors per month on average, of whom the repository's Technical Office is well aware and track their performance.

Last but not least, in the repository's Technical Office we run maintenance and format checks in the database a few times per year so as to identify items whose files cannot be opened any more due to corruption or wrong file denomination.

Reviewer Entry

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

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:

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

Self-assessment statement:

The implementation of the OAIS system at an institutional level is at its early stages, with a dedicated and trained staff.

At a project level, we use standards as regards the metadata schema used (Qualified Dublin Core) and best practices to describe, cite, and organise data. For instance, we promote the usage of international/disciplinary controlled vocabularies for data (we keep a controlled vocabulary of academic publishers based on **Library of Congress** and **Spain's National Library Authority Catalogs**, another for scholarly journals based on Library of Congress Authority Files and a third controlled vocabulary for research funders which rests on the **RIOXX**, **VIAF** and **FundRef** vocabularies. These vocabularies can be consulted by the data depositor in the new submission interface in the repository's intranet and a brief explanation about them is available at the DIGITAL.CSIC deposits handbook at <http://digital.csic.es/handle/10261/20101> .

In addition, we have produced a description template which incorporates recommendations by **DataCite** (http://digital.csic.es/bitstream/10261/81323/11/Datasets_DC_plantilla.pdf , visited on May 6, 2015) and citation recommendations by **Force 11** initiative. In addition, we implement best practices as regards formats accessibility in the long run and raise awareness about tools and applications for format conversion and files renaming. Such information can be retrieved at <http://digital.csic.es/dc/politicas/politicaDatos.jsp#politica3> , <http://digital.csic.es/dc/politicas/politicaDatos.jsp#politica5> , <http://digital.csic.es/dc/politicas/politicaDatos.jsp#politica8> (accessed on July 30, 2015) as well as in recent CSIC internal workshops about data management in the repository, <http://digital.csic.es/handle/10261/112797> and <http://digital.csic.es/handle/10261/114141> (visited on May 6, 2015).

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

we would hope that future applications for the DSA would include a mapping of your activities to the OAIS model.

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

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:

DIGITAL.CSIC uses the same End User Licence for data as with the rest of research outputs typologies that are being uploaded into the platform. This is a non compulsory, free, non exclusive distribution licence that allows the repository to disseminate, organise and preserve the items for the future, without requiring any transfer of copyright/exploitation rights. The license states that DIGITAL.CSIC will make best efforts to preserve all digital objects, including the possibility of formats conversion to do so, and should such commitment not be respected, the objects would be returned back to authors or would become part of a greater institutional digital archive. The text of the license is available at <http://digital.csic.es/politicas/#politica5> (visited on May 6, 2015).

The repository promotes the usage of **Creative Commons** and **Open Data Commons** licenses should the data copyright owners wished so, and its Technical Office provides online and offline training on license assignment and legal and reuse implications behind. Likewise, we raise awareness on options and international standard licenses for confidential data and train on how to anonymize data before giving open access. Last DIGITAL.CSIC training effort in this sense was delivered in March 2015 (<http://digital.csic.es/handle/10261/112797> , visited on May 6, 2015) and similar initiatives are already scheduled for 2016 due to the great interest amongst attendants. Last but not least, the upgraded version of the repository went live on July 22, 2015 and amongst new functionalities stands the option to create a **Creative Commons** license on the fly at the moment of data uploading. It goes without saying that this is the choice of the copyright owner and it is not compulsory at all.

Regarding licenses to sensitive data, the repository has not yet in place a technical mechanism to deal with those cases. What the team's repository does is to provide useful resources about international best practices in the management of such data and related restrictive licenses to institutional data producers who are interested in learning more about these options. The DIGITAL.CSIC hands on workshop delivered in March 2015 (<http://digital.csic.es/handle/10261/112797>) directly addressed these international practices widely, as a first step before developing a technical solution in the repository as a new service for institutional users.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments:

Data Seal of Approval Board

W www.datasealofapproval.org

E info@datasealofapproval.org

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:

DIGITAL.CSIC general policy clearly states the standard terms of use of contents, permitted uses (free access and reuse for research/educational purposes) and also commits to taking immediate action should a misuse (copyright infringement) or unsuitable items (not in line with DIGITAL.CSIC contents policy, corrupt files, works accused of plagiarism, etc) be identified (<http://digital.csic.es/dc/politicas/#politica4> and <https://digital.csic.es/dc/politicas/#politica6>, visited on July 30, 2015). As of today we have not faced yet any data upload request where protection of human subjects are implied or in need for anonymization but we have a few cases in the pipeline. In this sense, we make best efforts to raise awareness about related international protocols, best practices, model licenses for data with restricted access and common strategies to anonymize confidential data in a specific section of our data policy (<https://digital.csic.es/dc/politicas/politicaDatos.jsp#politica7>, visited on July 30, 2015) and training material (<http://digital.csic.es/handle/10261/112797> , visited on May 7, 2015).

We strongly recommend data producers to state clearly the reference (preferably its URL) of a reuse license in the metadata of the item they are uploading should they want to grant external users enhanced permissions other than those indicated in the repository's data reuse policy (that is to say, free access and reuse for private research/educational purposes only, with the need to contact the copyright holders directly for enhanced uses). In order to promote this, there is a metadata (dc.rights.license) which allows depositors to indicate that they are assigning a user license that goes beyond the by default DIGITAL.CSIC user license. Beyond that, it is not the mission of the repository to track misuses and copyright breaches of the contents it houses, as the repository is not the legal copyright holder of its digital objects (that is to say, the openly accessible attachments associated to the bibliographic records).

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 standard data license in DIGITAL.CSIC states that contents may be accessed and reused for free for research/educational purposes only. For the rest of reuses data consumers ought to get in touch with the data copyright owners. At the same time, we strongly recommend the use of **Creative Commons** and **Open Data Commons** licenses in our policy and best practices page (<https://digital.csic.es/dc/politicas/politicaDatos.jsp#politica7>, visited on July 30, 2015) and give training and support to researchers to decide on the most appropriate for their data. As a result, most data available in DIGITAL.CSIC go well beyond our standard use licence, with a slight preference for Creative Commons 4.0 International over Open Data Commons (for instance, <http://digital.csic.es/handle/10261/112946> , <http://digital.csic.es/handle/10261/113294> , <http://digital.csic.es/handle/10261/28394> , all visited on May 7, 2015).

Regarding licenses handling for sensitive data please note the enhanced answers provided in previous questions related to the same issue.

Reviewer Entry

Accept or send back to applicant for modification:

Accept

Comments: