

Information recommended to supplement the DATA MANAGEMENT PLAN under a research project financed by the National Science Centre

1.1 How will new data be collected or produced and/or how will existing data be re-used?

Example:

Research data will be acquired from:

- measuring equipment: Philips CM-20 transmission electron microscope, Kuma Diffraction monocrystalline diffractometer and, as defined in the application form, gas chromatograph, in a format matching the data acquisition software. The data will be exported from the relevant software to formats commonly used in this field of research, e.g. to CSV tables, TIFF images, markdown-formatted documents, PDF files; measurement results will be operator-validated;
- repository of microscopic photographs and analytical tests (chromatographic and spectroscopic) at the research centre's disposal and submitted by the partner. The data will be converted into the formats pre-assumed under the research process by taking the potential information loss into account;
- R/Matlab computational software;
- *in silico* experiments using the Molpro/Gaussian software;
- printed and electronic forms used to document the preparation of samples and the course of measurement processes.

1.2 What data (i.e. types, formats, volumes) will be collected or produced in the project?

Example:

High-resolution transmission microscope photographs in files named according to the sample coding convention, in the MRXS format typical of the microscope, and in TIFF files (3,000) with the total size of 2 TB.

Results of chromatographic tests in output files typical of the apparatus used, in the ... format and exported to CSV files; graphs/diagrams saved into the PNG graphical files with the total size of 1 GB, named as per the sample coding convention.

Results of computational experiments: input file and output files in the software's native format; output data converted into open formats: CSV and JSON; visualisations in the PNG, JPG and TIFF graphical files (400 GB in size); file naming convention explicitly indicating the sample number, software, computational method and key parameters.

Data extracted from forms: tables associated with individual forms documenting the course of the research process saved into the XLS format or into linked CSV files (30 MB).

2.1 What metadata and documentation (for example methodology or data collection and way of organising data) will accompany data?

The project will make use of the services rendered under the E-SCIENCE.PL platform, maintained by the Wrocław Centre for Networking and Supercomputing [*Wrocławskie Centrum Sieciowo-Superkomputerowe*] of the Wrocław University of Science and Technology. Files will be stored in a shared network resource accessible via the CIFS, NFS, SFTP and HTTPS protocols.

The pre-assumed file and catalogue naming convention will be consequential to individual stages of the research process. The documentation will be developed within a closed MediaWiki mechanism dedicated for the project and supporting version control. An indexing service covering both the content and metadata of files, embedded into the e-science platform, as well as the MediaWiki index will facilitate information mining. Additional metadata being outcomes of the experiment, yet not contained in the output files, will be saved into sidecar files and linked with the MediaWiki mechanisms. The metadata profiles to be used will match the preferences of the experiment at hand. What has been assumed with regard to certain metadata is the use of thesauruses such as e.g. UMLS, AgroVoc, GEMET or other commonly available ones, as well as registers such as GeoNames and Wikidata.

The ontologies of schema.org and DublinCore will be used to render data sets available, while the proprietary azonOntology solution is to be used to describe metadata serialized in formats such as JSON-LD, RDF/XML, or Turtle. Tabular data will be converted according to the *W3C Tabular Data and Metadata on the Web* recommendations into the JSON-LD format.

2.2 What data quality control measures will be used?

The documentation of the data acquisition procedures and the data validation will be placed on the project's wiki. With regard to measuring procedures, the validation component is assumed to be handled by the operator after or during the measurements, or it is to be based on documented guidelines. It is assumed that data processed in computational software will be subject to validation procedures at the preprocessing stage and the postprocessing stage. The data must be secured against unauthorised access and modification using an access control mechanism conforming with the NFSv4 ACL model, as well as snapshot and backup copies.

3.1 How will data and metadata be stored and backed up during the research process?

Data will be stored using the E-DRIVE service provided via the E-SCIENCE.PL platform. It provides sufficient technical and physical safety measures, enabling the entire infrastructure to be kept in server rooms of the Wrocław Centre for Networking and Supercomputing [WCSS] subject to continuous monitoring and safeguarding, ensuring redundant power supply, cooling, network connectivity, etc. using disk arrays, service server farms, a flexible access control model, encrypted connections as well as snapshot and backup copies produced in accordance with WCSS's policy. Each project is provided with the storage capacity of 4 TB, with an option to expand. An in-house archiving service rendered by WCSS is intended for purposes of long-term storage, while the AZON [Atlas Zasobów Otwartej Nauki]

platform (zasobynauki.pl) will allow for data to be rendered available under the *opendata* model. Individual protocols of the E-DRIVE service, i.e. SFTP, CIFS, HTTPS, or the Sync&Share service will be used to ensure data acquisition and access from portable computers and mobile devices. The authorisation model for granting rights to individual disk spaces is based on a group mechanism which maps the actual project teams. The groups are managed by the project manager or the group leader using the E-SCIENCE.PL platform's mechanisms.

3.2 How will data security and protection of sensitive data be taken care of during the research?

Data access will be controlled in line with pre-defined access control lists (ACL) based on a breakdown into project groups or individual users. Prior to being granted access rights to a service, each user is subject to an authentication procedure using a login and a password conforming with the E-SCIENCE.PL platform's policy or the X509v3 certificate. The user can autonomously recover data from snapshot copies. In the event of a storage system failure, data will be recovered from backup copies maintained by WCSS. The operations of authentication and granting of access rights via the HTTPS protocol are logged in the Central Authentication Service (CAS) provided by WCSS.

4.1 If personal data are processed, how will compliance with legislation on personal data on data security be ensured?

In 2017, the University adopted its "Personal Data Security Policy", and ever since it has been updating the relevant rules ensuring compliance of data processing with the applicable regulations, and appointed a Data Protection Supervisor (DPS) with whom all agreements concerning data processing are consulted. The University has been keeping records of all processing activities and activity categories, as well as records of persons holding data processing authorisations. I have been holding such an authorisation since On ... (state the approximate date), I attended a training workshop (or e-learning workshop) on personal data handling.

In fulfilment of my information obligations, I use the clauses provided on the websites of the Wrocław University of Science and Technology and the National Science Centre. When involved in the project implementation, I will make use of, and the following are the safeguards applied there:
.....

e.g. ... **services provided via the E-SCIENCE.PL platform and the E-DRIVE service.** The safeguards applied there allow for data access through encrypted transmission channels or via VPN, data sharing and mining; access control; rendering data publicly available with time-set limitations; version restoring; updating and data synchronization.

PLUS (where applicable to any project assuming the research to be conducted using personal data):

I had consulted all new and significantly altered activities pertaining to data processing with the DPS before the actual processing commenced. I conducted the risk assessment on (date), which I documented by means of I have been monitoring the relevant risks on an ongoing basis, and with regard to the risk recognised as high, I have assessed the consequences of data processing for the rights and freedoms of the persons whose data are processed.

4.2 How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

The proprietor of the data produced, including databases and intellectual property rights, will be the Applicant, while copyrights to individual pieces of scientific work may be vested in the authors for purposes of publication, possibly open. Data or results of data processing will be rendered available, unless there are substantial reasons to protect them. The procedure which applies in such cases is defined in the intellectual property management regulations of the Wrocław University of Science and Technology.

Select as appropriate:

No third party data are planned to be used subject to specific limitations *or*

Data from *[data source name]* will not be rendered available *or*

Data from *[data source name]* will be rendered available *[identify method or licences]*.

5.1 How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

For purposes of the project as well as its partners and the subcontractors involved, data will be rendered available by means of the sharing mechanisms which the E-DRIVE service features. The data intended for open access will be placed in the AZON (zasobynauki.pl) repository along with the relevant descriptions and in the right file formats, as required by the service. The AZON platform provides mechanisms for rendering resources available in line with the Linked Open Data paradigm, which particularly applies to permanent identifiers, different metadata profiles, format conversion, the search feature and the presentation layer compatible with WCAG2.0, etc. The chosen open license type will depend on the nature of the resources to be rendered available, with preference given to the CC-BY-SA licence.

5.2 How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?

Data will be rendered available by means of the AZON (zasobynauki.pl) repository, where the FAIR Data principles and the Frictionless Data approach have been deployed. All project data will also be stored using the archiving service provided by WCSS, which will be maintained by WCSS for a period of at least 10 years.

5.3 What methods or software tools will be needed to access and use the data?

Placing the data on the E-SCIENCE.PL platform will provide the contractors with access to the data via network protocols (CIFS, NFS, SFTP, HTTP), and depending on the activities being performed, enable integration on the level of the operating system, software and computational or cloud resources.

5.4 How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI) to each data set be ensured?

A unique identifier will be generated for a data set being placed in the AZON (zasobynauki.pl) repository. The repository policy guarantees that each resource is assigned a unique, permanent and dereferable identifier, and it provides for integration with the DOI as well as with Handle.Net.

6.1 Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

The person responsible for the implementation of the data management policy is the project manager. The technical support and content-related assistance in the scope of data management is provided by the Wrocław Centre for Networking and Supercomputing.

6.2 What resources (for example financial and time) will be dedicated to data management and ensuring the data will be FAIR (Findable, Accessible, Interoperable, Reusable)?

The IT infrastructure required to manage data over the entire project life cycle as well as to render them available is ensured by the Wrocław Centre for Networking and Supercomputing as a part of the services they provide. WCSS also provides the technical assistance and expertise in the scope of data acquisition and storage, preparation of metadata profiles, format conversion, long-term archiving, interoperability assurance and processing of data, in such a manner as to enable them to be used again in a service stated under other costs.