

## PLAN ZARZĄDZANIA DANYMI

dr Justyna Boniecka Katedra Genetyki Wydział Nauk Biologicznych i Weterynaryjnych Uniwersytet Mikołaja Kopernika

### 1. Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych (Data description and collection or re-use of existing data)

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

New data will be generated as a result of the experiments designed and proposed in this project. Data will be produced in digital form by specialized software tools (i.e. Mass Lynx, Empower, LightCycler 480, Image Lab Biorad, Dualex Scientific and others) that are used to control scientific equipment. Raw instrumental data will be collected and processed by a person responsible for a particular task using appropriate software tools to obtain user-friendly numeric data that will be used/reused by our team.

Well described and reproducible data generated in the course of this project, either in a raw form (if useful for others) or mostly processed, will be stored in a certified repository for reuse by others, under specified conditions (Creative Commons [CC] licences).

We also planned to reuse majorly data available on plant RNA/DNA sequences from National Center for Biotechnology Information (NCBI) databases as well as other data present there needed to conduct the project. This data will be used in accordance with NCBI restrictions on molecular data usage (<https://www.ncbi.nlm.nih.gov/home/about/policies/>).

#### What data (for example the kinds, formats, and volumes) will be collected or produced?

Data will be collected mainly, but not exclusively, as:

- raw chromatograms and mass spectra (as .dat files or as records in equipment-specific Oracle databases),
- raw qRT-PCR data as .ixo objects,
- western-blot images (as .tif, .jpeg, .png, .bmp and other graphic formats and densitograms),
- manually entered numeric and text records stored as tabularized database in .xlsx format or text files in .docx.
- Raw instrumental data will be processed by a person responsible for the particular task using appropriate software tools. This will help us to obtain user-friendly numeric data, and when possible expressed in SI units, among others:
  - concentrations, relative concentrations and relative amounts
  - absolute and relative expressions of mRNA and proteins,
  - nucleotide sequences,
  - fluorescence intensity and relative fluorescence intensities,
  - percentages values.

Existing data will be reused after a manual or automatized data import from digital and analogue data sources. In the case of re-use of data from analogue sources (e.g. data in a notebook, an old thesis or a publication) they will be transformed to digital form, and stored as tabularized database in .xlsx format.

If possible, data files will be converted to open formats (e.g.: .odt, .ods), prior introduction into a repository. At the moment, it is hard to estimate data volumes, especially considering the fact that data will also be selected.

## **2. Dokumentacja i jakość danych (Documentation and data quality)**

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

Raw instrumental data will be organized in a software-specific manner and stored in folders assigned to each task, designated only for the data obtained in this project. Raw data will be accompanied by technical metadata (e.g. acquisition date, sample number) and whenever possible, if software allows for additional data introduction after data acquisition, by additional metadata like sample full name, sample experiment date and details, authors, and others. Processed, manually entered and reused data will be organized in tabular or text datasets accompanied by the above-mentioned metadata.

Data intended for a repository will be accompanied by metadata in standard Dublin Core Metadata (or others): Title, Keywords, Subject, Description, Creator, Date, Type, Format, Source, Language, Relation, Contributor, Grant number or others if necessary. In that case, the introduced metadata will depend on metadata options proposed in a repository.

### **What data quality control measures will be used?**

Data will be obtained by a qualified person responsible for a particular task using calibrated tools/equipment and validated analytical methods. When possible and needed, at least three biological and two technical replicates will be performed. Data will be released, when meet validation criteria (usually: relative standard deviation of technical replicates less than 15%). To avoid mistypes, manually entered data will be cross-checked by a second person involved in the project. Software tools included in a database and statistical software packages will be used to check data consistency and if they are in appropriate formats. In depth explanation on samples used, methodologies and data processing applied in the course of the study that lead to the production of the data being uploaded in a repository will be given in README files.

## **3. Przechowywanie i tworzenie kopii zapasowych podczas badań (Storage and backup during the research process)**

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

Data will be stored in soft copies at storage devices of the equipment used for data collection (usually the personal computer of both the person responsible for a particular task and the project manager and a workstation if applicable). Raw data will be backed up by a person responsible for the particular task in regular manner, not rarer than monthly, using operating system-included tools, on independent data storage device (e.g. external hard drives), including network share when possible. External hard discs will be stored at the institute and closed in a locker. Processed data will be backed up in a similar manner, as soon as possible after creation, and it will be a responsibility of the person that obtained the data. All network shares will be backed up on a safe mass storage device in a distinct location. Some data (only created by us) will also be stored and saved in a repository.

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

We will not produce and store sensitive data.

**4. Wymogi prawne, kodeks postępowania (Legal requirements, codes of conduct)**

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

Not applicable

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

Copyrights to our data will belong to us, as creators, in accordance with the grant-maker's recommendations, some of them (readable for recipients) will be disseminated in the open access model under CC licenses. In the case of reused, external data, it will be disseminated in accordance with restrictions of the data disseminating party.

Appropriate form of intellectual property rights protection, in agreement with employer and funder rules, will be chosen, depending also on the option of CC licenses given by a particular repository. When possible, we will encourage applying Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA 4.0) license schema.

The results will be presented in an open access repository following the FAIR rules, published and shared. The data will be disseminated in a way to make it findable, accessible, interoperable and reusable.

Before the project starts, between all creators an agreement on the rules of dissemination of data obtained in the project will be signed. In the document it will also be stated that project manager will be responsible for raw and processed data sharing upon request and overall data management.

**5. Udostępnianie i długotrwałe przechowywanie danych (Data sharing and long-term preservation)**

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

Some of the raw data will be shared only on reasonable personal request directly from project manager. Raw data, if in our opinion useful and needed, will be disseminated maximum one year after the release of the official publication of the main findings.

We will publish, as much as possible, in well-established peer reviewed open access journals with impact factor and, if required by the journal, include raw and/or user friendly processed datasets as supplementary data. If not, data will be shared in chosen repositories and databases appropriate to data format (e.g. Repository for Open Data – repOD <https://repod.pon.edu.pl/> or one focused on dissemination of data on a chosen topic). Majority of the data disseminated via repository will be prior

selected, organized, and prepared in a user friendly manner (FAIR rules). Such data will be timely released, generally no later than the release through publication of the main findings.

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

For long term, raw and processed data will be preserved on hard drives and if possible on university servers and in scientific repositories. The raw data will be archived for minimum ten years and on request available minimum ten years after data release. Repositories-shared data will be available as long as the repository will serve.

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

To access raw data-on-request specialized software may be necessary. When possible, data will be exported to formats readable by open-source software. Processed data will be accessible using standard office applications or open standards (e.g.: .odt, .ods, .pdf,). Depending on technical possibilities in the near future, if needed, appropriate actions towards format migration will be undertaken.

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

Data will be deposited and shared only in repositories offering the unique and persistent identifier Digital Object Identifier (DOI), or others specific to a particular repository, for free.

**6. Zadania związane z zarządzaniem danymi oraz zasoby (Data management responsibilities and resources)**

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

Project manager will be responsible for overall data management, including preparation of final files and their upload into a chosen repository along with metadata. Each person responsible for a particular task will also be responsible for proper data management in the scope of their task.

**What resources (for example financial and time) will be dedicated to data management and ensuring the data will be FAIR?**

The proper data management, ensuring that the data will be FAIR, will be a responsibility of each person in the project. The payment to people responsible for data management will be incurred for the costs of the project during its implementation and was already included while calculating their salaries. After completion of the project, such work will be carried out by the project manager who is a statutory employee of the University. The costs of deposition of data in specialized repositories may be incurred for the costs of the project when necessary. Hard drives dedicated to project-related data backup will be purchased from the project's funding.