

## 1. Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych

Sposób pozyskiwania i opracowywania nowych danych i/lub ponownego wykorzystania dostępnych danych

Do przygotowania przez każdego z badaczy:

Zrobić to można najprościej odpowiadając na pytania:

Jakie rodzaje danych będą wygenerowane w trakcie realizacji projektu?

Mogą to być wyniki:

- a) analiz spektrofotometrycznych, spektrofluorymetrycznych, luminescencyjnych
- b) wyniki sekwencjonowania metodą Sanger, NGS
- c) wyniki analiz elektroforetycznych czy też mikroskopowych
- d) wyniki sekwencjonowania białek metodą MALDI-TOF
- e) opis eksperymentu to też dane

Jakie dostępne dane mogą być wykorzystane do realizacji projektu:

- sekwencje białek zdeponowane w bazie uniprot, PDB
- struktury krystalograficzne białek zdeponowane w bazie uniprot, PDB
- sekwencje DNA zdeponowane w bazach .....

### PRZYKŁAD 1:

*Podczas badań realizowanych w ramach zgłoszonego projektu zostaną wygenerowane następujące dane:*

- *wyniki pomiarowe spektrofotometrycznych*
- *wyniki z analiz NGS*
- *wyniki analiz elektroforetycznych*
- *dane będące opisami procedur eksperymentalnych*

### PRZYKŁAD 2:

*The data will be collected by experienced researchers within **PODAJ NAZWY KATEDR**. The collected data will be obtained as the result of **PODAJ NAZWY EKSPERYMENTÓW** experiments. Therefore the data divided into these **LICZBA** categories will be stored. All the data will be stored on computer discs/cloud drives and specially designated lab books to allow data analysis. All the data will be divided into texts, figures, schemes, tables, raw data and elaborated sections. The equipment applied for the study will be calibrated to allow the reproducibility.*

### Pozyskiwane lub opracowywane dane (np. rodzaj, format, ilość)

Na podstawie informacji zawartych wcześniej musimy opisać dane, które uzyskamy podczas realizacji eksperymentów. W celu ułatwienia pracy przedstawiono opis danych najczęściej wykorzystywanych w badaniach z nauk biochemicznych:

Rodzaj danych	Format	Rozmiar pojedynczego pliku
Dane spektrofotometryczne	XLS lub CSV lub format danych odczytany ze sprzętu	max 1 Mb na plik
Dane spektrofluorymetryczne	XLS lub CSV lub format danych odczytany ze sprzętu	max 1 Mb na plik
Sekwencje DNA metodą Sangera	scf, ab1	max 1 Mb na plik
Sekwencjonowanie metodą NGS	Fastq	około 15 Gb na jeden run/chip
Elektroforegramy	sgd, ptl lub format danych odczytany ze sprzętu	10 Mb/plik
Opisy eksperymentów	docx, txt, RTF	2 Mb na plik
Sekwencje DNA, białek pobrane z baz danych	wpisujemy rozszerzenie pliku pobranego z bazy	Wpisujemy sumaryczną wielkość pliku

#### Przykład:

- wyniki pomiarowe spektrofotometrycznych otrzymane podczas pomiarów z wykorzystaniem spektrofotometry nanodrop zostaną zapisane w postaci plików tekstowych \*.txt nie więcej niż 50 MB
- wyniki z analiz NGS zostaną zapisane w pliku o formacie \*.fastq nie więcej niż 15 GB
- wyniki analiz elektroforetycznych zostaną zapisane w formacie w plikach o formacie \*.sgd nie więcej niż 10 MB
- dane będące opisami procedur eksperymentalnych zostaną zapisane w formacie dokumentów tekstowych \*.docx nie więcej niż 100 MB

#### PRZYKŁAD 2:

All the laboratory procedures, including reliable documentation of research, will comply with standard operating procedures (SOPs) elaborated in recent years at PUMS labs. The NAZWA DANYCH data will be stored in excel or PODAJ FORMAT format. The PODAJ TYP DANYCH data will be elaborated with the use of NAZWA OPROGRAMOWANIA software. The lab data and elaborated reports will be stored

*in the form of Microsoft Office doc or docx files as well as Acrobat Reader pdf files. The amount of produced data will not exceed 1 TB.*

## **2. Dokumentacja i jakość danych Metadane i dokumenty (np. metodologia lub pozyskiwanie danych oraz sposób porządkowania danych) towarzyszące danym**

**Tu odpowiadamy na pytania.**

- **Jak nasze dane „zbierzemy” ?**
- **Jak nasze dane opiszemy? Ważny jest opis danych w taki sposób by osoba postronna zrozumiała co mamy za dane.**

**Pamiętajmy o:**

- **plikach opisujących tryb akwizycji danych (powiększenia mikroskopów, systemów obrazowania, rozdzielczość etc.)**
- **wieloplatformowości danych (systemy UNIX, WINDOWS, LINUX, MACOS) plik opisu danych musi dać się otworzyć na każdym systemie operacyjnym więc najlepiej zrobić to w pliku txt, pdf**

**Przykład 1:**

*Dane generowane podczas realizacji projektu będą pozyskane z wykorzystaniem oprogramowania obsługującego sprzęt pomiarowy.*

*- dane spektrofotometryczne zostaną otrzymane podczas pomiarów:*

*a) stężeń kwasów nukleinowych metodą spektrofotometryczną,*

*b) pomiarów absorbancji podczas testu ELONA*

*- sekwencje oligonukleotydów zostaną ustalone za pomocą sekwenatora NGS (usługa zewnętrzna)*

*- wyniki analiz elektroforetycznych zostaną zapisane w postaci plików graficznych podczas wizualizacji żeli poliakrylamidowych*

*Dane będą zapisane w odpowiednich folderach, które dodatkowo zostaną opisane w pliku readme.txt, który będzie zawierał odnośniki do dziennika laboratoryjnego (opis eksperymentu) oraz wykaz oprogramowania niezbędnego do podglądu danych.*

**Przykład 2:**

*The data obtained within the study will be stored in lab books containing particular enumeration, the description of the reagents, solvents as well as conditions applied for the study. The proper and reproducible procedure will be attached to each performed experiment. Every 3 months the*

*contractors will prepare Quarterly reports in doc/docx or pdf files, which will be collected by PI in folders, and file names will be clearly assigned. All samples will be collected in a specially prepared space in refrigerator of cupboard. All files produced within the experiments will contain initial of contractor\_number of lab book\_page in lab book.*

### **Stosowane środki kontroli jakości danych**

All chemical experiments will be optimized and scaled up to ensure reproducibility. All biological experiments will be repeated in 2 to 4 replicates at least 2-3 times in order to ensure proper statistical correctness. The raw data will be available for contractors and NCN experts to allow analysis and control of the results.

### **3. Przechowywanie i tworzenie kopii zapasowych podczas badań**

#### **Przechowywanie i tworzenie kopii zapasowych danych i metadanych podczas badań**

All the data, including the unprocessed raw data produced during the realization of the project will be stored on offline password-protected hard discs and *computers/cloud drives* within **PODAJ LICZBĘ WSPÓŁPRACUJĄCYCH KATEDR** Departments participating in the project. Also, the backups will be performed regularly every week. Reports and data for publications will be stored in a particular local network repository at PUMS. Data and metadata will be stored on a server instance specially dedicated for this purpose together with the database system. Currently, the university's resources allow the use of up to 200 GB of disk capacity for databases created for scientific projects. Backups are created in two stages. The first stage is a daily copy of the databases themselves to a dedicated separate matrix. The second stage is a copy of the entire virtual server images to a tape library. All contractors will be specially trained in information security practice management.

#### **Sposób zapewnienia bezpieczeństwa danych oraz ochrony danych wrażliwych podczas badań**

The security of data is connected with the special regulations at the PUMS. Following the regulations, all computers need to be password protected and equipped with regularly updated antivirus software. The IT PUMS Department provides secure internet and internal network connection. File server resources will be created and made available to the project team. In the event of accidental deletion of files, IT department employees will be able to recover them from an automatically created copy. IT staff will take care of technical data, including backup and recovery. The IT department will also contact the project manager to determine the frequency of creation and the

number of backups. Access to data will be given to authorized employees. An appropriate signed authorization will be issued for each person. This group will include research workers involved in the project, system/server administrators.

#### **4. Wymogi prawne, kodeks postępowania Sposób zapewnienia zgodności z przepisami dotyczącymi danych osobowych i bezpieczeństwa danych w przypadku przetwarzania danych osobowych**

PUMS established intellectual property rules, which are defined, and employees and students are informed with all issues related to intellectual property management. The PUMS has and applies various policies to increase the security of processing, storing and sending information, including:

- Procedure for managing incidents related to information security
- Procedure for analyzing the risk of critical IT systems failure
- Procedure for archiving and restoring IT systems
- Work safety policy at computer workstations
- Procedure for granting, changing and withdrawing privileges in IT systems
- Clean desk policy
- Clean screen policy
- Information security policy
- The policy of using private equipment for business purposes
- Policy on the destruction of data carriers
- Computer delivery procedure
- IT system management manual

#### **Sposób zarządzania innymi kwestiami prawnymi, np. prawami własności intelektualnej lub własnością.**

During the project, legal issues will be handled by the Department of Science, Project Management and International Cooperation in cooperation with the Law Firm/Patent Attorneys providing services to the University. Intellectual property and proprietary rights are regulated by legal acts in force at UMP.

#### **5. Udostępnianie i długoterwale przechowywanie danych Sposób i termin udostępnienia danych. Ewentualne ograniczenia w udostępnianiu danych lub przyczyny embarga**

All the data obtained within the project will be carefully collected and published or presented during meetings and scientific conferences, making them fully accessible to the broad scientific and

nonscientific audience interested in the performed study. The formats of all files will be standard and will allow access without any special software.

Experimental data relevant to intellectual property will be made available to the public once the procedures for applying for protection of potential inventions have been completed **(ważne, by to zapisać ; nie możemy ujawnić nic przed wysłaniem zgłoszenia patentowego do UPRP)**, which will precede the publication of results. After publication of the results, data will be available for a minimum of 5 years from the date of completion of the project.

### **Sposób wyboru danych przeznaczonych do przechowania oraz miejsce długotrwałego przechowywania danych (np. repozytorium lub archiwum danych)**

The data received during the project will be preprocessed (removing poor quality data) and then grouped into packages that will be uploaded to the repository.

List of repositories below

<https://www.teamscopeapp.com/blog/6-repositories-to-share-your-research-data>

### **Metody lub narzędzia programowe umożliwiające dostęp do danych i korzystanie z danych**

**Ważne, by tak udostępnić dane, aby jak najwięcej osób mogło z nich skorzystać bez posiadania specjalistycznego oprogramowania, dlatego warto wszystko przekonwertować (tam gdzie to możliwe) do pików typu CSV, TXT które można otworzyć na większości systemów operacyjnych.**

#### **Przykład:**

*In order to share key results of the project, data will be made available in formats supported by open source software as well as original measurement files. Each data package will have a readme.txt file describing how to handle the data.*

### **Sposób zapewniający stosowanie unikalnego i trwałego identyfikatora (np. cyfrowego identyfikatora obiektu (DOI)) dla każdego zestawu danych**

The selected repository for data sharing guarantees a unique DOI identifier.

### **6. Zadania związane z zarządzaniem danymi oraz zasoby**

**Osoba (np. funkcja, stanowisko i instytucja) odpowiedzialna za zarządzanie danymi (np. data steward)**

**Zależnie od wielkości projektu:**

- *W przypadku projektów , w których jest mniej niż 3 badaczy, osobą odpowiedzialną za zarządzanie danymi może być Kierownik Projektu:*

The PI will be responsible for the data management related to the project at PUMS. The PI will decide on the sharing of the data within contractors group as well as presentation of the results and data to the broader audience.

- *W przypadku większych projektów nie może to być Kierownik Projektu. W takiej sytuacji należy wskazać, że zadanie będzie delegowane na jednego z wykonawców grantu:*

Data management tasks will be delegated to one of the project staff/contractors. This employee will be properly trained by the training providers of future data stewards.

**Środki (np. finansowe i czasowe) przeznaczone do zarządzania danymi i zapewnienia możliwości odnalezienia, dostępu, interoperacyjności i ponownego wykorzystania danych**

The data will be stored on the hard drive of PI, which will be available at the Department during and after the realization of the project. The IT division at PUMS guarantees reliable assistance and management as well as access to the data. The long-term safe storage of all data will be performed at least up to five years after the realization of the project.