Toruń 17 październik 2019

Przykładowy "Plan zarządzania danymi medycznymi" – model do grantów zagranicznych opracowany przez Daniela Gackowskiego z Collegium Medicum UMK i Bożenę Bednarek-Michalską z Biblioteki Uniwersyteckiej.

1. 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 collected during experiments designed by team members. Data will be produced in digital form by specialize software tools (i.e. Mass Lynx, Empower, Diva, CytExpert, Gentools, LightCycler 480, BaseSpace, MinKNOW and others) used to control scientific equipment.
Data will be mainly, but not exclusively, collected as:

- raw chromatograms and mass spectra (as .dat files or as a records in equipment-specific Oracle databases),
- raw cytometric data (.fcs in at least 3.0 format),
- raw RT-PCR data as .ixo objects,
- western-blot images (as .tif, .jpeg, .bmp and other graphic formats and densitograms),
- manually entered numeric and text records stored as tabularized database in . xlsx format,
- raw genomic data.

Raw instrumental data will be processed by qualified personnel (data steward) using appropriate software tools to obtain human-friendly numeric data, when possible expressed in SI units, among others:

- concentrations, relative concentrations and relative amounts,
- cell counts, percentages, fluorescence intensity and relative fluorescence intensities,
- absolute and relative expressions of mRNA and proteins,
- variant ratios,
- mutations types and frequencies,
- variant allele frequency,
- nucleosides sequences.

Existing data will be re-used after a manual or automatized data import from digital and analog datasources. In the case of re-use of data from analog sources (i.e. paper medical records) they will transformed to digital form, and stored as tabularized database in .xlsx format.

2. Documentation and data quality
What metadata and documentation (for example methodology or data collection and way of organising data) will accompany data? What data quality control measures will be used?

Raw instrumental data will be organize in a software-specific manner. When possible raw data will be accompanied by metadata like sample code, acquisition date, experiment date and details, data, authors, and others. Processed, manually entered and reused data will be organized in tabularized datasets accompanied by metadata: *Title, Subject, Description, Creator, Publisher, Contributor, Date, Type, Format, identifier, Source, Language, Relation, Coverage and Rights, Grant number* or others if it will be necessary.

Data will be obtained by qualified personnel using validated analytical methods. When possible and appropriate, technical and biological replicates will be performed. Data will be released, when meet validation criteria (usually: RSD - relative standard deviation - of technical replicates less than 15%). Manually entered data will be cross-checked by second person to avoid mistypes. Software tools included in database and statistical software packages will be used to check data consistency and appropriate format.

3. Storage and backup during the research process

How will data and metadata be stored and backed up during the research process? How will data security and protection of sensitive data be taken care of during the research?

Data will be stored in electronic form at mass storage devices of equipment used for data collection (usually personal computer or workstation). Raw data will backed up in regular manner, not rare than one per moths, using operating system included tools on independent data storage device, preferably on network share when possible. Raw data backup will be administered by data servant. Processed data will be backed up immediately after creation on network share. Processed data backup will be administered by data servant. All network shares will be backed up on safe mass storage device in distinct location automatically and administered by IT.

Access to all devices used to data generation and storage will be controlled at user level using standard MS Windows domain protocols. Sensitive data will be collected in a "as-less-as-necessary" manner. All sensitive data will be password-protected in addition to user-level access. Passwords will be distributed using other communication channels than used for data transfer.

4. Legal requirements, codes of conduct

If personal data are processed, how will compliance with legislation on personal data and on data security be ensured? How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

At the Nicolaus Copernicus University in Toruń, personal data is protected at all levels, (in accordance with Polish law). There are procedures that will be used in the project. In case of doubt, there is Data Protection Inspector who provides advice and assistance in this regard (https://www.umk.pl/uczelnia/administracja/?name=Samodzielne-Stanowisko-Pracy---Inspektor-Ochrony-Danych).

Copyrights to the data will belong to their creators, in accordance with the grant-maker's recommendations, some of them (readable for recipients) will be disseminated in the open access model under the license. All data authors will be free to choose appropriate form of intellectual property rights protection, in agreement with employer funder rules. When possible, we will strongly encourage applying CC-BY-SA 4.0 license schema (https://creativecommons.org/licenses/by-sa/4.0/deed.pl).

The results will be presented in open repository following the FAIR rules (to make data findable, accessible, interoperable and reusable https://www.go-fair.org/fair-principles/), published and shared without disclosing personal data. We will also keep the medical secret in the publications concerning the tests carried out. All personal data, genetic data, health information relevant for the study (diagnosis, treatment history and outcome) and samples will be pseudonymized before processing by clinical coordinator under supervision of data steward in accordance to General Data Protection Regulation (GDPR). The identification key will be handled by data servant and stored in the Clinic of Hematology NCU. The key will not be available to the laboratory investigators.

UoO and Akershus University Hospital (our partners) has already established well working procedures. We believe the joint research project is the best way to understood and adapt already working good practices. University of Oslo: Data processing and storage will adhere to the GDPR directive. No data production or data processing will commence until the work and data protection plan is approved by the Data Protection Office at Akershus University Hospital. All patients will receive a study-Id number and patient data will be handled pseudonymised. The PI will not have access to patient information beyond what is required for the study (diagnosis, treatment history and outcome). The data will be processed and stored at a secure server

at the University of Oslo service for secure data storage (https://www.uio.no/english/services/it/research/sensitive-data/index.html).

A data processing agreement will be signed between the collaborating institutions and sharing of data between the study participants will be done according to this. The data will be deleted according to the approval given by the regional ethics committee and the data protection officer at Akershus University Hospital.

5. Data sharing and long-term preservation

How and when will data be shared ? Are there possible restrictions to data sharing or embargo reasons? How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)? What methods or software tools will be needed to access and use the data? How will the application of a unique and persistent identifier (such us a Digital Object Identifier (DOI)) to each data set be ensured?

Raw data will be shared on reasonable personal request directly from data author or via data steward. Raw data should be released maximum half year after the release through publication of the main findings. Processed data will be shared in chosen open repositories and databases appropriate to data format (i.e. EMBL Nucleotide Sequence Database https://www.ebi.ac.uk/ena, repOD https://repod.pon.edu.pl/, Norwegian Centre for Research Data http://www.nsd.uib.no/nsd/english/index.html). Data will be timely released, generally no later than the release through publication of the main findings. Data will be deposited and shared only in repositories offering free unique Digital Object Identifier (DOI) and European standards for data descriptions (https://www.ebi.ac.uk/ena/standards-and-policies).

Data-on-request will be available minimum ten years after data release. Data shared via repositories will be available as long as repository will serve. 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.

6. Data management responsibilities and resources

Who (for example role, position, and institution) will be responsible for data management (i.e the data steward)? What resources (for example financial and time) will be dedicated to data management and ensuring the data will be FAIR?

The data steward responsible for data management will be employed during the implementation of the project. The remuneration of the data steward will be incurred for the costs of the project during its implementation. After completion of the project, such work will be carried out by statutory employees of the University. The costs of deposition of data in specialized repositories may be incurred for the costs of the project when necessary. Mass-storage devices dedicated to project-related data backup will be purchased when necessary.