Data Management Plan

1 Data collection and documentation

1.1 What data will you collect, observe, generate or re-use?
Questions you might want to consider:
- What type, format and volume of data will you collect, observe, generate or reuse?
- Which existing data (yours or third-party) will you reuse?
Briefly describe the data you will collect, observe or generate. Also mention any existing data that will be (re)used. The descriptions should include the type, format and content of each dataset. Furthermore, provide an estimation of the volume of the generated data sets. (This relates to the FAIR Data Principles F2, I3, R1 & R1.2)

We will record electrophysiological signals (microelectrode arrays, patch clamp), images of experimental preparations (cardiac cell cultures, markers for the analysis of the deformation of the culture substrates) and outputs of computer simulations with mathematical models. According to previous experience, we expect to generate 50-100 GB of experimental data and 50-100 GB of simulated data per year.
Electrophysiological recordings from microelectrode arrays will be archived in binary format, e.g. using the IEEE 754 standard (used and directly readable by most processors and systems); for third-party re-use, we will provide utility programs converting these data into a decimal text format (e.g., comma-separated values, CSV). Electrophysiological recordings from patch clamp experiments will be recorded using the software manufacturer format and exported/archived as CSV files for readability by other software/users. Images will be stored using the standard TIF format. Videos will be stored using the standard MPG/MP4 or AVI formats.
Corresponding metadata will be archived in binary as well as text formats. To facilitate data mining, we will develop a SQL database (MySQL or MariaDB, with the support of the IT engineers of our Department) into which existing and new data and metadata will be imported.
Simulation outputs will be stored/archived using the IEEE 754 format and utilities provided for conversion into decimal text files. Corresponding metadata will be archived in binary form as well as in text format (CSV) and imported into the SQL database.
All raw data will be archived.
For the re-use of third-party data (e.g. segmented morphological data of intercalated discs) we will directly collaborate with the owners of such data to define suitable formats.

1.2 How will the data be collected, observed or generated? Questions you might want to consider:
- What standards, methodologies or quality assurance processes will you use?
- How will you organize your files and handle versioning?
Explain how the data will be collected, observed or generated. Describe how you plan to control and document the consistency and quality of the collected data: calibration processes, repeated measurements, data recording standards, usage of controlled vocabularies, data entry validation, data peer review, etc. Discuss how the data management will be handled during the project, mentioning for example naming conventions, version control and folder structures. (This relates to the FAIR Data Principle R1)

Data will be collected and documented according to predefined and already existing guidelines and check-lists specifying what information (metadata) needs to be collected in conjunction to every experiment (e.g., day/time, name of investigator, aim of experiment, replication of experiment in a series, basic characteristics of the experimental preparation, experimental conditions (e.g., pharmacologic agents and concentrations), list and coordinates of electrodes, pacing protocols,. analog processing settings, etc.). Some of this information is already generated automatically as a metafile by the acquisition softwares that are in use. Care will be taken to use predefined and consistent vocabularies throughout the entire projects.

Calibration data and data from technical tests will be acquired/archived using the same formats and metadata structure.
Data/metadata will be reviewed by the principal investigator in a short timeframe after each experiment/simulation to check/validate the completeness of metadata entries and to check the consistency of the data.
Raw data and metadata will always be stored in a central place (departmental server) in a SQL database (see 1.1). The data structure will allow the access of the data by user, date, preparation number as well as other parameters. The metadata will be linked to the raw data files. The database has appropriate interfaces with analysis software and to the web server. The database will have a journaling capability, which gives information about the versioning. All changes and deletions to the data will be stored.

1.3 What documentation and metadata will you provide with the data?
Questions you might want to consider:
- What information is required for users (computer or human) to read and interpret the data in the future?
- How will you generate this documentation?
- What community standards (if any) will be used to annotate the (meta)data?
Describe all types of documentation (README files, metadata, etc.) you will provide to help secondary users to understand and reuse your data.
Metadata should at least include basic details allowing other users (computer or human) to find the data. This includes at least a name and a persistent identifier for each file, the name of the person who collected or contributed to the data, the date of collection and the conditions to access the data. Furthermore, the documentation may include details on the methodology used, information about the per-formed processing and analytical steps, variable definitions, references to vocabularies used, as well as units of measurement. Wherever possible, the documentation should follow existing community standards and guidelines. Explain how you will prepare and share this information. (This relates to the FAIR Data Principles I1, I2, I3, R1, R1.2 & R1.3)

For experimental recordings, the following minimal information will be generated (either automatically or acquired manually) as metadata/metafiles accompanying the raw data:
Automatically generated:
- date/time of the experiment
- for microelectrode array recordings: number of recorded channels, list of channels, coordinates of corresponding electrodes, stimulation protocol (position of stimulation electrodes, pacing times, stimulus intensity and duration)
- for patch clamp recordings: all parameters defining the voltage-clamp or current-clamp protocol
- sampling rate
- analog filter settings
Acquired manually:
- name of the investigator
- persistent identifier / file name
- parameters related to the preparation (e.g., type of culture/cells used, age)
- indications regarding the experimental protocol (e.g., temperature, bath medium, pharmacologic agents used and concentrations, etc.)
- for experiments using stretchable microelectrode arrays: parameters related to stretch/strain
For computer simulations, the source codes or scripts will be archived with the generated data.
For all the data that will be deposited on a public repository, the metadata will be converted and made available as ASCII text files (e.g., CSV) along with "README" text files explaining in detail the structure and format of the files to guarantee that a human or a human-assisted machine can read/interpret the data.

2 Ethics, legal and security issues

2.1 How will ethical issues be addressed and handled?
Questions you might want to consider:
- What is the relevant protection standard for your data? Are you bound by a confidentiality agreement?
- Do you have the necessary permission to obtain, process, preserve and share the data? Have the people whose data you are using been informed or did they give their consent?
- What methods will you use to ensure the protection of personal or other sensitive data?
Ethical issues in research projects demand for an adaptation of re-search data management practices, e.g. how data is stored, who can access/reuse the data and how long the data is stored. Methods to manage ethical concerns may include: anonymization of data; gain approval by ethics committees; formal consent agreements. You should outline that all ethical issues in your project have been identified, including the corresponding measures in data management. (This relates to the FAIR Data Principle A1)

There are no ethical, legal or security issues. The projects do not involve any personal data, sensitive data or data subjected to a confidentiality agreement.

2.2 How will data access and security be managed?
Questions you might want to consider:
- What are the main concerns regarding data security, what are the levels of risk and what measures are in place to handle security risks?
- How will you regulate data access rights/permissions to ensure the security of the data?
- How will personal or other sensitive data be handled to ensure safe data storage and -transfer?
If you work with personal or other sensitive data you should outline the security measures in order to protect the data. Please list formal standards which will be adopted in your study. An example is ISO 27001-Information security management. Furthermore, describe the main processes or facilities for storage and processing of personal or other sensitive data. (This relates to the FAIR Data Principle A1)

The projects do not involve any personal or otherwise sensitive data. There are no security issues or associated risks.
The data will be accessible to a web server. However, data access will be regulated by an individual user registration process (userID and password). Only read access will be granted to users. The database itself is open source (MySQL or MariaDB). Because it is based on SQL, data export is straightforward.
State-of-the-art security standards will be implemented on the servers. Our Department is protected by the University firewall, and our servers will be managed by the IT engineers of our Department. We will have full control over our data.

2.3 How will you handle copyright and Intellectual Property Rights issues?
Questions you might want to consider:
- Who will be the owner of the data?
- Which licenses will be applied to the data?
- What restrictions apply to the reuse of third-party data?
Outline the owners of the copyright and Intellectual Property Right (IPR) of all data that will be collected and generated, including the licence(s). For consortia, an IPR ownership agreement might be necessary. You should comply with relevant funder, institutional, departmental or group policies on copyright or IPR. Furthermore, clarify what permissions are required should third-party data be re-used. (This relates to the FAIR Data Principles I3 & R1.1)

According to regulations, the data are owned by the University of Bern.

It is not planned nor expected that the generated data will be subjected to any confidentiality agreement or restrictive intellectual property rights. The data will be shared under the Creative Commons Attribution-Noncommercial-Share-alike scheme (CC BY-NC-SA).
If third-party data is used (e.g. segmented morphological data of intercalated discs) we will clarify directly with the owners of such data what licenses apply and whether permissions are required.

3 Data storage and preservation

3.1 How will your data be stored and backed-up during the research?
Questions you might want to consider:
- What are your storage capacity and where will the data be stored?
- What are the back-up procedures?
Please mention what the needs are in terms of data storage and where the data will be stored. Please consider that data storage on laptops or hard drives, for example, is risky. Storage through IT teams is safer. If external services are asked for, it is important that this does not conflict with the policy of each entity involved in the project, especially concerning the issue of sensitive data. Please specify your back-up procedure (frequency of updates, responsibilities, automatic/manual process, security measures, etc.)

Immediately after each experiment or simulation, the generated data and metadata will be copied via a LAN connection on our departmental servers, on which capacity in the terabyte range is attributed to the research group. The data on the department servers is automatically backed up twice on tape every night. One tape is then stored in a fire- and waterproof cabinet in the Department and the second tape is storage under similar conditions in a separate building of the University.
This procedure is routine in our Department and the management of the servers and the backups is assured by the IT engineers employed by our Department. As we do not generate sensitive data, there are no concerns regarding this specific aspect.

3.2 What is your data preservation plan?
Questions you might want to consider:
- What procedures would be used to select data to be preserved?
- What file formats will be used for preservation?
Please specify which data will be retained, shared and archived after the completion of the project and the corresponding data selection procedure (e.g. long-term value, potential value for re-use, obligations to destroy some data, etc.). Please outline a long-term preservation plan for the datasets beyond the lifetime of the project. In particular, comment on the choice of file formats and the use of community standards. (This relates to the FAIR Data Principles F2 & R1.3)

Since the volume of our data remains in a tractable range, there is no necessity to delete any data, and all data from completed projects will be preserved/archived on dedicated drives/volumes on our departmental archive. There is no legal obligation to destroy data.
Since there is, to our knowledge, no well-established common standard for the archiving of the type of data we plan to generate (in particular microelectrode array recordings), data and metadata (see 1.3) will be archived in generic formats (e.g., transformed to ASCII text or including README documentation text files describing the structure of the data) which should permit to reanalyze/reinterpret the experiments also in a far future.
Data of highest relevance and potential value for re-use (e.g., data from successful experiments on which publications are based) will be transferred to a repository as described under 4.1.

4. Data sharing and reuse

4.1 How and where will the data be shared?
Questions you might want to consider
- On which repository do you plan to share your data?
- How will potential users find out about your data?
Consider how and on which repository the data will be made available. The methods applied to data sharing will depend on several factors such as the type, size, complexity and sensitivity of data. Please also consider how the reuse of your data will be valued and acknowledged by other researchers. (This relates to the FAIR Data Principles F1, F3, F4, A1, A1.1, A1.2 & A2)

In collaboration with other principal investigators and the IT staff (2 persons) at our Department, it is planned to establish a departmental repository (backup procedures as in 3.1). The data/metadata will be stored in the form of a SQL database and our IT staff will obtain unique digital object identifiers (DOI) to the datasets.
[Comment Sarah Jones, DCC, and Research Data Management Team, UB Bern: To set up an own repository a huge amount of resources and time is required. Furthermore, the repository should be compliant to the FAIR Data Principles, which means, that a long term running and preservation strategy has to be in place. It is therefore recommended to use established repositories whenever possible.]

Because there are no widely established standard formats, we will work in conjunction with the IT staff to define such formats, keeping in mind that format conversion should remain feasible until e.g., consensual international standards/guidelines are defined.
The tight collaboration with the IT engineers from our Department will allow more flexibility in the management of the repository and tailored adjustments of data/metadata file formats and contents, as well as search engines that can be integrated in the repository.

In parallel, we will consider following existing repositories, and contact the respective managers of these repositories to inquire about required formats and the possibilities/conditions to store our data:
PhysioNet (www.physionet.org): a repository for physiological signals, supported by the US National Institute of Biomedical Imaging and Bioengineering
Dryad (www.datadryad.org): general repository

For mathematical models, we will consider following well-curated repositories:
CellML Project (www.cellml.org), developed by the Auckland Bioengineering Institute (University of Auckland)
Physiome Project (www.physiome.org): hosted by the Department of Bioengineering, University of Washington

The repository and DOI of the data will be mentioned in the publications. The data will be shared under the Creative Commons Attribution-Noncommercial-Share-alike scheme (CC BY-NC-SA; see 2.3).

4.2 Are there any necessary limitations to protect sensitive data?
Questions you might want to consider:
- Under which conditions will the data be made available (timing of data release, reason for delay if applicable)?
Data have to be shared as soon as possible, but at the latest at the time of publication of the respective scientific output. Restrictions may be only due to legal, ethical, copyright, confidentiality or other clauses. Consider whether a non-disclosure agreement would give sufficient protection for confidential data. (This relates to the FAIR Data Principles A1 & R1.1)

Data will be made available to peer-reviewers at the time of manuscript submission.
Data will be made publicly available at the time of manuscript publication.

There are no confidential data and no anticipated restrictions due to legal, ethical, copyright, confidentiality or other clauses.

4.3 I will choose digital repositories that are conform to the FAIR Data Principles. [CHECK BOX]
The SNSF requires that repositories are conform to the FAIR Data Principles (Section 5 of the guidelines for re-searchers, SNSF's explanation of the FAIR Data Principles).
If there are no repositories complying with these requirements in your research field, please deposit a copy of your data on a generic platform (see examples).
If no data can be shared, this is a statement of principles.

Yes

4.4 I will choose digital repositories maintained by a non-profit organisation. [RADIO BUTTON yes/no]
If the answer is no: "Explain why you cannot share your data on a non-commercial digital repository."
The SNSF supports the use of non-commercial repositories for data sharing. Costs related to data upload are only covered for non-commercial repositories.

Yes