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)

I will collect photo diode signals, spectra or CCD images, i.e., one- or two-dimensional arrays of floating point numbers with up to a few Kbyte to Mbyte per individual measurement. Each array will be recorded as a function of different parameters, such as temperature, external field or pump fluence. The files will be stored in .txt format and converted into .dat or .mat format for data analysis using, for example, Matlab software. According to previous experience, we expect to generate approximately 50-100 GB of experimental data per year. All raw data will be archived on group network servers. For the re-use of third-party data (e.g. spectra from previously analysed molecules in specific conditions) I 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)

Collection will be performed using detector MOS and CCD and the produced signals will be digitalized and transformed in ASCII format to be used for further analysis. Data will be collected and documented by a read-me.txt specifying what information (metadata) needs to be collected in conjunction to every experiment (e.g., day/time, name of investigator, sample name, replication of experiment in a series, basic characteristics of the spectroscopic investigation (e.g., wavelengths, fluences, polarization etc…). Care will be taken to use predefined and consistent vocabularies throughout the entire project.
Meta-(data) will be reviewed by the principal investigator in a short timeframe after each experiment 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 (institute server). The data structure will allow the access of the data by the research team, e.g. by date, sample characteristics as well as other parameters. The metadata will be linked to the raw data files. The database has appropriate interfaces via direct access or through a 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 to accompany the raw data:
- Date/time of the experiment
- Parameters related to the preparation (e.g., LC species, embedding cell, molecule concentrations)
- Indications regarding the experimental protocol (e.g., temperature, pump-probe wavelengths, polarization, etc.)

This information will be stored as ASCII text files 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. This information will be generated in many cases automatically by the measurement software or if required by the person recording the data. Moreover, relevant data will be stored as well on BORIS, the University's repository, in form of open source electronic notebooks, such as OpenWetWare from MIT or sciNote. The data will be published based on the Dublin Core Metadata Schema (cf. 4.1). The schema includes amongst others information on authors, the title, abstract, DOI, publication date, format, as well as data access and reuse.

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 generate any personal or otherwise sensitive data. There are no security issues or associated risks.
The data will be accessible via direct access or via 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.
State-of-the-art security standards will be implemented on the servers. The Institute of Applied Physics is protected by the University's firewall, and our server is managed by the IT engineers of the Institute and University. Moreover, I will open a personal server for my subgroup, funded by my hosting group. 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 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

Immediately after each experiment, the generated data and metadata will be copied via a LAN connection on our institute servers, on which capacity in the terabyte range is attributed to the research group. The data on the institute servers is automatically backed up every night.
This procedure is routine in our Institute and the management of the servers and the backups is assured by the team members and, if needed, by the IT engineer. 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?

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 in our institute's 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, 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?

Generally, we generate two classes of data, those that turn out to be faulty due to different reasons, e.g., instrumentation failures during measurement et cetera. After quality control of the remaining data, relevant data sets are extracted. Those data sets, which often also end up in the corresponding publications, will be shared. In detail, the original data of relevance together with a detailed recipe describing the data evaluation process will be stored together with the corresponding publication on BORIS [https://boris.unibe.ch/], the University's repository, in form of open source electronic notebooks, such as OpenWetWare

from MIT or sciNote. This ensures open access to the data 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 and electronic notebook files will be made available to peer-reviewers at the time of manuscript submission and publicly at the 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 organization. [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