

## Template for the SNSF Data Management Plan for openBIS users at ETH Zürich.

*Scientific IT Services of ETH Zürich.*

### 1. Data collection and documentation

<p><b>1.1 What data will you collect, observe, generate or reuse?</b>          Questions you might want to consider:</p> <ul style="list-style-type: none"> <li>- What type, format and volume of data will you collect, observe, generate or reuse?</li> <li>- Which existing data (yours or third-party) will you reuse?</li> </ul>	<p style="color: red;"><i>To be filled by researchers.</i></p>
<p><b>1.2 How will the data be collected, observed or generated?</b>          Questions you might want to consider:</p> <ul style="list-style-type: none"> <li>- What standards, methodologies or quality assurance processes will you use?</li> <li>- How will you organize your files and handle versioning?</li> </ul>	<p style="color: red;"><i>To be filled by researchers.</i></p> <p style="color: red;">All files produced during this project will be stored in our Electronic Laboratory Notebook (ELN) and Laboratory Information Management System (LIMS) openBIS. In this ELN, each scientist has a personal folder where to organize projects and experiments. Each experiment is described in the electronic notebook and all data related to the experiment is directly attached to it, in so called "datasets". Each dataset is immutable, thus different file versions are stored in the lab notebook in different datasets with a manually generated version number. Very large datasets (100s of TBs) are not directly stored in openBIS datasets, but they are linked to the experimental description using an extension to openBIS called BigDataLink. This works similarly to the git version control software, so every time changes are made to the data, these need to be committed to openBIS, which automatically keeps track of the versioning.</p>
<p><b>1.3 What documentation and metadata will you provide with the data?</b>          Questions you might want to consider:</p> <ul style="list-style-type: none"> <li>- What information is required for users (computer or human) to read and interpret the data in the future?</li> <li>- How will you generate this documentation?</li> </ul>	<p style="color: red;"><i>To be filled by researchers.</i></p> <p style="color: red;"><i>To be filled by researchers.</i></p>

- What community standards (if any) will be used to annotate the (meta)data?

In the data management system (openBIS ELN-LIMS), metadata are provided as attributes of the respective datasets. Based on the defined metadata schema, openBIS ELN-LIMS will be configured so that the required metadata is automatically assigned to datasets and / or manually provided by the researcher.

## 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?

*To be filled by researchers.*

### 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?

All data generated in the project will be stored in our openBIS ELN-LIMS. This operates in a client-server model, which is installed and maintained by the ETH Zurich IT services on ETH Zurich infrastructure. Researchers can access openBIS via any of the most common web browsers. openBIS requires user authentication with ETH Zurich credentials and it provides user right management, so that different users can have different access to all or different parts of the system, as required. Below is a description of the default openBIS roles, which can be modified upon request:

1. Instance admin. Has full admin powers. Can customize settings, create, modify and delete entities, assign user roles, create data spaces.
2. Instance observers. Has read-only access to everything in openBIS.
3. Space admin. Can create, modify, delete entities and assign roles only within a given data space.
4. Space power user. Can create, modify and delete entities only within a given data space.
5. Space user. Can create and modify entities only within a given data space.
6. Space observer. Has read-only access limited to a given data space.

openBIS does not offer any specific option for sensitive data, but the data will be encrypted prior to upload to openBIS. Furthermore, all operations on the system (incl. which users log in and when) are logged, so that it is fully transparent who did what to the data and when.

The data stored in openBIS is physically located on a NAS (network attached storage) provided by the ETH Zurich IT Services. The access to the share's data is governed by the latest security best practices and only a limited number of employees of the ETH Zurich IT services have access to that share.

**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?

*To be filled by researchers.*

### 3. Data storage and preservation

<p><b>3.1 How will your data be stored and backed-up during the research?</b></p> <p>Questions you might want to consider:</p> <ul style="list-style-type: none"><li>- What are your storage capacity and where will the data be stored?</li><li>- What are the back-up procedures?</li></ul>	<p>Our data will be stored in openBIS ELN-LIMS, maintained by the ETH Zurich IT Services. openBIS uses a postgres database that stores all metadata. This database is backed up (“pg_dump”) every night with a 7 days retention of the dumps and fully backed-up twice a week with a backup retention of 20 days. The full backup procedure includes a point-in-time recovery that allows a finer granularity (up to minutes) of data recovery in case of a disaster. The database backup is stored on the NAS (network attached storage) provided by the ETH Zurich IT services. The same NAS is used to store the data uploaded to openBIS. This network attached storage is snapshot every night with a 7 days retention, and data is backed up on a proprietary tape library with a retention of 90 days.</p> <p>Data which is no longer actively needed is moved to the long term storage (i.e. tapes). The tape library where openBIS moves the data has a read-only replica in a different geographical location in order to minimize any data loss.</p> <p><i>For data linked to openBIS with the BigDataLink tool, please provide details of the data location and back-up.</i></p>
<p><b>3.2 What is your data preservation plan?</b></p> <p>Questions you might want to consider:</p> <ul style="list-style-type: none"><li>- What procedures would be used to select data to be pre-served?</li><li>- What file formats will be used for preservation?</li></ul>	<p><i>To be filled by researchers.</i></p>

## 4. Data sharing and reuse

<p><b>4.1 How and where will the data be shared?</b> 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?</p>	<p><i>To be filled by researchers.</i></p>
<p><b>4.2 Are there any necessary limitations to protect sensitive data?</b> Questions you might want to consider: - Under which conditions will the data be made available (timing of data release, reason for delay if applicable)?</p>	<p><i>To be filled by researchers.</i></p>
<p><b>4.3 I will choose digital repositories that are conform to the FAIR Data Principles.</b> [CHECK BOX]</p>	<p><i>To be filled by researchers.</i></p>
<p><b>4.4 I will choose digital repositories maintained by a non-profit organization.</b> [RADIO BUTTON yes/no] _If the answer is no: "Explain why you cannot share your data on a non-commercial digital repository."</p>	<p><i>To be filled by researchers.</i></p>