**Roberto Fray da Silva**

**Danilo de Souza Miguel**

Data Life Cycle Plan

Project PCS5031 – University of São Paulo - Brazil

Roles and responsibilities

**1. Outline the staff/organizational roles and responsibilities for implementing this data management plan.**

Roberto Fray da Silva and Danilo de Souza Miguel

Custodians of the repository: Cloud (Roberto Fray da Silva), Backup (Danilo de Souza Miguel), Github repository: <https://github.com/rfsilva1/wsn-analysis>.

**2. Who will be responsible for data management and for monitoring the data management plan?**

Roberto Fray da Silva.

**3. How will adherence to this data management plan be checked or demonstrated?**

The adherence to the plan was checked will be checked weekly, during the 7 stages of the data life cycle. After the project ends, a DOI is going to be generated for the database, and this will be made available on the project's repository on Github.

**4. What process is in place for transferring responsibility for the data?**

All the analysis will be stored on Github (codes / scripts), but the original dataset belongs to the Sensorscope project (<http://lcav.epfl.ch/op/edit/sensorscope-en>), and is maintaned on their servers.

**5. Who will have responsibility over time for decisions about the data once the original personnel are no longer available?**

Similar to the past question, the dataset is stored on the Sensorscope project servers, as it belongs to that research group. Our analyses and codes will be stored in our Github repository, which will remain open for the public.

Expected data

**1. What data will be generated in the research?**

The data generated on the original research (by the Sensorscope research group) is related to data collected by sensors (ambient and soil temperatures, ambient and soil humidities, solar radiation, wind direction and velocity). These will be processed and analyzed using R, and the analyses will be made available on the Github repository.

**2. What data types will you be creating or capturing?**

The data is mainly structured, related to data collected by the sensors.

**3. How will you capture or create the data?**

The data was already collected, we will just use it for analysis.

**4. If you will be using existing data, state this and include how you will obtain it.**

The data was collected from the Sensorscope project website (http://lcav.epfl.ch/op/edit/sensorscope-en).

**5. What is the relationship between the data you are collecting and any existing data?**

The data is part of the Sensorscope project, which is related to the deployment of wireless sensor networks in Switzerland. All the dataset that were generated on this project are available on its website.

**6. How will the data be processed?**

The data will be converted to CSV and imported in a NoSQL database, probably MongoDB. The pre-processing, processing and data analysis will be done using the R statistical software. The software LibreOffice Calc will also be used on the data analysis. The R Shiny software will be used for data visualization.

**7. What quality assurance & quality control measures will you employ?**

The QC measure that will be adopted is weekly checks on the database.

Period of data retention

The codes and data generated will be maintained for 10 years.

Data formats and metadata

**1. Which file formats will you use for your data, and why?**

The data collected will be maintained in a NoSQL database. The figures and charts will be stored in JPEG format, and the spreadsheets in XLSX.

**2. What form will the metadata describing/documenting your data take?**

After analyzing the patterns that are available, we chose to use the SensorML schema to describe the sensors, and the TransducerML schema to describe the processes and readings.

**3. How will you create or capture these details?**

The data will be collected from the Sensorscope website. During the analysis, this information will be stored in XML files.

**4. Which metadata standards will you use and why have you chosen them? (e.g. accepted domain-local standards, widespread usage).**

Already answered on question 2.

**5. What contextual details (metadata) are needed to make the data you capture or collect meaningful?**

The description available on the Sensorscope project website are sufficient for the description we propose to do.

Data dissemination and policies for public access, sharing and publication delays

**1. How and when will you make the data available? (Include resources needed to make the data available: equipment, systems, expertise, etc.)**

The codes used for processing the data, import it on the database, export it, and visualize it, are going to be available on the project's Github repository.

**2. What is the process for gaining access to the data?**

The repository will be publicly available.

**3. Will any permission restrictions need to be placed on the data?**

No. The raw data will not be made available in our repository (as its not our property). It should be downloaded on the Sensorscope project website. We will make all our analyses and codes available in our repository.

**4. Are there ethical and privacy issues? If so, how will these be resolved?**

This is not a relevant question for this project.

**5. What have you done to comply with your obligations in your IRB Protocol?**

This is not a relevant question for this project.

**6. Who will hold the intellectual property rights to the data and how might this affect data access?**

The intellectual property belongs to EPFL and the project Sensorscope researchers.

**7. Which bodies/groups are likely to be interested in the data?**

Researchers from different areas: computer engineering, information science, biology, and meteorological science.

**8. What and who are the intended or foreseeable uses/users of the data?**

The same as described in question 7.

**9. Do you plan on publishing findings which rely on the data?**

Yes, in an event on the computer engineering area.

**10. If so, do your prospective publishers place any restrictions on other avenues of publication?**

No.

**11. How long will the original data collector/creator/principal investigator retain the right to use the data before opening it up to wider use?**

This does not depend on us. This depends on the data owners (EPFL and the Sensorscope project).

**12. Explain details of any embargo periods for political/commercial/patent or publisher reasons.**

This is not a relevant question for this project.

Data storage and preservation of access

**1. What is the long-term strategy for maintaining, curating, and archiving the data?**

The processed data and the analyses will be stored in our personal computers, in an external HD, and in our Github repository.

**2. Which archive/repository/database have you identified as a place to deposit data?**

This question was already answered before.

**3. What procedures does your intended long-term data storage facility have in place for preservation and backup?**

This is not a relevant question for this project.

**4. How long will/should data be kept beyond the life of the project?**

10 years.

**5. What data will be preserved for the long-term?**

Data and analyses.

**6. On what basis will data be selected for long-term preservation?**

All the data and analyses should be preserved.

**7. What metadata/documentation will be submitted alongside the data or created on deposit/transformation in order to make the data reusable?**

The metadata schema SensorML and TransducerML will be used to increase the data and analyses distribution.

**8. Outline the staff/organizational roles and responsibilities for implementing this data management plan.**

This question was already answered before.

**9. How will adherence to this data management plan be checked or demonstrated?**

This question was already answered before.