

h e p i a Haute école du paysage, d'ingénierie et d'architecture de Genève



MIcroclimate DAta Stories (MIDAS)

Territoire en Recherches 2021-2023

Data Management Plan

04.2023

Marielle Guirlet, Scientific Collaborator, HEG-Genève, HES-SO René Schneider, PI, HEG-Genève, HES-SO

h	e	p	i	а	
Haute école du paysage, d'ingénie et d'architecture de Genève					



Date	Version	by
02.12.2022	draft	Marielle Guirlet
03.02.2023	consolidated	Marielle Guirlet
04.04.2023	consolidated	Marielle Guirlet
27.04.2023	finalized and validated	Marielle Guirlet, René Schneider

This document describes the management plan of data collected and produced in the context of the project <u>MIDAS</u> between Avril 2021 and April 2023. It is based on the <u>Twenty-questions for research data management</u> by David Shotton (Shotton 2012).

The nature of your data

1. What is the subject discipline (domain, field) to which your research data relates ?

The data collected and handled in the frame of the MIDAS project relate to environment in a broad sense. One type of data relates to phenology (blossoming dates of tree leaves in Geneva). Another type relates more specifically to atmospheric sciences. Those data are measurements of meteorological quantities and of atmospheric composition. They are provided by scientific institutions and organizations (Meteoswiss, SABRA, Climacity data) and by the citizen association TSF (Terrasses sans Frontières). TSF data come from measurements made on the roofs of several Geneva buildings.

2. What is the exact nature (range, scope) of your research data ?

The research data collected and analyzed are:

- Blossoming dates of the first leaf of the official Geneva chestnut tree in springtime (see <u>here</u> for more information)
- Data describing the physical and chemical atmospheric conditions:
 - Meteorological data : temperature, atmospheric pressure, relative humidity, ... (Meteoswiss, Climacity and TSF)
 - Air quality data : atmospheric pollutants : NO₂, O₃, PM2.5, PM10 (SABRA)
- Contextual information for TSF temperature and relative humidity measurements

In the second stage of the project, we produce Data Stories based on the data previously collected and analysed. These Data Stories include several aspects : data visualization and statistical analysis of the aforementioned research data, along with text. We consider these data stories as the main research outputs of our project. As such, like our research data, they need management of their own as described in this document.

3. In what format(s), will you store your data in the short term after acquisition?

Most data are stored in .txt, .csv and .xlsx files on my PC:

- Blossoming dates are stored in .xlsx and .csv files
- Meteoswiss data samples are stored in .txt files
- SABRA data samples are stored in .csv files
- Climacity data samples are stored in .txt and .csv files
- TSF data are stored in .csv files (temperature and relative humidity measurements), .pdf files (measurement tracks) and text files (information about the measurement context)

The extended datasets and the TSF ones are also provided through a <u>dashboard</u>. They may be downloaded from the dashboard in .txt or .csv format depending on the source of measurement.







The Data Stories are of different types (e.g., annotated chart, Data Comic, presentation with slides). Thus different formats:

- Annotated chart: source file in Tableau-software native format (twb); image version in png
- Data Comic: source file in Adobe Photoshop native format (psd); image version in jpg or png
- Presentation with slides: source file in Microsoft PowerPoint (ppt); image version in pdf.

The source files may be exported as images in pdf, jpg or png files.

4. Who owns the data arising from your research, and the intellectual property rights relating to them?

Some data collected and handled in the frame of the project are provided by our project partners (e.g., TSF, HEPIA for Climacity). Other ones come from external institutions (e.g., Geneva city, Meteoswiss, SABRA).

The data provided through the dashboard are already made accessible with no restriction by their source institution (or with little restriction for the Meteoswiss ones: they should be used for scientific research purposes only). Users are asked by the data providers to acknowledge the source of their data as follows:

- Meteoswiss: "source: MétéoSuisse"
- SABRA: "source: Etat de Genève SABRA"
- Climacity: "source: HEPIA"

About the blossoming dates, we acknowledge the "Ville de Genève" for the data.

About TSF data, we propose to acknowledge them as: "source: TSF/HEPIA".

Potential users of the dashboard are informed that they should use these statements to acknowledge the source of the data. We also include these statements in the Data Stories when they are based on such data.

The Data Stories and other related products of our research project are owned by our institution of affiliation (HEG Genève, HES-SO).

Data descriptions (metadata, "data about data")

5. How will your research datasets be described?

Collected data:

- Climacity, Meteoswiss and SABRA data: the metadata are written in the header of the files. Each measurement is time-stamped. The exact measurement time is provided along the other measured quantities at each measurement step.
- TSF data: the measurement location and day are stated in the header and also in the filename. Measurements are time-stamped. The exact measurement date (hh:mm.ss) is provided along the other measured quantities at each measurement step. If available, contextual information (measurement tracks, possibly a picture of the roof, other details) is stored in separate files.

Research products (e.g. Data Stories):

The Data Stories produced so far are based on a single topic and on the same data. Some information about the data we use is stated on the Data Stories themselves (data source, time span, location, physical quantity, ...). As a "by-side product" of those Data Stories, we also produce a <u>log book</u> (in French) which describes our workflow and gives all details on those data.

6. How will these descriptive metadata be created or captured ?







TSF sends us the context information by e-mail along with the measurement files. From the information and text provided in separate e-mails, I produce a description text file (one file per measurement location). I also store the pictures or photographs in separate files (measurement tracks, pictures of the building roofs).

The descriptive metadata of the Data Stories are inferred from the data themselves (measurement time span and location) and stated in the Data Stories as text elements.

Data sharing

7. With whom will you share your research data in the short term, before publication of any papers arising from their interpretation?

At first, the data were shared among the team members via switchdrive (sample of collected data and working versions of the Data Stories).

With the dashboard, more complete data sets are now made available for visualization and for download to anyone aware of the dashboard (also see 3.). The <u>report on TSF measurements</u> (in French) presents the results of the analysis of these data and also includes some contextual information for those measurements.

The Data Stories can be accessed from the project web site.

At the end of the project, the Data Stories and their underlying data will be made open to anyone on the archive <u>Yareta</u> (Organizational Unit: HEG IS MIDAS 2021-2023, title: Microclimate Data Stories (MIDAS)).

Data storage and backup

8. Where will you store your data in the short term, after acquisition?

The data are stored on my computer and on my laptop, and on my colleague's laptop. The data made available via the dashboard are stored on a specific server to which the dashboard is linked.

In the frame of our project, a hackathon was organized in October 2021. The data made available to the participants (sample sets) were copied after the hackathon to two hard disks. Those disks are physically stored in the office of my colleagues.

9. Who is responsible for the immediate day-to-day management, storage and backup of the data arising from your research?

Myself alone for the data stored on my computer and my colleagues in charge of the dashboard.

10. How frequently will your research data be backed up for short-term data security?

Regularly, as often as our computer accounts and the server are backed up by the automatic back up service of our institution.

Data archiving

11. Where will your research data be archived for long-term preservation?

The Data Stories and their underlying data will be archived on the long-term preservation archive Yareta.

12. When will your research data be moved to a secure archive for long-term preservation and publication?

The Data Stories and their underlying data will be moved to the archive when they are finalized, at the end of the project at the latest (April 2023).



13. Who will decide which of your research data are worth preserving?

Myself in agreement with other team members and our project PI.

14. How (i.e. by what physical or electronic method) will you transfer your research datasets to their longterm archive, under the curatorial care of a separate third-party, e.g. a data repository?

By completion of the submission form and uploading of the data files over the Internet.

Data publication

15. For how long will you embargo your research data before it is published for others to see and use?

We will allow immediate public access to the Data Stories and to the data.

16. Why is public access to your research data to be restricted (if indeed it is)?

All research data will be made publicly available, since no access restriction applies.

17. Under what data-sharing license will you publish your research data?

CC-BY-NC-SA.

18. What persistent identifiers will be used to permit correct citation of your datasets?

The accession number of the dataset issued by the database Yareta to which it is submitted (a DOI).

19. What metadata will be published with the data to make them interpretable and reusable?

As mentioned in 5., all details on the Data Stories (workflow, content, data underlying them) are stated in the log book. These details relate to the workflow, the content, the data underlying them, and the context within which the Data Stories are to be used: audience type, interactivity level for instance.

Future data management

20. Who will be responsible for your data, once you have left your present research group?

The project PI is taking over responsibility for these data.

Bibliography

SHOTTON, David, 2012. Twenty Questions for Research Data Management. Data Management Planning [online]. 07.03.2012. 07.05.2013. [Viewed 27.04.2023]. Available from : https://datamanagementplanning.wordpress.com/2012/03/07/twenty-questions-for-research-datamanagement/