Data Stewardship Ex1

A Data Management Plan created using DMPonline

Creator: Gerald Weber

Affiliation: Universitaet Wien - University of Vienna (Austria)

ORCID iD: https://orcid.org/0000-0002-1221-8045 https://orcid.org/my-orcid

Template: TU Wien

Grant number: none

Project abstract:

Analysis of two dataset and creation of DMP. The datasets should be computed and stored in a reproducable way for other scientists or public review.

Last modified: 22-04-2019

Data Stewardship Ex1

Data Collection

What data will you collect or create?

For the analysis of bike usage patterns on the Fremont Bridge, we collect a dataset from data.seattle.gov.

The dataset contains of 56.9K rows and 3 columns and describes the number of hourly bicycle count on the Fremont bridge since October 2012. The contained columns are Date/Time, East pathway count and west pathway count.

The direction of the bicycles is not covered, but it is supposed, that most of the bicycles on the east pathway are traveling northbound while bicycles on the west pathway are traveling southbound.

The choosen format is CSV because it is simple to import that data with Python (delimiter ,).

Date of the data retrieval was 2019-04-19. The last line of the data is from 2019-03-31 11:00 PM.

Size of the data is 1597 kB. Due to the small size of data, there should be no problem sharing the data.

The result of the analysis will be a jupyter notebook as well as a PDF export of the jupyter notebook to make the results available to users without a jupyter installation.

How will the data be collected or created?

For this analysis, the data is retrieved only once by download link on the Seattle open data platform.

https://data.seattle.gov/Transportation/Fremont-Bridge-Hourly-Bicycle-Counts-by-Month-Octo/65db-xm6k

The name is provided by the platform: Fremont_Bridge_Hourly_Bicycle_Counts_by_Month_October_2012_to_present.csv

The name does not contain any information about the retrieval time.

Data was last updated on April 2, 2019 and provided by Seattle Department of Transportation. There is no exact information about update interval (update frequency: other). The Jupyter Notebook contains an automatic download of the data, if no filename is provided or the file could not be found on disk. No further data preparation is required for the analysis.

Documentation and Metadata

What documentation and metadata will accompany the data?

The documentation contains the used software libraries and their versions.

Metadata is generated automatically by issuing several statements to retrieve the used software and software versions.

The used statements are supplied in the README.md file. Information about the versions are provided in separate files containing the raw output of the provided statements. These files were created by calling the statements and supplied with the research.

Metadata is supplied with research result metadata.xml

Ethics and Legal Compliance

How will you manage any ethical issues?

The data is part of the open data initiative of the city of Seattle and does not contain any sensitive data. It is open for research, entrepreneural uses and general information to city residents.

http://www.seattle.gov/tech/initiatives/open-data/about-the-open-data-program

http://www.seattle.gov/tech/initiatives/privacy/data-we-collect

How will you manage copyright and Intellectual Property Rights (IPR) issues?

Data has been provided by Seattle Department of Transportation and its licence is public domain.

Storage and Backup

How will the data be stored and backed up during the research?

The data will be uploaded to an DOI repository and can therefore be shared and recovered by the use of DOI.
https://doi.org/10.5281/zenodo.2648564

How will you manage access and security?

While the data and the analysis is public domain and is open for everyone.
No access restrictions required.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

To validate the findings, the same dataset should be used. It can be retrieved from the Zendoo storage via DOI
https://doi.org/10.5281/zenodo.2648564

What is the long-term preservation plan for the dataset?

For our research we will use Zendoo as our long term data storage. No additional costs charged. While we have choosen CSV as our file format we do not need any software to transform the dataset.

Data Sharing

How will you share the data?

Data is linked in the research and is publicly search and findable via Zendoo. The dataset is public domain and can be used freely.

Are any restrictions on data sharing required?

None.

Responsibilities and Resources

Who will be responsible for data management?

Retrieval and storage of the dataset is done by Gerald Weber.
The author of the research (Gerald Weber) is responsible for checking the availability of the dataset twice a year (latest 31.1. & 31.7. of each year) and take action to restore the availability in case of problems.

What resources will you require to deliver your plan?

No additional expertise is required.
Infrastructure for storage is provided by external providers. Their availability will be checked by the author twice per year.