DMP Template for the Social Sciences

A DMP in a nutshell

A Data Management Plan (DMP) is a structured guideline that describes the comprehensive lifecycle of data, from conception to storage, analysis, and preservation. DMPs help researchers to think through all relevant questions concerning the data their research will generate, and ensure attention remains focused on the long-term accessibility and subsequent reusability of their data assets. DMPs provide a basic description of what kind of data will be produced and collected, and details about what will happen to the data both during a project and after it has been completed. This includes statements about the provenance of data, contextual information surrounding the data collection process, how data are conceptually related to data sets produced by similar studies, the infrastructures used to store and manage data, as well as information regarding the publication, citation, long-term access and, if necessary, destruction of data when the research lifecycle is complete. Humanities and social sciences data are unique in that they often consist of private information contributed by individual study participants, thus various questions regarding data protection, copyright attribution, exploitation rights, and licensing are also addressed in this template.

FAIR data principles

The template is also compliant with the FAIR principles to improve the findability, accessibility, interoperability, and reuse of the data. This implies that research data and contextual tools like software should be stored and made available for use in a suitable repository or archiving system and data should be provided with persistent identifiers. Data must be identifiable, accessible, traceable, interoperable, and whenever possible, available for subsequent use. In compliance with intellectual property rights, and if no third-party rights, legal requirements or property laws prohibit it, research data should be assigned a licence for open use.

Note on the handling of the DMP

Please consider the DMP as a research instrument, helping to structure and plan the research process and define the responsibilities within a joint research project. It can vary in length and detail depending also on the type of data and project-stage. Thus, not all questions might be relevant for you, especially at the beginning of a project. Rather regard the DMP as a dynamic document which can be updated until the end of the project. In order to keep track of different versions, the version number of each DMP should always be included in the administrative section below.

1. Administrative Data

Provide basic information to help identifying your research project and people involved.

Include: Principal investigator, project sponsor/ grant and number, project title, project coordinator (name, affiliation, email address, phone number, and IDs such as ORCID if available), author of the DMP (name, affiliation, email address, phone number, and IDs such as ORCID if available), start and end date of project, version of DMP, date of first DMP version, date of last update, short data summary, any policies you adhere to. You may add any additional basic information.

Version of DMP	1.0
Project coordinator	Yang Yucheng
Principal investigator	Yang Yucheng
Author of DMP	Yang Yucheng
Data officer and responsible for DMP	Yang Yucheng
Project title (Acronym)	
Start and end date of project	5.12.2024-30.6.2025
Grant number	1
Data summary	Urban form is an organic, dynamic and complex system which accumulates the urban growth, including population growth, constructions and daily activities, over several centuries. Having a solid understanding of urban form situation is one of the most critical factors of the conservation of historical urban landscape. However, there is still no common agreement about how to recognize the urban form precisely. In recent years, quantitative urban form studies have emerged with the development of information techniques. Because of the merits such as high efficiency and accuracy, more and more researchers are trying to use these approaches to address the recognition and category of urban form. In this research, I take Vienna as an example and attempt to find the principles of urban form distribution and classification in Vienna. Due to the long history of construction and human activities, Vienna possesses a rich accumulation of the urban form from different ages and styles, making it a great example for research

2. Data Characteristics

Provide information on the data collection/ generation process. Which data will be collected and how will it be collected. This also helps in evaluating which software and hardware will be necessary. This includes the type of data that will be generated, a description of methods and data handling as well as the formats that will be generated.

Include: data provenance and data sources, versioning of data, method of data collection, formats of data, amount of data expected to be collected (gigabytes, terabytes), software/ hardware used for data collection/ processing/ storage, aspects on reuse (format), (personnel) costs for data processing and data storage. Will you re-use any existing data and how? What is the origin of these data?

The source of the data is mainly from Geo-Information of Vienna City. The data will be collected by downloading from the website of Vienna City, and this will result in the format of CityGML 3.0 or shapefile. The building blocks information are also accessed from the website. Building height, an important factor for understanding urban form, also plays a critical role in my analysis. The formats generated will be readable by using Geographical software, such as ArcGIS Pro and QGIS. Remote sense images are also an important part of reference to study in the urban form. The city of Vienna also has the digital data in the website, and the precision of the data should be in 1-meter resolution.

The temporal coverage should also be considered in the research. The different period of the data from 2000-2024 in each 5 years would be compared to find the tendency of the urban form evaluation.

During the data processing stage, the Multiple Fabric Assessment (MFA) protocol based on Araldi and Fusco (2019) will be employed for the identification and characterization of urban fabrics, and the publication of the data will be in a Shapefile format to ensure long-term usability.

3. Documentation and Metadata

3.1. Metadata standards

Provide information on the metadata standards you will be adhering to (you may include metadata standards used by your chosen repository system). Metadata are a standardized scheme to describe datasets. A unified metadata standard benefits findability and usability. The standard used in social sciences is DDI (Data Documentation Initiative). Will you be using standard vocabularies for all data types present in your dataset, to allow inter-disciplinary interoperability and to allow the metadata to be machine-actionable? In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? What naming conventions do you follow? Also consider documenting the version number of the metadata standard you will be using.

Include: structure of metadata, metadata standards, persistent identifiers (e.g. DOI).

The metadata structure should include dataset name, description and keyword.

The metadata standard DDI(Data Documentation Initiative) and ISO 19115(Geographic Information) will be used, as these are the standard for spatial data. The data formats will include Shapefile, CityGML, TIFF(for remoting sensing images). And the coordinate reference systems should based on the original data from the city of Vienna, Austria_Gauss-Krueger_M34_NORD-5Mio

.The DOI and versioning including version numbers and update timestamps for the dataset will be used as the persistent identifiers.

3.2. Documentation

Provide information on which documents you will prepare (documents, protocols, related records….). The documentation outlines the research process and ensures integrity, understandability and transparency of the data collection process and facilitates correct interpretation (also consider machine readability of the data and documentation). Has the software needed to access the data been sufficiently documented? Is it possible to include the relevant software (e.g. in open source code)?

Include: methods reports, instruments of data collection (e.g. questionnaire, code), codebook, program code for data processing and data code for analysis and tools/methods of long-term archiving. You may also add informed consent forms, transcripts of an audio file as well as tools and software including the version number.

> The data collection and processing activities should include spatial data preparation and indicator computation. Spatial data preparation is about to explain the definition of the spatial units adapted for the study of Vienna. And indicator computation outlines the calculation of urban form, such as pedestrian ratio, building heigh, vegetation coverage, and the geoprocessing protocols for generating these metrics will be used.
>
> The code that was used to analysis the urban form should include analysis algorithms, for example, using Local Moran's I for detecting spatial pattern characters.
>
> Prepared codebook should include the lists and explains about all variables and indicators, including street morphologies such as street grid, vegetation metrics, building scales, and network-building relationship indicators. And some examples should be provided for each variable to ensure clarity and reproducibility.

3.3. Data quality control

Provide information on the quality assurance measures (protocols) you will implement.

Include: all data quality assurance processes, such as pretests, data entry validation, peer-review of data, repeat samples or measurements, definition of standardized processes and checklists, intercoder reliability measures. Provide information about deletion processes. Make sure to consider additional costs. In the event that research data and records are to be deleted or destroyed, either after expiration of the required archive duration or for legal/ ethical reasons, such action will be carried out only after considering all legal and ethical perspectives. The interests and contractual stipulations of third-party funders and other stakeholders, employees and partner participants in particular, as well as the aspects of confidentiality and security, must be taken into consideration when decisions about retention and destruction are made. Any action taken must be documented and be accessible for possible future audit.

> To ensure that the output of the data collections process will result in high-quality, valid that that can be replicated and reused, the following measures will be taken.
>
> Firstly, the research relies on the definition of indicators, such as urban morphology metrics including building contiguity, street network attributes. So, the clear definitions will enable the consistency in data collection and analysis. And geospatial analysis and statistical validation should also take into consideration.

4. Data Availability and Storage

4.1. Data release and sharing strategy

Provide information on how and when data will be released and made accessible for sharing (and further reuse).

Include: the repository/archiving system(s) and persistent identifiers you will assign, any restrictions in the data release, sharing and usage options, how data will be shared within the project, how data will be shared outside the project, any restrictions in data sharing and why these restrictions exist, access options (open access, restricted access, no-access, scientific use, educational purposes etc.), license options used to pursue sharing strategy (e.g. Creative Commons, General Public License, GNU), machine-readability of licences, embargo periods, informing potential users about the availability of data, search keywords that optimize possibilities for re-use, data discovery strategies, identification procedure for persons accessing the data, interoperability of data, open software applications used, procedure when data collected is combined with data stemming from other sources to ensure interoperability and future reuse, costs for making data traceable according to FAIR requirements (findable, accessible, interoperable, reusable).

> The data will be stored in a database under the „Creative Commons Namensnennung 4.0 Lience", such as Zenodo. So, the dataset should be available to those who is interested in through the website. Data should be storage in the widely accepted formats such as CSV, GeoJSON, Shapefile or TIFF.

4.2. Data storage strategy

Provide information on the processes and strategies in place to ensure safe data storage *during* the research process. Safe data storage also includes physical security of infrastructure.

Include: software, collaborator access, security, backups, collaborative workspaces, transfer of data from the field to storage, costs, etc.

> During the research, data will be stored in a database, such as Zenodo, which is free and open access with the support of geospatial data formats. And regularly back up date using cloud services or other hard drives is necessary in order not to lost the key data elements.

4.3. Data preservation strategy

Provide information on the processes and strategies in place to ensure safe data storage and access to data *after* the research project is completed. Describe which data will be archived long-term and how this decision will be made.

Include: the duration of the guaranteed storage period, what will happen to data not archived long-term (deleting or erasing data, deletion strategies and protocols), where data will be preserved to ensure permanent access (state the name of the repository), any associated costs (for archiving or while processing data before archiving, including legal and ethical questions and software considerations), which long-term formats will be used (does the repository suggest or prescribe any formats).

All data collected as part of the urban morphology research will be available for the interested researchers for a long period. The data should be saved in a public accessible repository such as Zenodo. And the stored file should be the file formats which is widely used including GeoJSON, csv to keep the data's accessibility.The repository will assign persistent identifiers to ensure its reliability of location and citation.

5. Legal and Ethical Aspects

5.1. Legal Aspects

Provide information on any issues that may arise during data collection, storage, release, sharing and publishing.

Include: considerations on legal or ethical barriers to sharing data, data ownership, planned license for reuse or replication, any restrictions on data reuse or replication and why, for projects with international partners consider different national legislation, who has permission to publish data (copyright, ownership, Intellectual Property Rights, data protection), legal situation concerning copyright, exploitation and individual rights, permission to collect data (informed consent), permission for third-party data used in research, processes in case of breaches, personal information been used in the research and is anonymization, pseudonymization or recoding necessary, additional costs for legal questions.

Data collected in this research project is owned by city of Vienna. Permission to collect data was granted by Creative Commons Namensnennung 4.0 Lience. And this license will allow to reuse and replication of the data under the permission of the original authors.

Sensitive geospatial data may have some legal restrictions under the data protection laws.

In this research, no personal data will be used.

5.2. Ethical Aspects

Provide information on any issues that may arise during data collection, storage, sharing, release and publishing.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

Include: ensuring compliance with informed consent, scientific standards and research integrity, necessity of ethical review board, protecting identity of participants, storage and transfer of sensitive data to permanent storage, how to ensure respondents are not negatively affected by participating in the research project.

There is no personal data used in the research. Data storage systems will be used and have access control method to keep the data safe. Data containing sensitive information including urban population, sensitive places like military base may be restricted to prevent misuse or misrepresentation. A documented protocol will be in place to address potential breaches, including notifying affected parties and corrective measures.