TWIN2EXPAND

DATA MANAGEMENT PLAN. UPDATE.

DELIVERABLE 5.3

twinning towards
research excellence
in evidence-based planning
and urban design

Co-funded by the European Union

UK Research and Innovation

surf society and urban form research lab

TWIN2EXPAND

Document Description: This deliverable describes the updated plan for the management of data used, collected and produced through the project.

This deliverable constitutes Deliverable 5.3 for Work Package 5 of the TWIN2EXPAND project.

December 2024

General Contact:
TWIN2EXPAND Project, Dr. Nadia Charalambous, Department of Architecture, University of Cyprus, Nicosia, Cyprus

For more information: charalambous.nadia@ucy.ac.cy

Please refer to this publication as follows:
University of Cyprus, Department of Architecture (2023). Data Management Plan. Update. (Deliverable 5.3). Nicosia. TWIN2EXPAND project: 101078890 (Horizon Europe), 10052856 & 10050784 (UKRI).

This project has received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No. 101078890.

This project has received funding from the UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee under grant numbers 10052856 and 10050784.

Partners

University of Cyprus | UCL The Bartlett School of Architecture | CHALMERS UNIVERSITY OF TECHNOLOGY | Politecnico di Torino | Space Syntax

Version Control Sheet

Co-funded by the European Union | UK Research and Innovation | surf society and urban form research lab

Deliverable number	D5.3
Deliverable name	Data Management Plan
WP / WP number	WP5, Task 5.2
Delivery due date	31/12/2024 - Month 24
Type	R – Document, report
Dissemination level	PU - Public
Lead beneficiary	UCY - University of Cyprus
Contributors	All partners
Reviewer(s)	Nadia Charalambous
Changes with respect to the Grant Agreement	No changes
Dissemination and uptake	Internal use

Revision history

Version	Date	Summary of changes	Initials	Changes marked
1	19/12/2024	Draft Report	IGe	
2	30/12/2024	Final Review	NC	

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

Executive Summary

This TWIN2EXPAND updated Data Management Plan (DMP) describes the way in which the project consortium is managing the use and reuse of datasets from external sources, within the project and all datasets that will emerge from the project. The TWIN2EXPAND DMP follows the structure of the Horizon Europe DMP template reflecting the status of the data that will be reused, collected, processed, or generated. It details the practices adopted in terms of metadata and archiving to ensure that the data will be findable, accessible, interoperable and reusable (FAIR) for other potential users. Furthermore, the DMP provides information about the datasets and other research outputs that the consortium is aiming to preserve, specify the format type, the secure data storage systems, including the allocation of resources for the preservation of data and related data management activities.

This document replaces D5.2 Data Management Plan, submitted in M6, and its subsequent interim update D5.3_DMP_Update1_20240514.

This version of the DMP defines the policy and approach to data management in TWIN2EXPAND for handling data management related issues at the administrative and technical level, including data and metadata collection, publication and deposition of open data, the data repository infrastructure and compliance with the Open Access Infrastructure for Research in Europe (OpenAIRE).

Acronyms

CA = Consortium Agreement
CC = Creative Commons
CERN = European Council for Nuclear Research
CSA = Coordination and Support Action
Dn.n = Deliverable (followed by number)
DMP = Data Management Plan
DOAR = Default Open Access Repository
DOI = Digital Object Identifier
DPO = Data Protection Officer
EBDP = Evidence-Based Urban Design and Planning
EC = European Commission
FAIR = Findable, Accessible, Interoperable, Reusable
GA = Grant Agreement
GDPR = General Data Protection Regulation
HO = Host Organisation
IP = Intellectual Property
IPR = Intellectual Property Rights
OpenAIRE = Open Access Infrastructure for Research in Europe

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

ORCID = Open Researcher and Contributor ID
PC = Project Coordinator
POIs = Points of Interest
QA = Quality Assurance
SC = Steering Committee
SDG = Sustainable Development Goal
TB = Terabyte
UCY = University of Cyprus
WGS = Working Group Studio
WP = Work Package

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

Contents

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

Introduction

Open science practices are embedded within the project's activities through CSA8 (Active interaction with EBDP stakeholders and NGOs in Cyprus) and are designed, managed and monitored through WGSc (working group studio on research impact). The first WGSc was held ahead of the 1st stakeholder meeting, in M7 and M9 respectively, once the first Data Management Plan (DMP), as well as the Dissemination, Exploitation and Communication Plan (D5.4) were completed, after setting up a project identity and initiating engagement through communication channels for increased effectiveness of in-person meetings and stakeholders' input into research activities. TWIN2EXPAND's open science practices essentially relate to two aspects of the project: the objective of having a broader impact on society and the research framework, methods and findings linked to the research project.

To this extent, the following practices have been applied with regards to the research components of TWIN2EXPAND:

- Production of the present DMP, its regular revision and update.
- preregistration of the research project's research question(s), hypothesis and methodology on TWIN2EXPAND's web site and on a suitable preregistration platform. The registration platform selected for use in this project is OSF. This platform was selected because it provides templates appropriate for qualitative research projects. The project's DOI is 10.17605/OSF.IO/S5YM9
- documentation of any changes made in the methods during the course of conducting the study, published on the project's web site and selected platform; to date no changes have been necessary.
- open data for all non-confidential data and all models created for the project.
- Gold Open Access model for at least 4 project publications; Green Open Access model for publications beyond the target, released either through institutional repositories or through the EC's Open Research Europe.

The types of data generated by the project include summaries from workshops with stakeholders, contextual information, analyses of policy practices, analyses of case studies, ethnographic data of sites (space use, user make-up), including quantitative counts from observations, recorded material (photos, videos, sound), policy recommendations, conceptual frameworks, geo-referenced spatial models, geo-referenced data of the physical environment. The data will be stored on secure institutional computers and shared among the partners through a dedicated system. The PC will be responsible for the appropriate management and dissemination of project data. Open access shall apply under the terms and conditions laid down in the Grant Agreement (GA), securing the FAIR (Findable, Accessible, Interoperable, and Re-usable) principles, enabling greater real-world impact, which requires the open sharing of research data and findings. Therefore, all data will be described with rich metadata meeting all relevant domain standards and deposited in a trusted repository (requiring authentication and

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

authorisation steps), which will assign a DOI. The ORCID of the authors and contact information will be provided for all datasets. All spatial data will be in well-known and open formats and include a readme file containing a description of each file, the relationship between files and any tabular, visual or textual material in accompanying publications, data processing steps and description of any associated datasets. Data usage licence and provenance information will be provided along with documentation at the project-level (context, aim of study, hypothesis, methodology), file-level (relationship between files constituting a data set), and item-level (meaning of variables).

The legal framework of the Consortium and IPR issues are represented by the contract of the Partners with the EU Commission implemented by the Consortium Agreement (CA). Ownership, protection and guarantee of knowledge produced by the Consortium will be addressed throughout the project, aiming to identify IP generation and the measures needed to transform knowledge and IP into fair exploitation. IPR and Open Science training will be organised for all researchers involved in the project as part of the Research Management Training (RMT) organised by POLITO, ensuring that appropriate research practices are carried out and that the value of IP is not inadvertently reduced. The CA defines how all IPR issues will be managed inside the project, specifying arrangements for access rights to pre-existing know-how and knowledge, dissemination rules, and IPR applied among the partners and their affiliates. It also defines the principles of IPR Management adopted within the project (partner data set reference and name; data set description; standards and metadata; data sharing archiving and preservation, including storage and backup) and sets out provisions for background and foreground IP and the rights of each partner for their use within and after the project, as well as provisions for the settlement of any IP disputes within the partnership. All background IP in existence prior to the CA will remain the property of the originating party or parties; all foreground IP generated by the project will be defined by an exploitation agreement where needed and no partner shall derive unfair advantage in relation to exploitation.

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

Data summary

The research data within the scope of the TWIN2EXPAND project relates exclusively to *WP4 Research Project: Application of EBDP in Different Planning Contexts*. The assessment of the applicability of evidence-based urban design and planning (EBDP) will be performed through the analysis of case studies in the partners' countries (Cyprus, Sweden, UK). This requires the collation, collection, production, processing and storage of a variety of data sets relating to the urban form of the neighbourhoods/cities in which the case studies are located, the physical infrastructure (e.g. parcels/plots, building footprints), usage of the infrastructure (e.g. land uses), usage of space (e.g. traffic counts, pedestrian movement), socio-economics and demographics (e.g. census data), biodiversity (e.g. recording of animal activity) as well contextual and technical information relating to the case studies (e.g. project briefs, design drawings). With the exception of contextual information, all data will be geo-referenced for analysis and processed within the open source GIS software QGIS.
The consortium performed a data audit to identify existing datasets within the geographical scope of the project case studies to then set data collection procedures.

Data collection

Data collection (sources, methods and tools) is detailed within the OSF registration and subdivided according to the three interrelated research strands described in the Research Project Plan (D4.1). The data collection protocol for each strand is reported below.

1. A lightweight modelling approach for EBDP in the EU: enhancing replicability and comparative analysis

The data in the study comes from readily available datasets with coverage for the EU, with the prerequisite that these are open so that the derivative data and workflows can be shared or replicated. This consists of a combination of:

- Overture Maps: Several layers are used to define streets ("transportation"), places (Points of Interest), and buildings.
- Eurostat 2018 Urban Clusters dataset consisting of high-density urban clusters: To define town and city boundaries in a consistent manner.
- Eurostat 2021 Census grid population count: The population density per square kilometre.
- Urban Atlas Land Cover 2018: To define urban blocks. Blocks denoted as green space are used to define green areas.
- Urban Atlas Street Tree Layer 2018: To define tree canopies.

The underlying data is retrieved from readily available online sources through a combination of API access and download links. Python workflows are used to extract information

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

overlapping the EU high density clusters. The administrative boundaries are buffered to include a larger area avoiding edge effects that would otherwise trigger issues in the analysis.

Data is stored in Postgres databases with the PostGIS extension and is processed using Python and an ecosystem of open geospatial Python packages. Derivative data from the analysis steps is likewise stored in the database from where it is processed for statistical and other forms of downstream analysis.

2. Applicability of evidence-based design and planning (EBDP) in different planning contexts and scales

The project will leverage various forms of spatial, social, and environmental data associated with the built environment of the island of Cyprus. The data sources for existing data relied upon can be categorized into three main groups: Official European databases, including the European Environmental Agency, the Copernicus Land Monitoring Service, and Eurostat; national databases, such as the Open Data Portal for Cyprus, the National Statistical Service, and the National Cadastral Service; and the volunteer geographic information database OpenStreetMap. For original data sources, the project will rely on field work and stakeholders focus groups.

The project will employ data collection methods primarily focused on existing data sources, relying on open source, public databases and governmental institutions. The data will be sourced from various repositories, including those of the Republic of Cyprus, Eurostat, OpenStreetMap, and the Copernicus Land Monitoring Service. Additionally, non-open data will be purchased from official governmental sources to test the extent of the applicability and performance of spatial models in the context of Cyprus. Depending on the relevant scale, data will be collected from the open sources –or purchased from official agencies- with its related metadata or will be generated through the development of urban spatial models. These chosen data collection/generation methods align with the overarching objective of the project to test the applicability of Evidence-Based Design and Planning (EBDP) in contexts where data availability may pose challenges. As for original data, the project will collect spatial and social data through stakeholders focus groups for the case studies at all scales and through field work for the project scale only.

The project will employ a Data Audit that entails an examination of the quality, completeness and reliability of urban data available. The data audit serves as a tool to ensure the quality and appropriateness of the urban data available for Cyprus.
Data collection tools for original data for the case studies at all scales will rely on stakeholders focus groups and for the case studies at project scale only through field work.
1. Field work:

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

- Ground level photographic survey (descriptive)
- Arial level photographic survey (descriptive)
- Pedestrian and Cyclist movement counts (Van Nes and Yamu, 2021; Space Syntax Limited observations manual)
- Building level land use survey around the project scale case studies (1km buffer) (Guidelines and Taxonomy)
- Questionnaires

2. Stakeholders focus groups.

3. Evidence-based design and planning (EBDP) for and under climate emergency

This research will use different types of data such as existing GIS (Geographic Information System) data (building footprints, street lines, plot polygons, etc.) but also will seek to collect novel empirical data about species diversity, presence (e.g. captured from sound recordings), as well as human movement and activity (e.g. from mobile phone data).

The GIS data will be collected from open platforms such as OSM (Open Street Map) data and official sources in Sweden (e.g. Trafiklab). However, empirical data regarding species diversity/abundance and human movement using in-field low-cost sensors and mobile phone data will be collected.

For example, to collect empirical data about species diversity/presence, we will investigate the possibility of employing bioacoustics and deploy low-cost full-spectrum acoustic sensors in different green spaces in Gothenburg. The collected audio data can then be processed, and machine learning algorithms can be applied to detect different animal species based on their voicing frequency.

Date reuse

A number of existing datasets are reused for the purposes of this project:

- Space syntax models of the case study areas (previously generated by the partner institutions). These datasets will be used to perform spatial accessibility analysis of the case studies.
- Open access datasets of building footprints (e.g. EUBUCCO) and data sets under licence of parcels/plots, building footprints, land uses, etc. (e.g. datasets from relevant national agencies). These datasets will be used to perform land use analyses, correlational analyses with spatial characteristics, and for the creation of additional dataset of physical characteristics of the built environment (e.g. use of geometries from the datasets to store collected data)

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

- Open access datasets of road centre lines (e.g. OpenStreetMap). These datasets will be used to develop new spatial models of the case study areas where these are not already available open access, under licence or from the partner institutions.
- Copernicus data on density and tree coverage. These data will be used to perform ecological analysis of the case study areas.
- Open access socio-economic and demographic data (e.g. Eurostat). These data will be used to perform population reach and socio-economic analysis of the case study areas.

Data types and formats

The project reuses and generates primarily georeferenced quantitative data and categorised data linked to different geometries:

- Accessibility (space syntax values) data linked to line geometries.
- Points of Interest (POIs) data linked to point geometries.
- Land use and land coverage data linked to polygon geometries.
- Physical characteristics data (e.g. building height, frontage, access points, etc.) linked to relevant geometries for the purposes of visualisation: polygon, line or point. These data may be the product of aerial photogrammetry and the resulted point cloud.
- Space use data linked to line or point geometries.
- Socio-economic and demographic data, linked to polygon geometries at various scales depending on the resolution of available data for each case study, or linked to a raster grid where linking to geometries is not possible.

All the above data will be stored as GeoPackages (.gpkg) files.

The project also generates qualitative data, resulting from analysis of focus groups, policy documents or project-related documents, in text format stored as .txt files. It may, additionally, generate data in the form of photographic, video and sound material which will be stored as .jpg and .mp3 files.

All project deliverables in the format of reports and scientific articles will be stored as .pdf files.

Data structure and evolution

The geospatial models used and created by the project are of varying intensity: lightweight models and heavyweight models.
For the lightweight models, the data transitions through different stages:

1. Data acquisition: raw data collected from open sources
2. Data wrangling: network cleaning and establishment of data schemas and finally overall data organisation
3. Data processing: metrics such as centralities, densities, statistics, etc. for various units of analysis and scales of aggregation

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

4. Data post-processing: statistical data analysis, models, visualisation requirements
5. Final data outputs (e.g. plots, GeoPackages)

For the heavier-weight models, the data transitions through similar stages, but the types, units, scales of aggregation, etc. respond to the research questions developed for each research strand. In some cases, the transition requires manual editing by researchers following a protocol based on an existing body of literature. In these cases, the related metadata provide references to the literature and identify the researchers who performed manual editing through their ORCID.

Purpose of data generation and reuse

The purpose of data generation and reuse is to provide a comprehensive, spatially detailed and accurate description of the built environment of the case study areas to analyse their characteristics and qualities in terms of social, economic and ecological performance. The types of data used and generated enable the measurement of key performance indicators including centrality, accessibility and density, also allowing for comparative analysis of the case studies, as well as for the development of descriptive and predictive statistical models. Linking the quantitative data and analyses to the contextual information and qualitative data will enable the consortium to respond to the key research questions relating to the applicability of EBDP in the different case study contexts, its relevance towards achieving SDG11, the identification of EBDP tools best suited to different types and scales of projects given a specific level of data availability and quality. The data will therefore also enable and be used within the production of policy briefs (D4.4) including infographics of the research.

Data size

The consortium estimated that the total size of data generated and reused will amount to a maximum of 2TB.

Data origin

The origin of the data used in the project are:

- Data from partner institutions
- Data from open access sources
- Data from national databases/agencies
- Data from focus groups
- Data from analysed texts
- Data collected from fieldwork.

Data utility

The data is expected to be useful to different end users:

- Academics, researchers and students in the same or related fields

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

- Practitioners working on the case study projects or case study areas
- Local authorities of the case study areas
- Relevant national authorities (e.g. planning departments, transport departments)

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

FAIR data

For making the datasets FAIR, this DMP has considered and taken into account standards and principles based on (Wilkinson et al, 2016)[1], the OpenAIRE[2] project and technical infrastructure, the European Open Science Cloud (EOSC)[3], the FOSTER[4] portal and the GOFAIR initiative[5].

Making data findable, including provisions for metadata

In order to make the datasets findable, the PC will ensure that all metadata will have a persistent identifier, rich metadata and are searchable and discoverable online. Therefore, the datasets will be assigned a Digital Object Identifier (DOI) in order to make content easily and uniquely citable.

Furthermore, the datasets will be accompanied by complete, rich, machine readable and persistent metadata in order to be easily findable by both machine and computer, with information to the origin of the dataset, in order to be findable even in the case the actual data are deleted. The metadata will also be assigned a globally unique and persistent identifier (PID) (eg. DOI) and will clearly and explicitly include the identifier of the data they describe.

The DOI for datasets will be generated through the project Default Open Access Repository (DOAR). The consortium has selected Zenodo as the DOAR. The datasets will be assigned a DOI through the DOAR, while the DOI for 'gold' open access publications will be assigned by the publishers of the project research papers.

All file names will include at least a version number and/or a timestamp. All project deliverables will use the following naming conventions: T2E_[number of deliverable]_[acronym of deliverable]_[version date in yyyy/mm/dd format or FINAL]; e.g. T2E_D5.3_DMP_Update_FINAL. An associated metadata file will detail the meaning of the acronyms.

All TWIN2EXPAND results deposited in the DOAR will provide search keywords together with their metadata. Keyworks for open data will be selected from controlled vocabularies that are suitable for the specific type of data. With regards to geospatial datasets, the consortium is considering following the EN ISO 19115 standard, following the guidelines provided in the document "INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO

[1] Wilkinson, M., et al. (2016). "The FAIR Guiding Principles for scientific data management and stewardship", Scientific Data 3. https://doi.org/10.1038/sdata.2016.18
[2] https://www.openaire.eu
[3] https://eosc-portal.eu
[4] https://www.fosteropenscience.eu
[5] https://www.go-fair.org

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

19115 and EN ISO 19119"[6]. Regarding text, photographic and video data, since there are no specific disciplinary standards, the consortium will use consider using the DOAR's deposition metadata domain model; in the case of Zenodo, this is based on DataCite's metadata schema[7]. All preliminary, but not finalised data generated during the course of the project are accompanied by "readme" style metadata[8], stored as a .txt file which describes key aspects of the data. An example of metadata for a dataset generated by the project based on open-source data is provided in Appendix I, formatted for readability.

The metadata of each record will be indexed and searchable directly in the DOAR immediately after publishing; Zenodo automatically sends metadata to DataCite servers during DOI registration for indexing.

Making data accessible

To make the datasets findable, the PC will ensure that all metadata are retrievable by their identifier using a standardised communications protocol: the protocol will be open, free and universally implementable, and it will allow for an authentication and authorisation procedure where necessary and finally metadata will be accessible, even when the data are no longer available (eg. Out of date dataset, but metadata linking to origin).

Therefore, all data generated by the project (foreground IPR) which does not include sensitive information (individual personal data) will be made openly accessible via the DOAR, which will automatically assign a DOI, if this has not already been issued by a publisher, as detailed in the section above. Data and metadata will be available for the lifetime of the repository; in the case of Zenodo this is currently defined to be of at least 20 years.[9] Metadata will remain accessible after the data are no longer available, thus the datasets will remain accessible through replicable scripts run against open data sources.

Any data which contains personally identifiable information (e.g. from interviews, photographs or videos) is collected only when informed consent has been granted by the research subject (see section on Ethics), and will be made accessible only in anonymised format. If consent to public access in anonymised format is not granted by the research subject, the data will then be stored exclusively at the partners' premises in password-protected drives and deleted 6 months after the end of the project. Under these circumstances only the related metadata will be deposited in the DOAR.

All data reused by the project (background IPR) will be detailed in a 'list of background' including both backgrounds owned by project participants and rights owned by third parties.

[6] https://inspire.ec.europa.eu/documents/inspire-metadata-implementing-rules-technical-guidelines-based-en-iso-19115-and-en-iso-1
[7] https://schema.datacite.org/
[8] https://www.openaire.eu/what-is-metadata
[9] https://about.zenodo.org/principles/

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

When open data or any dataset covered by background IPR are used for the generation of new data, the resulting data will be made open access; the metadata describing all workflows will be fully open access.

Zenodo registers a DOI representing the specific version of a record and a DOI representing all of the versions of a record (Concept DOI). The Concept DOI resolves to landing page of the latest version of the record.[10]

The metadata will include information to enable the user to access the data as well as reference about the software needed to read the data. At the present stage, all data formats are accessible and readable through well-known open access software.

Embargos on the project data are not currently envisaged. Likewise, it is not expected that a data access committee will be necessary, as any data from focus groups or fieldwork (videos, photos, sounds) comprising personal information will be anonymised. If any issues regarding restricted access to anonymised data arise, the TWIN2EXPAND's Steering Committee (SC) will act as data access committee and seek clarification. Access to the data and verification of user identity will follow the standardized access protocol of the DOAR.

Making data interoperable

In order to make the datasets interoperable, the PC will ensure that all metadata will use a formal, accessible, shared and broadly applicable language for knowledge representation; the metadata will use vocabularies that follow FAIR principles and they will include qualified references to other metadata.
The exact form and schemas of the datasets has developed organically during the project: geospatial data is interoperable through the use of open-source geospatial tools (e.g. GDAL / GEOS based python stack) and open data formats (e.g. GeoPackage). During the project data is identified and collected, all open data sources, dates of access and data licences are documented.

In the project, the consortium uses standard file formats and every dataset has metadata following selected disciplinary standards (the Zenodo domain model standard). TWIN2EXPAND did not introduce new project specific ontologies or vocabularies to date, and it is not expected to do so. For any datasets generated by the project built on an existing dataset, a qualified reference will be provided.

Increase data re-use

[10] Resolving in Zenodo is expected to change to a landing page specifically representing the concept behind the record and all its versions. See https://help.zenodo.org/

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

All datasets will be deposited with metadata in the form of 'readme' files following agreed standards as described above, including comprehensive information on software, data usage licence and provenance information along with documentation at the project-level (context, aim of study, hypothesis, methodology), file-level (relationship between files constituting a data set), and item-level (meaning of variables, units of measurement, etc). Documentation of replicable workflows, use of open tools and data formats will facilitate use in downstream research.

All datasets generated by the project (foreground), not built upon previous datasets, will be made freely available in the public domain under a Creative Commons (CC-By Attribution 4.0 International) licence. As such, the data produced will be usable by third parties even after the end of the of the project for as long as the data remain available, aligned with the protocols of the DOAR, which, in the case of Zenodo, is expected to be at least 20 years.

Quality Assurance in TWIN2EXPAND is covered through WP1 Project Management and Coordination and detailed in the Project Management Plan (D1.1). The quality assurance procedure which applies to project deliverables will also apply to datasets to ensure that the right version of the dataset is used at the right time and place. Therefore, the metadata will also include the following QA information:

- **Done by**: Appropriately qualified member of staff working on the project that has worked on the document. If different members of staff have worked on the document, then all names must be listed here. This information will also be recorded through the DOAR by linking the researchers' ORCID to the datasets.
- **Quality assured by:** The quality assurance of the dataset happens at a WP Leader level first. The WP Leader reviews the contents and quality of the datasets and metadata to be in compliance with the agreed standards.
- **Approved by**: The final approval is completed by the Project Coordinator (PC) who reviews that that the datasets and metadata are in compliance with the requirements of the Grant Agreement and GDPR standards. Once approved, the datasets are uploaded on the DOAR.

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

Other research outputs

Other research outputs from the project include only deliverables in the form of reports (digital pdfs). At present, no other types of outputs are expected from the project.

The other research outputs which will be deposited along with metadata in the DOAR, following the same standards for the datasets, are the following:
D3.1 Scientific paper: State-of-the-art of EBDP, conceptual model and methods
D3.2 Report on EBDP conceptual model and methodologies
D4.2 Scientific paper: Applicability of EBDP in different planning contexts
D4.3 EBDP Matrix
D4.4 Policy Briefs
D4.5 Report on the applicability of EBDP
D5.6 Scientific paper: Knowledge Translation
D5.7 Non-academic publication

The three scientific papers and non-academic publication will be published with the 'gold' open access model and arrangements will be made with the publishers to deposit pre-prints in relevant repositories as soon as possible after acceptance of the paper by the journal. If any further scientific articles result from the project research activities, these will follow the 'green' open access model for publication and be released through institutional repositories and through the EC's Open Research Europe.

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

Allocation of resources

Direct costs for making the data FAIR are estimated to be zero. The selected DOAR (Zenodo) is free of charge. Costs related to open access for research publications are eligible for reimbursement during the duration of the project; these costs have been included in the project budget and estimated to be 16,000.00€ in total for the four publications (D3.1, D4.2, D5.6 and D5.7).

Data management activities concern the whole project and need to be coordinated and monitored both at project and work package level. Data management is also linked to publication of project results and thus dissemination activities. Therefore, the following roles and responsibilities are identified:

1. The **Project Coordinator**, Nadia Charalambous, is responsible for:
    - Approving all versions of the DMP
    - Approving the final versions of data sets and metadata
    - Approving the final version of deliverables to be deposited in the DOAR

2. The **Project Data Manager** (WP4 Lead; Task 5.2 Lead), Meta Berghauser-Pont, is responsible for:
    - Quality checking all versions of the DMP
    - Coordinating the technical realisation of WP4, including data collection, data generation, data storage, etc.
    - Monitoring data management activities (both collection and publication) and deadlines
    - Ensuring that GDPR policies are applied to dataset
    - Providing support to WP Data Officers
    - Providing data management solutions for specific issues in accordance with open science practices, the DMP, the CA and the GA

3. **Work Package Data Officers** [WP3: Gareth Simons; WP4: Iason Giraud] are responsible for:
    - The implementation of the data management policy in their respective WPs
    - Supporting the PDM in monitoring data management activities
    - Creating the metadata for datasets and deliverables relating to their WP
    - Ensuring that results and metadata are deposited in the DOAR
    - Monitoring that results are properly interlinked and cross-referenced
    - Raising any ethical and privacy issue that may forbid the publication of data to the PDM

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

4. The **Dissemination Manager** (WP5 Lead), Ilaria Geddes, is responsible for:
    - Offering assistance in choosing the right publication path for 'gold' open access and depositing pre-prints
    - Offering help and further guidance for publishing, disseminating and communicating scientific publications

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

Data security

Data collected and used in different locations will be stored on partners' institutional servers, which all have protocols for data security and recovery. For any data which may include identifying information, the following security protocol and process will be applied in order to anonymize as quickly as possible and produce a secondary dataset that is safe and abides to FAIR standards:

- If taking video or sounds s for the purposes of collecting data on movement or biodiversity then the faces or sound will be blurred at first opportunity, analysis then performed by researchers in a restricted environment, then output data aggregated to CSV files per days / times / counts. Aggregation helps protect privacy.
- If doing surveys, the surveys should use anonymous identifiers and should strictly steer clear of sensitive data. Results likewise tabulated to CSV files in synthesised / aggregated form.
- In both cases, the raw data never becomes public and is deleted as soon as transformed into CSV to remove longer term risk.
- While being processed, the datasets will be stored on secure institutional servers hosted in the EU. Researchers' access will use strong passwords with two factor authentication for access. Dataset logging will be encouraged: e.g. keeping track of who has been given access and under what conditions, and logging when the datasets have been deleted once done.

The output of aggregated data as CSV files could be assigned metadata and shared more publicly with reduced concern for personally identifiable information.

The consortium aims at data versioning with secure daily data backups and retention of backups for a set duration (e.g. 30 days). Sensitive data will not be shared electronically among the partners.

All open data will be stored in the trusted repository for long term preservation and curation. If any data including personal information is collected, these datasets will be anonymised to comply with GDPR regulations (see section below on Ethics). The personal information of any participant will be stored separately from any questionnaire, survey, observation or assessment, and it will be identified with the collected measures only based on a unique code. The information linking the code with the name will be stored in a secure, password-protected, electronic repository at each institution which collected the data, to which only one appointed person in each partner institution will have access. Any such dataset including personal information will be deleted as soon as they are no longer needed for the purposes of data analysis.

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

Ethics

All research related to this project will be undertaken with scientific integrity and responsibility in mind. Based on the HORIZON guidance on ethics and data protection[11] the only ethical issue arising in this process entails the collection of data which may include identifying personal information in particular through interviews with stakeholders and/or public space users, or through recording of video or photographic material of space use in public spaces. At present, it is not known if and the extent to which such data will be collected, as this will be defined in the Research Project Plan (D4.1) due in M8. If personal data are recorded for research purposes, the following procedures will be followed:

Procedures for obtaining informed consent: Each project participant will be provided detailed information about what will happen to the personal data that they will be collected and will complete an informed consent form, which provides information about the right to refuse consent, to withdraw at any time, and appropriate authorities, unrelated to the study, to which complaints can be directed, including contacts at the Host Organisation Authorities. The right to refuse consent and to withdraw at any time will be provided, along with information about appropriate authorities, unrelated to the study, to which complaints can be directed, including contacts at the Host Organisation (HO), University of Cyprus, the project's coordinating country and organization.

Storage of data including personal information: In order to ensure the safety of sensitive data, a secondary database will be created based on pseudonymisation or anonymisation practices. The personal information of any participant is stored separately from any questionnaire, survey, observation or assessment, and it will be identified with the collected measures only based on a unique code. If it is necessary to retail a link between the research subjects and their personal data, the information linking the code with the name will be stored in a secure, password-protected, electronic repository, hosted and managed by UCY, to which only one appointed person in each partner institution will have access. This will ensure necessary and effective data sharing among the partners without compromising anonymity of the participants or increasing data risks by storing the data in multiple locations. The database linking participant identifying information to alphanumeric codes will be password-protected and stored separately. The participants' names will not be mentioned in any form or manner throughout the stages of research or publication, and the personal identity of the data will be secured and protected. The researchers will inform participants of the privacy and data protection measures being taken. Only non-identifiable data will be shared, and this data will be shared in an aggregate form.

[11] https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

All the data involving personal information will not be kept longer than needed; all data will be archived with highest safety and confidentiality. Moreover, all the necessary precautions will be taken to protect and safeguard vulnerable populations. Under the GDPR rules, personal data will be processed in accordance with principles and conditions that aim to limit the negative impact on the people involved and ensure fairness, transparency and accountability of the data processing, data quality and confidentiality.

Depositing datasets: only anonymised data in aggregated format may be shared. The datasets will be anonymised following a standardised protocol (e.g. through Amnesia[12]).

[12] https://amnesia.openaire.eu/

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

Other issues

UCY has an appointed a Data Protection Officer (DPO), who advises on and preapproves that the security of personal data and the protection of all data subjects' rights in relation to the processing of their personal data under the project is fully compliant with GDPR. The DPO will assure all personal data collection and processing by the Consortium will be carried out according to EU and national legislation.

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

Appendix I: Metadata example

Title	Nicosia's non-motorised segmented street network - manually edited	
Description	A dataset containing spatial and attribute data for non- motorized segmented street network	
Keywords	GIS, Spatial analysis, Street network, walkability	
File Name	Nicosia_NM_network_sgm_ME_241107	
Date created	01 November 2023	
Last update	07 November 2024	
File Format	Geopackage .gpkg	
Size	16.8 MB	
Schema	Field name Data type Description	Street_id Integer Unique identifier for each street segment
Associated File(s)	Nicosia_NM_unlinks_ME_241107 Nicosia_NM_network_ME_241107	
Coordinate System	CGRS93	
Version	1.0	
Source	SURF Research Lab	
Licence	Open Data Commons Open Database License (ODbL)	
Project Identifier	DOI 10.17605/OSF.IO/S5YM9	
Data Identifier	To be added after dataset DOI is created	
Contact	charalambous.nadia@ucy.ac.cy	
Citation / Credit	Ricchiardi A. and Geddes I. (2024) *Nicosia's non-motorised segmented street network - manually edited; V1.0.* SURF Research Lab, Nicosia, Cyprus. / ©SURF Research Lab 2024	

Authors	**Authors ORCID**
Ana Ricchiardi	0000-0002-6418-5142
Ilaria Geddes	0000-0002-4463-8597
Quality Assured (QA) by	**QA ORCID**
Approved by Project Coordinator (PC)	**PC ORCID**
Nadia Charalambous	0000-0003-1977-2599

File name acronyms:
NM = non-motorized
sgm = segmented
ME = manually edited

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

Zone extent:
The extents of the dataset refer to the Urban Morphological Zone (UMZ) of Nicosia and a buffer of 20 km around it.

Non-motorized Network:
The non-motorised network dataset is based on Open Street Maps (openstreetmap.org, http://download.geofabrik.de, date of download 01-11-2023).
We include all streets and paths that are accessible for people walking or cycling, including those that are shared with vehicles. All streets where walking or cycling is forbidden, such as motorways, highways, or high-speed tunnels, are not included in the network.

Editing and cleaning:
The original road-centre line map was edited based on the basic representational principles of Space Syntax Methodology. We followed the same editing and generalizing procedure to remove errors and increase comparability between networks. This process included removing duplicate and isolated lines, snapping and generalizing. The aim was to create appropriate line-segment maps to be analysed using Angular Segment Analysis, a network centrality analysis method introduced in Space Syntax.

The road-centre line map was cleaned and simplified using the Space Syntax Toolkit (SST) version 0.3.10 for QGIS (version used LTR 3.28) using the default settings, and manually edited by researchers to ensure that:

- all roads and paths were represented as one line irrespectively of the number of lanes or type, meaning that parallel lines representing a street and a pedestrian or a cycle path running on the side, are reduced to one line[13]
- all footpaths were included in the models
- roundabouts and intersections were represented correctly
- the unlinks[14] produced by the SST were represented correctly

[13] The reason is that these parallel lines are nor physically or perceptually separated, and thus are accessible and recognized from pedestrians as one “line of movement” in the street network. If there are obstacles or great distance between parallel streets and paths, then the multiple lines remain. This representational choice follows the Space Syntax methodology in representing the public space and the street network.

[14] The network (cleaned and simplified road-centre line map) layer is complemented with an “Unlink points” layer; a GIS point layer with the locations of all non-level intersections, such as overpasses and underpasses, bridges, tunnels. The Unlink point layer is necessary to generate the segment model to conduct network analysis that takes into account the non-planarity of the street network.

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab

The road-centre line map was then processed with the Place Syntax Toolkit version 3.2.5 to create line segment maps which could be used in the analysis.

Co-funded by the European Union
UK Research and Innovation
surf society and urban form research lab