European Research Council
Executive Agency

Established by the European Commission

Horizon 2020
European Union Funding
for Research & Innovation

European Research Council (ERC)

urbisphere Data Management Plan

(ERC-SyG)

Version 2.0
23 September 2020

European Research Council
Established by the European Commission

ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

Project Acronym	Project Number
urbisphere	855005

SUMMARY

DMP for ERC Synergy project *urbisphere* Grant Agreement number 855005

- Appropriate management of research data is important to maximize the impact of the project. The basic elements of the DMP have been agreed by the *urbisphere* team. *urbisphere* participates in H2020 Open Research Data Pilot (ORDP).
- *urbisphere* aims to follow the FAIR Data Principles (i.e. to make data findable, accessible, interoperable and reusable).

- New data will be generated using many methods: (1) collection (2) analyses (including modelling).
- Examples of methods and analyses include: Earth Observation (EO) data, instruments and interview questionnaires, auxiliary data (e.g. census, cadastral), numerical model and analytical workflows. All are related to the objectives of the project and are inherently part of *urbisphere*.

- Examples of data that to be collected and used are:
  1. public (e.g. USGS/NASA, ESA, EEA, DLR, Met Office, LAQN, Meteo France, DWD, local city government agencies),
  2. commercial providers (e.g. Planet, MAXAR, AirBUS, Netatmo),
  3. field campaigns that are of both short (e.g. by drone, interviews) and long duration. The latter will collect time series data from a variety of in situ sensors (e.g. atmospheric LiDARs, flux towers, weather stations, radiation, air quality).

- Data generated using new methodological approaches will be submitted for peer review, typically via papers, to ensure that the methods are robust (i.e. published in scientific articles).
- Datasets for all papers will be archived/published separately (e.g. https://about.zenodo.org/).
- Other communications via scientific conferences, workshops and wider audience (newspapers, magazines, TV, radio social media) will occur. An archive of their occurrence will be made.

- The Data Management Plan is a ''living'' and ''active'' document that will be updated as the research progresses, and new data are produced within *urbisphere*.

1. Making Data Findable

- Datasets will be generated by the *urbisphere* team to address the project's objectives. *urbisphere* Principal Investigators (PIs) will discuss and agree what and when data will become open access and findable, as described in the *urbisphere* Collaboration Agreement (art 8, hereafter CA). Data will be made "as open as possible, as closed as necessary".

- Each final dataset will have a unique persistent identifier **UPI** (e.g. DOI, ARKs, Handles, ORCiDs) with rich metadata. There are several initiatives for assigning dataset DOIs (e.g. Zenodo).

- Data and code (where appropriate) that underpin peer-review articles will be archived in **searchabl**e repositories (e.g. Zenodo).
- Keywords will be used that relate to the project (e.g. *urbisphere*), place (e.g. city name) and process (e.g. temperature, exposure) to facilitate the use of searchable repositories.
- Database of the following with be made accessible via the website (urbisphere.eu):
  1. publications,
  2. archived/ published datasets with locations,
  3. GitHub sites.

2. Making Data Openly Accessible

Within *urbisphere there are the following d*ata types that need to be considered with respect to open-access:

1	**Commercial** data (e.g. satellite)
2	**Proprietary** data from external to *urbisphere* sources (e.g. cadastral data, aerial imagery, LiDAR, in situ data from agencies)
2.1	Publicly available
2.2	Shared confidentially to the *urbisphere* team
3	Sensitive data collected within *urbisphere* (e.g. parts of household interviews)
4	All other data collected or created within *urbisphere* that are not ethically sensitive

- **All datasets (data, code, etc) associated with *urbisphere* scientific publications will be archived/** published **(e.g. zenodo).**
- *urbisphere* peer-reviewed articles generated: will be published either as gold (open access publishing) or green (self-archiving of the accepted but not yet published manuscript) open access publications
- Research data created in the project are owned by the *urbisphere* PIs (as per the CA)

- Data types #1-3 have a range of data protection rules or ethical standards or licensing agreements that require the data to be retained in a protected manner (refer to the CA Art. 8).
- Data Type #4 will become openly accessible (refer to the CA Art. 8).

Example Repositories that will be used for making data available

1	Zenodo	*Zenodo.org (accessed 23 Aug 2020)* *Why use Zenodo?* • Safe — your research is stored safely for the future in CERN's Data Centre for as long as CERN exists. • Trusted — built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science. • Citeable — every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable. • No waiting time — Uploads are made available online as soon as you hit publish, and your DOI is registered within seconds. • Open or closed — Share e.g. anonymized clinical trial data with only medical professionals via our restricted access mode. • Versioning — Easily update your dataset with our versioning feature. • GitHub integration — Easily preserve your GitHub repository in Zenodo. • Usage statistics — All uploads display standards compliant usage statistics
2	centaur	https://centaur.reading.ac.uk/ • Green publications will be lodged here
3	GitHub	https://github.com/ • Code used for processing data will be archived into GitHub repositories. • This allows version control and links to Zenodo
4	ALUFR	• a private (i.e. protected) mySQL database for real-time visualization, storage, quality control and processing of research data collected from in-situ sensor systems and ground-based remote sensing systems operated as part of *urbisphere* field campaigns. • rich web-based meta-data description system and interacts through an API with various open access programming languages (e.g. Python, R). • From the database, we will create final, version-controlled datasets in open (.csv or netCDF) formats for storage on Zenodo.
5	USTUTT	• University of Stuttgart data repository (https://darus.uni-stuttgart.de) • Provides flexibility and richness of the metadata • share data within the project without publishing, but without the github-Integration feature of Zenodo. • https://www.izus.uni-stuttgart.de/en/fokus/darus/.
6	FORTH	• A private (i.e. protected) data server for storage of proprietary spatial data (i.e. cadastral, satellite observations, drone and aerial imagery) • From the data server, we will create, final geospatial data as associated data supplement scientific publications for storage on Zenodo

3. Making Data Interoperable

- *urbisphere* will adhere to appropriate global data formats as much as possible.

Metadata will:

1. be indexed in a catalogue carried with the data,
2. be compliant with INSPIRE (https://inspire.ec.europa.eu/),
3. key words: will use appropriate vocabularies such as:
   - GEMET (https://www.eionet.europa.eu/gemet/en/themes/),
   - EnvThes (https://www.ecopotential-project.eu/images/ecopotential/documents/D5.6.pdf),
4. acknowledgements will include European Union (EU) ERC Synergy Grant *urbisphere* (855005).

Data Formats will:

- be readable from a large range of open-access software (e.g. QGIS, R, Python, Fortran etc.).

Directories and File Names will:

1. have standard conventions that consider City/location/year/variables etc.,
2. be documented on GitHub sites,
3. details of format will be given in all data sets metadata.

Types of data

Class of data	**Data standards**	**Software**
Interviews	csv and/or netCDF (tabular)	SPSS, R
Earth Observation, Spatial data	• GeoTIFF (raster) • ESRI shapefile and/or GeoPackage open standard (vector) • csv and/or netCDF (tabular)	QGIS
Atmospheric observations	netCDF, ascii, csv, proprietary to instrument, SQL	Python, R, scripts, SQL
Model outputs	netCDF, ascii, csv	
Socio-economic data	csv and/or netCDF (tabular), GeoTIFF (raster)	SPSS, R, QGIS

4. Increase Data Re-Use

- Datasets will be available for re-use through both private and public repositories (see section 2).

- Open access data license: Exact one will be decided later (e.g. Creative Commons (CC)).
- Scientific publications:
  1. papers will be made available gold or green access,
  2. papers will be announced via social media with #urbisphere.

- Publication data and code: will be archived/published for all publications (e.g. Zenodo).

- Metadata (see also section 3) will:
  1. be supplied with data,
  2. include relevant paper references,
  3. describe quality assurance processes, precision, and uncertainty.

- Selected systems (e.g. Sun Photometer, ceilometers) will be incorporated into international networks (e.g. AERONET, EPROFILE) if possible and manageable to ensure re-use of the data in global climate research and data assimilation systems.

5. Allocation of Resources and Data Security

Allocation of resources
Cost for hardware (e.g. servers, data storage), software, Journal publication are included in the project budget.

Responsibilities for data management in the project
Each *urbisphere* PI is responsible for data management in their area of expertise.

DS – dedicated servers; DSt – Dedicated storage

Group	Responsible for	Comments	DSt/DS
FORTH	EO data	versioning on the secure data server	DSt
USTUTT	socio-economic data	e.g. expert/household interviews, statistical offices e.g. in the pilot cities	University systems
ALUFR	instruments	• continuous operation of near real-time communication, processing, storage and management systems to handle data from atmospheric observations collected with *in-situ* sensors, networks and ground-based remote sensing sensors • Develop and run mySQL database with web-based query and visualization products	urbisphere.uni-freiburg.de
UREAD	Modelling code	• Development and running computer models • GitHub – version control of software development	University systems

Data security
All *urbisphere* data, apart from other repositories, will be secured in decentralized internal data storage facilities with daily backup. Therefore, data recovery will be possible at any time and all the data will be safely stored in for long term preservation and curation. Selected core systems will be equipped with georedundant backup. Publically available data from the scientific publications will be deposited on Zenodo and therefore data security in terms of recovery, storage and safety are ensured. This provides options for making data openly available and other data restricted access as required.

Ethical aspects
- Ethical or legal issues can impact data sharing. Raw sensitive data cannot and will not be made publicly available.
- Data from that may be sensitive because of household and/or cohort data (that might encompass direct and indirect personal data) for example in the context of interviews, indoor sensors and wearable sensors.

These types of data files will have:
1. restricted access within *urbisphere,*
2. be pseudo-anonymized as outlined by agencies such as ZENDAS,
3. to secure confidentiality and regulate access to data, a system with login names and passwords will be implemented. Histories of data entry and modification will be recorded.

- Interviews (e.g. household interviews), indoor measurements and wearable sensor deployments – participants will all be voluntary.
- Data collection will be conducted on a fully free informed consent.
- Informed consent forms and information sheets (in language and terms intelligible to the participants) templates under EU GDPR regulation of will be prepared prior to the release of the questionnaires or any sensing systems.
- Data and answers from interviews / questionnaires will be confidential (name of individual will not be recorded).
- Tracking the answer to a specific person is not possible.
- To ensure the methodology meets the ethical considerations and the protection of personal data, the appropriate committees (e.g. the Ethics Commission of the University of Stuttgart, ZENDAS) will be involved in ethics-related matters and approval sought where required.
- Detailed description of the ethical aspects will be provided in the Ethic deliverables (due M12).

DISCLAIMER. Please note that the ERC Data Management Plan is not a part of the Ethics Review. It is the responsibility of the Principal Investigator to inform the ERCEA Ethics Team of any ethics issues/concerns regarding the collection, processing, sharing and storage of data in relation to the project. The Principal investigator can also be asked to submit an Ethics Data Management Plan (Ethics DMP).