Action full title:

Universal, mobile-centric and opportunistic communications architecture

Action acronym:

UMOBILE

Deliverable:

D.6.10 - Data Management Plan

Project Information:

Project Full Title	Universal, mobile-centric and opportunistic communications architecture
Project Acronym	UMOBILE
Grant agreement number	645124
Call identifier	H2020-ICT-2014-1
Topic	ICT-05-2014 Smart Networks and novel Internet Architectures
Programme	EU Framework Programme for Research and Innovation HORIZON 2020
Project Coordinator	Prof. Vassilis Tsaoussidis, Democritus University of Thrace

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

Deliverable Information:

Deliverable Number-Title	D.6.10 - Data Management Plan
WP Number	6
WP Leader	Francesco Amorosa and Angela D'Angelo
Task Leader(s)	Angela D'Angelo
Contributing Partners	AFA Systems, DUTH, FON, Senception, Copelabs
Authors	Angela D'Angelo, Sotiris Diamantopoulos, Alberto Pineda, Rute Sofia, Paulo Mendes
Contact	dangelo@afasystems.it
Due date	31/07/2015
Actual date of submission	31/07/2015

Dissemination level:

PU	Public	X
CO	Confidential, only for members of the consortium (including the Commission Services)	
CI	Classified, as referred to in Commission Decision 2001/844/EC	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

Document History

- **Version:**1.0
  - **Date:** July 24th
  - **Description:** First draft to the consortium for comments and feedback
- **Version:**1.1
  - **Date:** July 30th
  - **Description:** Updated with partners comments and feedback. Final Version.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

1 Executive Summary

Open Access Model garantees free access for users and free dissemination of knowledge. UMOBILE participates in the "Pilot on Open Research in HORIZON 2020": participating projects are required to develop a Data Management Plan (DMP), in which they specify what data will be open.

This Data Management Plan explains which of the research data generated in UMOBILE will be made open, how data will be shared and which procedures will be put in place for long-term preservation of the data.

Following "Guidelines on Data Management in Horizon 2020", the DMP clarifies that scientific generated research data will be easily:

1. Discoverable
2. Accessible
3. Assessable and intelligible
4. Useable beyond the original purpose for which it was collected
5. Interoperable to specific quality standards

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

2 Open access to scientific publications

Open access to scientific publications refers to free of charge online access for users. Open access wil be achieved through the following steps:

1. Any paper presenting the project results will acknowledge the project: The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124-UMOBILE and display the EU emblem.
2. Any paper presenting the project results will be deposited at least by the time of publishment to a formal repository for scientific papers. If the organization hasn't a formal repository (`https://www.openaire.eu/participate/deposit/idrepos`), the paper can be uploaded in the European sponsored repository for scientific papers: `http://zenodo.org/`.
3. Authors will ensure that the publisher accepts open access via self-archiving in their departments formal repository or via `http://zenodo.org/`. Usually they do accept; if not, they will try to negotiate with them. In case of no success they will not publish via self-archiving.
4. Authors can choose to pay "author processing charges" to ensure open access publishing, but still they have to deposit the paper in a formal repository for scientific papers (step 2).
5. Authors will ensure open access via the repository to the bibliographic metadata identifying the deposited publication. More specifically, the following will be included:
   - The terms "European Union (EU)" and "Horizon 2020";
   - "Universal, mobile-centric and opportunistic communications architecture-UMOBILE", Grant agreement number 645124;
   - Publication data, length of embargo period if applicable; and
   - A persistent identifier.
6. Each case will be examinated separately in order to decide if self-archiving of paying for open access publishing.

3 Open access to research data

Open access to research data refers to the right to access and re-use digital research data generated by projects.

EU expects funded researchers to manage and share research data in a manner that maximizes opportunities for future research and complies with best practice in the relevant subject domain, that is:

- The dataset has clear scope for wider research use
- The dataset is likely to have long-term value for research or other purposes

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

- The dataset have broad utility for reference and use by research communities
- The dataset represents a significant output of the research project

Openly accessible research data, generated during UMOBILE project, will be accessed, mined, exploited, reproduced and disseminated free of charge for the user. Specifically, the "Guidelines on Data Management in Horizon 2020" clarifies that the beneficiaries must:

- *(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:*
  - *(i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;*
  - *(ii) other data, including associated metadata.*

It is useful to categorize the data as in the following table (which provides also an exampe of the dataset).

Category	Description	Examples
Raw Collected Data	Obtained data that has not been subjected to any quality assurance or control	Measuments collected from devices (Hotspots, Smartphones, UAVs, Videocameras, . . .)
Validated Collected Data	These are the raw data that has been evaluated for completeness, correctness, and conformance/compliance of a specific data set against the standard operating procedure (verified), as well as reviewed for specific analytic quality (validated)	Images and videos collected with UAVs, which are verified (content verification) and filtered (quality enhancement)
Analyzed Collected Data	Validated data are then analyze, through statistical operations, based on a specific target or application scenario	Patterns of smoke or fire found in the video collected from UAV
Generated Data	The data needed to validate the results presented in scientific publications (pseudo-code, libraries, workflow, naming schemes, . . .)	Naming scheme associated to the analyzed data (i.e: geolocalization, fire dimension, . . .)

The followings sections describe some sample datasets that we are planning to collect and generate in UMOBILE. The provided datasets are, at this early stage of the project, possible examples which are probably subject to change with the evolution of the project.

For each of the dataset that we are going to share in the project lifetime, policies for access and sharing as well as policies for re-use and distribution, will be defined and applied. A generic guideline is provided in sections "Data sharing" and "Archiving and preservation".

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

4 Dataset 1: Message delay

4.1 Data set reference and name

UMOBILE.MES_DELAY

4.2 Data set description

Message delay is a Key Performance Indicator in computer networks. Data produced by simulation tools and/or by real life trials will be used as a means to quantify the performance advantages the UMOBILE architecture offers compared to current practices. Message delay is measured in seconds and it may range from milliseconds to minutes or even hours in scenarios involving disruptive communication environments. Scientific publications related to the UMOBILE project may include Message delay data.

4.3 Standards and metadata

Metadata will include the simulation tool used to create the message delay data and the configuration parameters.

5 Dataset 2: AAA logs

5.1 Data set reference and name

UMOBILE.AAA_LOGS

5.2 Data set description

AAA logs are written by the AAA server in order to record all the events that happen during while the server is running. They contain information about the authentication requests and they are very useful in order to detect problems in the testing phase or even to extract information about users behavior.

These logs contain private information about the users that must be handled with care. Even if the information has been collected in a testing phase, user rights have to be respected. Therefore, and because of the open nature of the data managed in this project, the information in the logs must be anonymized before releasing it.

5.3 Standards and metadata

There are no standards for these logs. A possible solution are RADIUS servers as AAA servers. In this case, the logs would include the attributes defined by RADIUS.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

6 Dataset 3: Social Network Reports

6.1 Data set reference and name

UMOBILE.SOCIAL_REPORTS

6.2 Data set description

These reports contain personal information about the users' and information about their behavior. This information can be used for statistical purposes and this is especially valuable in some use cases of UMOBILE.

These reports contain private information about the users that must be handled with care. Even if the information has been collected in a testing phase, user rights have to be respected. Therefore, and because of the open nature of the data managed by this project, the information in the reports must be anonymized before releasing it.

6.3 Standards and metadata

There are no standards for this type of dataset. The kind of the information provided in these reports depends on the information needed in each situation and the availability of each social network.

7 Dataset 4: Affinity Networking

7.1 Data set reference and name

UMOBILE.AFFINITY_SETS

7.2 Data set description

These traces shall contain contact data related with: visits of devices to UMOBILE hotspots; direct contact between devices (Bluetooth and Wi-Fi). Aspects kept relate with average visit/contact time; social strength computation derived from the association and exchange of data between devices; whether or not the owners of devices were acquainted before, etc. The data shall be provided both in sql format as well as in text. There is NO private data concerning the users kept. The MACs are hashed, and the IPs are hidden. This data is useful to better understand the evolution of affinity networks based on short-range wireless technology, over time and with different time granularity (e.g. days, weeks, months).

7.3 Standards and metadata

The data is expected to be provided in ANSI SQL, XML, or text (ASCII) format. For this data set, data citation and metada practices derived from CRAWDAD shall be considered (`http://www.dlib.org/dlib/january15/henderson/01henderson.html`)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

8 Dataset 5: Social Context

8.1 Data set reference and name

UMOBILE.SOCIAL_CONTEXT_SETS

8.2 Data set description

These traces shall contain contact data related with: UMOBILE users physical activity (walking, running, standing, driving); surrounding environment (noisy, calm, number of talking events); relative distance among UMOBILE devices; social interaction among UMOBILE devices (strength of social ties); Traces can also include information about the overall social context of an UMOBILE user, such as social isolation. The data shall be provided both in sql format as well as in text. There is NO private data concerning the users kept, since the identity of the user is not collected nor stored. This data is useful to better understand the context of each UMOBILE users in different scenarios. For instance, such traces will help to understand how to improve social daily routines (e.g. with the goal of reducing social isolation), and will allows us to consider information about the users' context aiming to improve the efficiency when reacting to emergency situations, or civil protection cases, or even the dissemination of micro-blogs.

8.3 Standards and metadata

This data set may help to better understand what is the semantics and mandatory/optional fields that should be considered in a data dissemination protocol. Related to for instance: draft-irtf-icnrg-ccnxsemantics-00, draft-irtf-icnrg-ccnxmessages-00

9 Data sharing

Open access to research data wil be achieved in UMOBILE through the following steps:

1. Write, and update as needed, the "Data Management Plan" (current document)
2. Select what data we'll need to retain to support validation of the project finding (the datasets described in the above section)
3. Deposit the research data into a **online research data repository**. In deciding where to store project data, the following choice will be performed, in order of priority:
   - An institutional research data repository, if available
   - An external data archive or repository already established in the UMOBILE research domain (to preserve the data according to recognised standards)
   - The European sponsored repository: `http://zenodo.org/`

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

- Other data repositories (searchable here: `http://www.re3data.org`), if the previous ones are ineligible

4. License the data for reuse (Horizon 2020 recommendation is to use CC0 or CC BY)
5. Provide info on tools needed for validation: everything that could help third party in validating the data (workflow, code,. . .)

Independent of the choose, the authors will ensure that the repository:

- Gives the submitted dataset a persistent and unique identifier, to make sure that research outputs in disparate repositories can be linked back to particular researchers and grants
- Provides a landing page for each dataset, with metadata
- Helps to track if the data has been used by providing access and download statistics
- Keeps the data available in the long term, if desired
- Provides guidance on how to cite the data that has been deposited

Even following the previously described steps, each case will be examinated separately in order to decide which online repository to choose.

9.1 Policies for Access and Sharing

As suggested from the Euporean Commission, the partners will deposit **at the same time the research data needed to validate the results presented in the deposited scientific publications**. This timescale applies for data underpinning the publication and results presented: research papers written and published during the funding period will be made available with a subset of the data necessary to verify the research findings. The consortium will then make a newer, complete version of data, available within 6 months of project completion. This embargo period is requested to allow time for additional analysis and further publication of research findings to be performed.

Other data (not underpinning the publication) will be shared during the project life following a granular approach to data sharing, releasing subsets of data at distinct periods, rather than wait until the end of the project, in order to obtain feedback from the user community and refine it as necessary.

An important aspect to take into account, is **who is allowed to access the data**. It could happen that some of the dataset shouldn't be publicly accessible to everyone. In this case, a control mechanisms will be established. These include:

- Authentication systems that limit read access to authorized users only
- Procedures to monitor and evaluate, one to one, access requests: user must complete a request form stating the purpose for which they intend to use the data.
- Adoption of a Data Transfer Agreement that outlines conditions for access and use of the data

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

Each time a new dataset will be deposited, the consortiun will decide on who is allowed to access the data. Generally speaking, anonymised and aggregate data will be made freely available to everyone, whereas sensitive and confidential data will only be accessed by specific authorized users.

9.2 Policies for Re-use, Distribution

A key aspect will be **how users will learn of the existence of data** and the content it contains. People will not be interested in a set of unlabelled files published on a website. To attract interest, partners will describe accurately the content of published dataset and, each time a new dataset will be deposited, the information will be disseminated using the appropriate mean (i.e.: mailing list, press release, facebook, website), based on the type of data and on the interested target audience.

Research data will be made available in a way that can be shared and easily reused by others. That means:

1. sharing data using **open file format** (whenever possible), so that they can be implemented by both proprietary and open source software;
2. using format based on an underlying open standard
3. using format which is interoperable among diverse internal and external platforms and applications
4. using format which does not contain proprietary extensions (whenever possible)

Documenting datasets, data sources, and methodology by which the data were acquired establishes the basis for interpreting andappropriately using data. Each generated or collected and then deposited dataset, will include documentation to help users to re-use it.

As recommended, the license that will be applied to the data is CC0 or CC BY. If some limitations will occur on the generated data, these restrictions will be clearly described and justified.

Potential issues, that could affect how data can be shared and used may include the need to: protect participant confidentiality, comply with informed consent agreement, protect Intellectual Property Rights, submit patent applications, protect commercial confidentiality. Possible measures that may be applied to address these issues include: encryption of data during storage and transfer, anonymisation of personal information, development of Data Transfer Agreements that specify how data may be used by an end user, specification of embargo periods, and development of procedures and systems to limit access to authorized users only (as already explained).

10 Archiving and preservation

Dataset will be maintained for 5 years following project completion.

To ensure high-quality long-term management and maintenance of the dataset, the consortium will implement **procedures to protect information over time**. These procedures will permit a broad range of users to easily obtain, share, and properly interpret both active and archived information, and they will ensure that information are:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124

- kept up-to-date in content and format so they remain easily accessible and usable;
- protected from catastrophic events (e.g., fire and flood), user error, hardware failure, software failure or corruption, security breaches, and vandalism.

Regarding the second aspect, solutions dealing with disaster risk management and recovery, as well as with regular backups of data and off-site storage of backup sets, are alway integrated when using the official data repositories (i.e.: `http://zenodo.org/`); the partners will ensure the adoptions of similar solutions when choosing an institutional research data repository.

Partners are encouraged to claim costs for resources necessary to manage and share data; these will be clearly described and justified. Arrangements for post-project data management and sharing must be made during the life of the project. Costs associated with long-term curation and preservation, such as POSF (Pay Once, Store Forever) storage, will be purchased before the close of the project grant

11 Conclusion

The purpose of the Data Management Plan is to support the data management life cycle for all data that will be collected, processed or generated by the UMOBILE project. The DMP is not a fixed document, but evolves during the lifespan of the project. This document is expected to mature during the project; more developed versions of the plan could be included as additional deliverables at later stages. The DMP will be updated at least by the mid-term and final review to fine-tune it to the data generated and the uses identified by the consortium since not all data or potential uses are clear at this stage of the project.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645124