Management Of Networked IoT Wearables – Very Large Scale Demonstration of Cultural Societal Applications
(Grant Agreement No 732350)

D11.2 Open Data Management Plan

Date: 2017-11-15

Version 1.0

Published by the MONICA Consortium

Dissemination Level: Public

Co-funded by the European Union's Horizon 2020 Framework Programme for Research and Innovation under Grant Agreement No 732350

Document control page

Document file: D11.2 Open Data Management Plan 1.0.docx
Document version: 1.0
Document owner: CNET

Work package: WP11 – Socio-economic & Business Processes
Task: T11.3 – Open Data Management
Deliverable type: [R]

Document status: ☒ Approved by the document owner for internal review
☒ Approved for submission to the EC

Document history:

Version	Author(s)	Date	Summary of changes made
0.1	Peeter Kool (CNET),	2017-05-03	ToC Created
0.3	Peeter Kool (CNET)	2017-06-07	Added content
0.6	Peeter Kool, Peter Rosengren (CNET)	2017-08-12	Added Template
0.8	Peeter Kool (CNET)	2017-09-22	Added initial DMP Components
0.9	Peeter Kool, Peter Rosengren (CNET)	2017-10-23	Draft for internal and external Review
1.0	Peeter Kool (CNET)	2017-11-15	Final version submitted to the European Commission

Internal review history:

Reviewed by	Date	Summary of comments
Robert Dupré KU	2017-10-31	Some minor clarifications needed
Dr. Michael Fischer FHH-SC	2017-11-06	Some minor comments

Index:

1 Executive Summary

The purpose of the current deliverable is to present the 1st Data Management Plan (DMP) of the MONICA project and to provide the guidelines for maintaining the DMP during the project.

The scope of the DMP is to describe the data management life cycle for all data sets to be collected, processed or generated in all Work Packages during the 36 months of MONICA project. FAIR Data Management is highly promoted by the Commission and since MONICA is a data intensive project, relevant attention has been given to this task. However, the DMP is a living document in which information will be made available on a more detailed level through updates and additions as the implementation of MONICA project progresses and MONICA is deployed at pilot sites.

2 Introduction

The Data Management Plan methodology approach that has been used for the compilation of the D11.2 has been based on the updated version of the "Guidelines on FAIR Data Management in Horizon 2020 1 version 3.0 released on 26 July 2016 by the European Commission Directorate – General for Research & Innovation" (EC, 2016).

The MONICA DMP addresses the following issues:

- Data Summary
- FAIR data
    - Making data findable, including provisions for metadata
    - Making data openly accessible
    - Making data interoperable
    - Increase data re-use
- Allocation of resources
- Data security
- Ethical aspects
- Other issues

According to the EU's guidelines regarding the DMP the document may be updated - if appropriate - during the project lifetime (in the form of deliverables). In MONICA where the pilot sites are in different countries and they are also of very different types and the deployment and usage of the deployed MONICA functionalities is not yet defined. Therefore, we will need to update the DMP with the data that is being collected/created at each pilot site according to its usage and whether it can be published as Open Data.

3 Methodology

The Data Management Plan regards all the data sets that will be collected, processed and/or generated within the project. The methodology the consortium follows to create and maintain the project DMP is as follows:

1) Create a data management policy.
    a) Using the elements that the EC guidelines (EC, 2016) proposes to address for each data set.
    b) Adding the strategy that the consortium uses to address each of the elements.
    c) Adding a Self-Audit process for the DMP.
2) Create a DMP template that will be used in the project for each of the collected data sets, see Appendix 1 MONICA Template for DMP.
3) Creating and maintaining DMPs
    a) If a data set is collected, processed and/or generated within a work package, a DMP should be filled in. For instance, training data sets, example collections et c.
    b) For each of the pilots, when it is known which data will be collected, the DMP for that pilot should be filled in.
4) The filled DMPs should be added to this document as updates in section 5.
    a) This document is the living document describing which data is collected within the project as well as how it is managed.
5) Towards the end of the project an assessment will be made of which data is valuable to keep as Open Data after the end of the project.
    a) For the data that is considered to be valuable an assessment of how the data can be maintained and the cost involved will be made. We expect that in the MONICA project the participating cities can accommodate most of this data within their existing Open Data infrastructure.

4 MONICA Open Data Management Policy

The responsible party for creating and maintaining the DMP for a data set is the partner that creates/collects the data. If a data set is collected, processed and/or generated within a work package a DMP should be created.

Before each pilot execution it should be clear which data set is collected/created in the pilot and how the data will be managed, i.e. the DMPs for the pilot data must be ready and accepted. This will be done individually for each of the pilots because of the difference between the pilots being in different countries and of different types of events, i.e. closed, open et c.

4.1 Naming and identification of the Data set

To have a mechanism for identifying the different collected/generated easily we will use a naming scheme. The naming scheme for MONICA data sets will be a simple hierarchical scheme including the country, pilot, creating or collecting partner and a describing data set name. This name should be used as the identification of the data set when it is published as Open Data in different open data portals.

MONICA_{Country or WP}_{Pilot Site or WP}_{Responsible Partner}_{Description}_{Data Set Sub Index}

Figure 1: MONICA Data Set Naming Scheme

The parts are defined as follows:

- **MONICA:** Static for all data sets, identifying the project.
- **Country:** The two letter ISO 3166-1 country code for the pilot were data has been collected or generated.
- **Pilot Site:** The name of the pilot site where the data was collected without spaces, i.e. Tivoli, RheinInFlammen et c.
- **Responsible Partner:** The partner that is responsible for managing the collected data, i.e. creates and maintains the Open Data Management plan for the data set. Using the acronyms from the DoA.
- **Description:** Short name for the data set without spaces, i.e. SoundPressure, PeopleCount et c.
- **Data Set Sub Index:** Optional numerical index starting on 1. The intention is that data sets created/collected at different times can be distinguished and have their individual meta data

MONICA_DK_Tivoli_CNET_AggregatedPeopleCount_1

Figure 2: Example naming of a MONICA data set

In the example above the Data set is created within MONICA in Denmark at Tivoli. CNET is responsible for Open Data Management plan for the data set. The data set contains aggregated people count and it is the first of a series of data sets collected at different times.

There can be situations where the data needs to be anonymised with regards to the location the data has been collected, for instance at some pilots it might not be allowed to publish people count data with the actual event location for security reasons. In these cases, the **Country** and **Pilot Site** will be replaced by string **UNKNOWN** when it is made available as Open Data**.**

For data sets that are not connected to a specific pilot site the **Pilot Site** should be replaced with the prefix WP followed by the Work Package number that creates and maintains the Open Data Management plan for the data set, i.e. **WP6**. The same applies to the **Country** part which also should be replaced with the prefix WP followed by the Work Package number in the cases where the data set is not geographically dependent, such as pure simulations or statistics.

4.2 Data Summary / Data set description

The data collected/created needs to be described including the following information:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected

- Specify if existing data is being re-used (if any)
    - Provide the identification of the re-used data, i.e. MONICA identifier or pointer to external data, if possible.
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

4.3 Fair Data

FAIR data management means in general terms, that research data should be “FAIR”, that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation solution.

4.3.1 Making data findable, including provisions for metadata

This point addresses the following issues:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism.
- Outline the naming conventions used.
- Outline the approach towards search keywords.
- Outline the approach for clear versioning.
- Specify standards for metadata creation (if any).

As far as the metadata are concerned, the way the consortium will capture and store this information should be described. For instance, for data records stored in a database with links to each item metadata can pinpoint their description and location.

There are various disciplinary metadata standards, however the MONICA consortium has identified a number of available best practices and guidelines for working with Open Data, mostly by organisations or institutions that support and promote Open Data initiatives, and will be taken into account. These include:

- Open Data Foundation
- Open Knowledge Foundation
- Open Government Standards

Furthermore, data should be interoperable, adhering for data annotation, data exchange, compliant with available software applications related to smart city, security, and acoustics.

4.3.2 Making data openly accessible

The objectives of this point address the following issues:

- Specify which data will be made openly available and if some data is kept closed explain the reason why.
- Specify how the data will be made available.
    - Will the data will be added to any Open Data registries?
- Specify what methods or software tools are needed to access the data, if a documentation is necessary about the software and if it is possible to include the relevant software (e.g. in open source code).
- Specify where the data and associated metadata, documentation and code are deposited.
    - Will the data be stored in external Open Data portals or will it remain in the MONICA cloud Open Data Portal?
- Specify how access will be provided in case there are any restrictions.

4.3.3 Making data interoperable

This point will describe the assessment of the data interoperability specifying what data and metadata vocabularies, standards or methodologies will be followed in order to facilitate interoperability. Moreover, it will address whether standard vocabulary will be used for all data types present in the data set in order to allow inter-disciplinary interoperability.

Within the MONICA project we will deal with data of many different types and from very different sources but in order to promote interoperability we use of the following guidelines:

- OGC SensorThings API model for time series data (OGC, 2017), such as environmental readings et c.
- If the data is part of a domain with well-known open formats that are in common use, this should be selected.
- If the data does not fall in the previous categories an open and easily machine-readable format should be selected.

4.3.4 Increase Data Re-use

This point addresses the following issues:

- Specify how the data will be licensed to permit the widest reuse possible.
    - Tool to help selecting license: https://www.europeandataportal.eu/en/content/show-license
    - If a restrictive license has been selected, explain the reasons behind it.
- Specify when the data will be made available for re-use.
- Specify if the data produced and/ or used in the project is useable by third parties, especially, after the end of the project.
- Provide a data quality assurance processes description, if any.
- Specify the length of time for which the data will remain re-usable.

4.4 Allocation of Resources

The objectives of this point address the following issues:

- Estimate the costs for making the data FAIR and describe the method of covering these costs.
    - This includes, if applicable the, cost for anonymising data.
- Identify responsibilities for data management in the project.
- Describe costs and potential value of long term preservation.

The MONICA project will host an Open Data portal in its cloud that will be used for storing the Open Data. The Data Protection Manager of MONICA is responsible for accepting the security solution used for the MONICA cloud solution. The actual maintenance of data will be the responsibility of the DMP owners, i.e. the creators of the data.

4.5 Data security

If data is not stored within the MONICA system, describe the mechanism for data security and backup. For sharing of sensitive data within consortium there are guidelines in chapter 4 of D10.5 The MONICA Ethical Guidelines that should be followed. For live pilot data it must be assessed by the Cyber Security risk management process defined in D10.11.

4.6 Ethical aspects

How does the data relate to the guidelines in deliverable D10.5 The MONICA Ethical Guidelines. An important point is: Does the informed consent, if any, cover the intended use of the data including long term preservation?

4.7 Self-Audit Process

Within MONICA project the ethical manager will be in charge of the execution of the defined data management plan and will supervising the compliance with legal and ethical issues in terms of information security, data protection and ethical issues. The existence of an auditing mechanism is deemed necessary in order to avoid the publication of non-validated data.

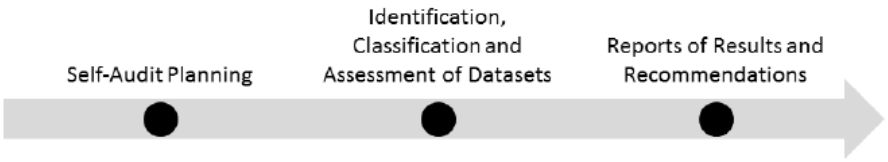

Figure 3: The self-auditing process of MONICA

The steps of the Self-Audit process that will be implemented are summarized below:

- Self-Audit Planning
  - Plan and Set-up Self-Audit
  - Collect Relevant Documents
- Identification, Classification and Assessment of Data sets
  - Analyze Documents
  - Identify Data Sets
  - Classify Data Sets
  - Assess Data Sets
- Report of Results and Recommendations
  - Collate and analyze information from the audit
  - Report on the compliance with the Data Management Plan
  - Identify weaknesses and decide on corrective actions

4.8 Other issues

Other issues will refer to other national/ funder/ sectorial/ departmental procedures for data management that are used.

5 Initial DMP Components in MONICA

During the next period each work package will analyse which DMP components are relevant in their work package. When the pilot definitions are ready with regards to which data is collected and how data is used, DMPs for the pilot needs to be created. This definition will follow the template in Annex 1. Below we present a first set of initial generic DMP components.

5.1 User Scenarios

DMP Element	Issues to be addressed
Identifier	MONICA_WP2_WP2_DEXELS_UserScenarios_1
DMP Responsible Partner	DEXELS
Revision History	**Date** / **Partner** / **Name** / **Description of change** **2017-07-08** / CNet / Peeter Kool / Created initial DMP
Data Summary	Definition of user scenarios for scoping of the initial requirements (Deliverable 2.1 Scenarios and Use Cases for use of IoT Platform in Event Management) is based on interviews, questionnaires and discussions with pilot partners. Collating data from end users is an integral part of the MONICA project – co-production of the final product will help to ensure that a useful product is created. The origin of the data is from pilot use case partners in the MONICA project. Written responses are likely to be fairly small in size (<1Gb over the course of the project).
Making data findable, including provisions for metadata	It will become both discoverable and accessible to the public when the consortium decides to do so. The report contains a table stating all versions of the document, along with who contributed to each version, what the changes where as well as the date the new version was created.
Making data openly accessible	The data are available in D2.1: Scenarios and Use Cases for use of IoT Platform in Event Management. The dissemination level of D2.1 is public. It is to be available through the MONICA wiki for the members of the consortium and when the project decides to publicize deliverables, it will be uploaded along with the other public deliverables to the project website or anywhere else the consortium decides.
Making data interoperable	Raw data cannot be made freely available because it contains sensitive information.
Increase Data Re-use	Engineers who want to build similar systems, could use this as an example.
Allocation of Resources	N/A
Data security	The Scenario and Use case report will be securely saved on the Fraunhofer premises and will be shared with the rest of the partners through the MONICA wiki.
Ethical aspects	N/A
Other Issues	

5.2 User Requirements

DMP Element	Issues to be addressed
Identifier	MONICA_WP2_WP2_DEXELS_UserRequirements_1
DMP Responsible Partner	DEXELS
Revision History	**Date** / **Partner** / **Name** / **Description of change** **2017-06-08** / CNet / Peeter Kool / Created initial DMP
Data Summary	Analysis and definition of User Requirements for scoping of the initial requirements (Deliverable 2.3 Initial Requirements Report which will be followed by a D2.4 Updated Requirements Report) are based on interviews, questionnaires and discussions with pilot partners (see previous DMP). The data is essential for the technical team to develop the MONICA platform; other partner teams throughout the project, as well as the wider research community will benefit when results are published.
Making data findable, including provisions for metadata	It will become both discoverable and accessible to the public when the consortium decides to do so. The report contains a table stating all versions of the document, along with who contributed to each version, a changelog as well as the date the new version was created.
Making data openly accessible	The data are available in D2.3: Initial Requirements Report. The dissemination level of D2.3 is confidential. It is available through the MONICA wiki for the members of the consortium and when the project decides to publicize deliverables, it will be uploaded along with the other public deliverables to the project website or anywhere else the consortium decides.
Making data interoperable	Raw data is recorded and formatted following the Volere template into the JIRA Issue tracker hosted at Fraunhofer premises.
Increase Data Re-use	Engineers who want to build similar systems, could use this as an example.
Allocation of Resources	N/A
Data security	The Initial Requirements Report will be securely saved on the Fraunhofer premises and will be shared with the rest of the partners through the MONICA BSCW document sharing system hosted by FIT.
Ethical aspects	N/A
Other Issues	

5.3 System Architecture

DMP Element	Issues to be addressed
Identifier	MONICA_WP2_WP2_ISMB_SystemArchitecture_1
DMP Responsible Partner	ISMB
Revision History	**Date** / **Partner** / **Name** / **Description of change** **2017-08-10** / CNET / Peeter Kool / Created initial DMP

Data Summary	A report describing the MONICA platform in detail containing information like component descriptions and dependencies, API descriptions, information flow diagram, internal and external interfaces, hardware requirements and testing procedures. This will be the basis upon which the system will be built.
Making data findable, including provisions for metadata	It will become both discoverable and accessible to the public when the consortium decides to do so. The report contains a table stating all versions of the document, along with who contributed to each version, a changelog where as well as the date the new version was created.
Making data openly accessible	The data are available in D2.2: Monica IoT architecture. The dissemination level of D2.2 is confidential. It is to be available through the MONICA wiki for the members of the consortium and when the project decides to publicize deliverables.
Making data interoperable	N/A
Increase Data Re-use	Engineers who want to build similar systems, could use this as an example.
Allocation of Resources	N/A
Data security	The Architecture report will be securely saved on the Fraunhofer premises and will be shared with the rest of the partners through the MONICA BSCW document sharing system hosted by FIT.
Ethical aspects	N/A
Other Issues	

5.4 Pilot Generated Data Sources

The main foreseen data sources from MONICA will come from the Pilot Use Cases and will be made available as open data as far as possible. However, there are also many data sets that cannot be made available due to its sensitiveness, for instance live video recordings. Here are an initial set of generic DMP Components that are not linked to individual pilots. But we foresee that for each component published as Open Data from a pilot will need its own individual DMP.

5.4.1 Sound Level Time Series

DMP Element	Issues to be addressed
Identifier	MONICA_WP4_WP4_B&K_SoundLevel_1
DMP Responsible Partner	B&K
Revision History	**Date** / **Partner** / **Name** / **Description of change** **2017-08-22** / CNet / Peeter Kool / Created initial DMP
Data Summary	The data collected from deployed acoustics sensors with sound level measurements. The data will also contain the position of the sensors with geocoordinates.
Making data findable, including provisions for metadata	The data will be made available in OGC SensorThings API format
Making data openly accessible	The data will be published on the MONICA Open Data portal.

Making data interoperable	The data will be made available in OGC SensorThings API format
Increase Data Re-use	Data scientists and sound engineers will benefit from being able to analyses the real time data streams as well as historic records..
Allocation of Resources	N/A
Data security	
Ethical aspects	N/A
Other Issues	

5.4.2 Surveillance Video Streams

DMP Element	Issues to be addressed
Identifier	MONICA_WP5_WP5_KU_VideoStream_1
DMP Responsible Partner	KU
Revision History	**Date** / **Partner** / **Name** / **Description of change** **2017-08-21** / CNet / Peeter Kool / Created initial DMP
Data Summary	Surveillance cameras installed at the different events will generate large amounts of data. This data will not be publicly available.
Making data findable, including provisions for metadata	The data will generally not be stored in the MONICA cloud since the amount of data is too large. Instead it will be stored locally if at all.
Making data openly accessible	In general, the data will not be publicly available. After careful analysis some video streams might be made available for research purposes if it is possible from a security and privacy perspective.
Making data interoperable	The data will be in a standard video format.
Increase Data Re-use	Data scientists and sound engineers will benefit from being able to analyses the data streams.
Allocation of Resources	N/A
Data security	Should follow the guidelines from D10.11
Ethical aspects	Local storage of video will follow the ethical guidelines and the local regulations and laws for the site.
Other Issues	

5.4.3 Wearables Positioning Streams

DMP Element	Issues to be addressed
Identifier	MONICA_WP3_WP3_Dexels_PositionStream_1
DMP Responsible Partner	DEXELS
Revision History	**Date** / **Partner** / **Name** / **Description of change** **2017-08-22** / CNet / Peeter Kool / Created initial DMP
Data Summary	UWB wristbands provide a stream of position information that is used for locating people for different purposes.

Making data findable, including provisions for metadata	The data will be made available in OGC SensorThings API format
Making data openly accessible	In general, the data will not be publicly available. After careful analysis some position streams might be made available for research purposes if it is possible from a security and privacy perspective.
Making data interoperable	The data will be made available in OGC SensorThings API format
Increase Data Re-use	Data scientists and sound engineers will benefit from being able to analyses the real time data streams as well as historic records.
Allocation of Resources	N/A
Data security	The data will be securely stored in the ATOS deployed MONICA cloud and using the guide lines from D10.11
Ethical aspects	Contains possible sensitive information about individual people movement.
Other Issues	

5.4.4 Common Operational Picture

DMP Element	Issues to be addressed
Identifier	MONICA_WP6_WP6_CNET_CommonOperationalPicture_1
DMP Responsible Partner	CNET
Revision History	**Date** / **Partner** / **Name** / **Description of change** **2017-08-23** / CNet / Peeter Kool / Created initial DMP
Data Summary	The Common Operational Picture is the most central data element in MONICA. It represents the current status of all relevant operations and process parameters at the event site such as number of visitors, current reported incidents, threat levels, sound levels, et c. The data will be accessed and used by the event operators and security personnel.
Making data findable, including provisions for metadata	The data will be made available in SQL and No-SQL formats. It will be searchable using ODATA 2.0.
Making data openly accessible	In general, the data will not be publicly available. Parts of can be made available if possible considering ethical and security concerns.
Making data interoperable	The data will be made available in standard formats such as OGC SensorThingsAPI depending on the type of information.
Increase Data Re-use	The information could be very useful for research in all the areas related to public outdoor events such as, people movement, incident detection et c.
Allocation of Resources	N/A
Data security	The data will be securely stored in the ATOS deployed MONICA cloud and using the guide lines from D10.11
Ethical aspects	Can contain sensitive data linking individuals to location and actions.
Other Issues	

6 Conclusion

The purpose of this document is to provide the plan for managing the data generated and collected during the project; The Data Management Plan. Specifically, the DMP described the data management life cycle for all data sets to be collected, processed and/or generated by a research project.

This document will be continuously updated with DMPs for each of the data sets during the project lifetime.

7 References

(EC, 2016) European Commission (2016). H2020 Programme, Guidelines on FAIR Data Management in Horizon 2020. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (Accessed 21 june 2017)

(OGC, 2017) Open Geospatial Consortium (OGC) SensorThings API. https://github.com/opengeospatial/sensorthings (Accessed 5 august 2017)

Appendix 1 MONICA Template for DMP

DMP Element	Issues to be addressed
Identifier	The identifier of the data set following the MONICA naming principles, see 4.1
DMP Responsible Partner	The Partner that is responsible for creating and maintaining the DMP
Revision History	**Date** / **Partner** / **Name** / **Description of change** **2017-06-08** / Xxx / N N / Created DMP
Data Summary	Guidelines in 4.2
Making data findable, including provisions for metadata	Guidelines in 4.3.1
Making data openly accessible	Guidelines in 4.3.2
Making data interoperable	Guidelines in 4.3.3
Increase Data Re-use	Guidelines in 4.3.4
Allocation of Resources	Guidelines in 4.4
Data security	Guidelines in 4.5
Ethical aspects	Guidelines in 4.6
Other Issues	Guidelines in 4.8