



Project title: Multi-Owner data Sharing for Analytics and Integration respecting Confidentiality and OWNeR control
Project acronym: MOSAICrOWN
Funding scheme: H2020-ICT-2018-2
Topic: ICT-13-2018-2019
Project duration: January 2019 – December 2021

D7.1

Data Management Plan

Editors: Pierangela Samarati (UNIMI)
Reviewers: Andrew Byrne (EISI)
 Tomasz Pawlowicz (MC)
 Daniel Bernau (SAP SE)
 Stefano Paraboschi (UNIBG)
 Rigo Wenning (W3C)

Abstract

This deliverable describes the data management plan of the MOSAICrOWN project. It identifies the different kinds of data managed and produced by the project, describing their representation, management, sharing, archiving and preservation.

Type	Identifier	Dissemination	Date
Deliverable	D7.1	Public	2019.06.30



This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 825333. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the European Commission.

MOSAICrOWN Consortium

1. Università degli Studi di Milano	UNIMI	Italy
2. EMC Information Systems International	EISI	Ireland
3. Mastercard Europe	MC	Belgium
4. SAP SE	SAP SE	Germany
5. Università degli Studi di Bergamo	UNIBG	Italy
6. European Research Consortium for Informatics and Mathematics	W3C	France

Disclaimer: The information in this document is provided "as is", and no guarantee or warranty is given that the information is fit for any particular purpose. The below referenced consortium members shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law. Copyright 2019 by Università degli Studi di Milano.

Versions

Version	Date	Description
0.1	2019.06.04	Initial Release
0.2	2019.06.25	Second Release
1.0	2019.06.30	Final Release

List of Contributors

This document contains contributions from different MOSAICrOWN partners. Contributors for the chapters of this deliverable are presented in the following table.

Chapter	Author(s)
Executive Summary	UNIMI
Chapter 1: Data Management Plan	UNIMI

Contents

Executive Summary	7
1 Data Management Plan	9
1.1 Data set: Data about the project	9
1.2 Data set: Data for the working of the project	10
1.3 Data set: Project results - Documents	10
1.4 Data set: Project results - Software tools	12

Executive Summary

This deliverable describes the data management plan establishing the policy regulating collection, management, sharing, archiving, and preservation of data in the MOSAICrOWN project. Data regulated by the plan are all data that are either managed or are produced by MOSAICrOWN (i.e., documents and software tools).

Introducing the MOSAICrOWN data management plan, we want to explicitly note that the project does not use or manage any real-life data. As a matter of fact, MOSAICrOWN strictly builds the technical framework and solutions for the digital data market, but neither collects nor uses any personal data either directly or indirectly (for more information see Deliverable D1.1 “POPD – Requirement No. 2”).

Therefore, data produced by MOSAICrOWN are technical solutions (in the form of documents or software tools).

The MOSAICrOWN Consortium is committed to timely and rapid distribution of the project’s results, making them widely available and openly accessible. MOSAICrOWN pursues an open-access policy, making results and publications publicly available. Data needed for the working of the project (such as work communications and progress reports on the technical work) are restricted to the project participants.

In this deliverable, we describe the data management plan with reference to the different kinds of data handled or produced by MOSAICrOWN.

1. Data Management Plan

Data produced by the project are technical solutions (in the form of documents or software tools). The project does not acquire or produce any specific data sets. Therefore, in the following we refer the data management plan to the different kinds of data managed or produced by the project distinguishing the following kinds of data: *i*) data/information about the project, *ii*) data for the working of the project itself, *iii*) documents (scientific papers, and deliverables/work documents) produced by the project, and *iv*) software tools produced by the project.

There are three main servers where data on the project are hosted. These are dedicated to manage the following:

- Web site of the project (<https://mosaicrown.eu>) (with a public and a restricted area),
- Project document repository SVN (<https://mosaicrown.eu/svn/mosaicrown/>) for the project's internal working and communication,
- Mailing lists (mosaicrown-...@di.unimi.it) for project working and communication, where ... instantiates to different lists and contexts (e.g., “mb” for the Management Board, “tb” for the Technical Board).

All servers reside within the premises of the project coordinator (UNIMI) and are managed by administrative personnel of UNIMI. They are backed up weekly.

1.1 Data set: Data about the project

Data set description

This kind of data refers to all the data describing the project that are to be made publicly available (for dissemination purposes) such as: fact sheet, consortium, objectives, vision, planned work and its organization in work packages. In addition to these static data about the project, this kind of data includes information that is continuously updated as the project progresses, such as news reporting related events, dissemination and exploitation relevant information.

Standards and metadata

Data are organized with a hierarchical structure allowing easy retrieval and navigation. Content management is handled with the Wordpress content management system.

Data sharing

Data in this data set are publicly accessible via the project web site <https://mosaicrown.eu>.

Archiving and preservation

Data are stored on a server residing within the premises of the project coordinator (UNIMI) and are managed by administrative personnel of UNIMI. The server is backed up weekly. Data will be maintained at least five years after the life of the project.

1.2 Data set: Data for the working of the project

Data set description

This kind of data refers to the information needed for the working of the project itself, such as name and contact information of people working in the project, meetings, communication among them, and intermediary progress reports. As established by the Consortium Agreement (art 10.9), for people working on the project, only contact information is acquired, and MOSAICrOWN will not make available any personal data to other parties or process any personal data on behalf of other parties.

Standards and metadata

Contact information is organized in different mailing lists to allow easy retrieval and references. Work in progress is organized in a hierarchical structure with nested directories and files with self-explanatory names. Mailing lists and SVN information is also accessible via the restricted area of the project web site using project login credentials.

Data sharing

Data sharing is restricted to project participants. Access to the SVN service is available to all project participants and is regulated with control of login and (randomly generated) password distributed to each individual participant by the administrator of the server. Access to the different mailing lists of the project is regulated with control of login and (randomly generated) password distributed to each individual participant by the administrator of the server.

Archiving and preservation

Data are stored on a server sitting within the premises of the project coordinator (UNIMI) and are managed by administrative personnel of UNIMI. The server is backed up weekly. Data will be maintained at least five years after the life of the project.

1.3 Data set: Project results - Documents

Data set description

This kind of data refers to all documents produced by the project. Among them we distinguish: deliverables/work documents and scientific papers.

Standards and metadata

All documents (in .tex or .doc format when in production) are made available in PDF format once completed.

Deliverables and work documents are identified following an easy-to-use notation with three parts: a letter that determines the kind of document (D for Deliverable, W for Work document); the number of the work package that coordinates its production; and a serial number that uniquely identifies the document within that work package.

All documents have .bibtex metadata, easing the search and can be exported for easy reference. These metadata include the ones prescribed in the Grant Agreement (art 29.2).

Data sharing

Intellectual Property Rights (IPRs) and dissemination of project results are regulated by the Consortium Agreement (art 8). IPR of results remain with the project's party that generated them. Sharing and dissemination follow an open access policy. For sharing and dissemination, we distinguish deliverables and scientific papers.

All deliverables of the project that are classified as public (PU) are made accessible to the general public following their final acceptance by the EC. This dissemination and sharing is made via the web site of the project (<https://mosaicrown.eu>, link "Results/Deliverables").

MOSAICrOWN embraces an open access policy, and also all scientific papers are made publicly available via the project website. Before undergoing public release, paper publication undergoes an internal process within the MOSAICrOWN Consortium, regulated by the Grant Agreement (art. 29) and Consortium Agreement (art. 8.3) of the project. In particular, "prior notice of any planned publication shall be given to concerned Parties at least 21 calendar days before the publication. Any objection to the planned publication shall be made in accordance with the Grant Agreement in writing to the Coordinator and to the Party or Parties proposing the dissemination within 15 calendar days after receipt of the notice. If no objection is made within the time limit stated above, the publication is permitted." At this point, the paper is made publicly available via the project public web site (<https://mosaicrown.eu>, link "Results/Publications").

Before papers are accepted as publication at scientific venues, sharing and dissemination are made possible via the partners' institutional archives and other open archives (e.g., arXiv). For papers accepted for formal publication by publishers, this public sharing and dissemination access is made possible via "green" open-access publication (i.e., "self-archiving"), which is in line with the copyright policy of major institutions and associations of the most selective and recognized conferences and journals. Within this policy, the publishers allow authors to post the final versions of accepted papers in their personal web site, the web site of their employers, or selected pre-authorized institutional web sites. Project publications are then hosted (and publicly accessible) from the partners' website or open archives and linked from the web site of the project, thus ensuring broad visibility and easy access.

Archiving and preservation

Deliverables are stored on a server sitting within the premises of the project coordinator (UNIMI) and are managed by administrative personnel of UNIMI. The server is backed up weekly. Deliverables will be maintained at least five years after the life of the project.

Scientific papers are linked from the project web site and are stored on partners institutional archives (also periodically backed up) or other open archives (like arXiv). This provides guarantees of continuity of service preservation of access.

1.4 Data set: Project results - Software tools

Data set description

This kind of data refers to all software tools produced by the project. Software tools to be developed within the project have all corresponding deliverables or work documents.

Standards and metadata

Code of software tools produced by the project are written using commonly-used programming languages (e.g., Java, C++, Python) or shell scripts. Every tool will have structured metadata associated, specifying information such as: unique identifier, creator/s, versions; allowing for easy reference and retrieval.

Data sharing

Intellectual Property Rights and dissemination of project results and dissemination of project results are regulated by the Consortium Agreement (art 8). IPR of results remain with the project's party that generated them.

As for sharing, software tools developed in the project, corresponding to work documents and deliverables, subject to the applicable IPR clauses, are classified as public (PU). Object code of tools developed by academic partners, subject to the applicable IPR clauses, will be made accessible to the general public following formal acceptance by the EC. Public dissemination and sharing is made via the web site of the project (<https://mosaicrown.eu>).

Source code of software tools developed by academic partners will be provided as building blocks under an open source license and as applicable under the relevant IPR agreement.

Archiving and preservation

Object code of tools produced within MOSAICrOWN will be stored on a server sitting within the premises of the partners that developed them. Object code of tools will be maintained at least five years after the life of the project. Object code of tools developed by academic partners may also be stored for dissemination in public repository (e.g., GitHub).