



"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813884".



Project Number: 813884

Project Acronym: Lowcomote

Project title: Training the Next Generation of Experts in Scalable Low-Code Engineering Platforms

DATA MANAGEMENT PLAN

Project GA: 813884

Project Acronym: Lowcomote

Project website: <https://www.lowcomote.eu/>

Project officer: Mercedes Leguey-Jimenez

Work Package: WP1

Deliverable number: D.1.3

Production date: 28/06/2019

Contractual date of delivery: 30/06/2019

Actual date of delivery: 30/06/2019

Dissemination level: Confidential, only for members of the consortium (including the Commission Services)

Lead beneficiary: IMT Atlantique

Authors: Massimo Tisi & Marie Chastanet

Contributors: all consortium members



"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813884".



HISTORY OF CHANGES		
Version	Publication date	Change
0.1	14.03.2019	▪ Initial version drafted
0.2	05.05.2019	▪ First review
0.3	28.06.2019	▪ Initial version completed



"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813884".



Project Abstract

Low-code development platforms (LCPD) are software development platforms on the Cloud, provided through a Platform-as a-Service model, which allow users to build completely operational applications by interacting through dynamic graphical user interfaces, visual diagrams and declarative languages. They address the need of non-programmers to develop personalized software, and focus on their domain expertise instead of implementation requirements.

Lowcomote will train a generation of experts that will upgrade the current trend of LCPDs to a new paradigm, Low Code Engineering Platforms (LCEPs). LCEPs will be open, allowing to integrate heterogeneous engineering tools, interoperable, allowing for cross-platform engineering, scalable, supporting very large engineering models and social networks of developers, smart, simplifying the development for citizen developers by machine learning and recommendation techniques. This will be achieved by injecting in LCDPs the theoretical and technical framework defined by recent research in Model Driven Engineering (MDE), augmented with Cloud Computing and Machine Learning techniques. This is possible today thanks to recent breakthroughs in scalability of MDE performed in the EC FP7 research project MONDO, led by Lowcomote partners.

The 48-month Lowcomote project will train the first European generation of skilled professionals in LCEPs. The 15 future scientists will benefit from an original training and research programme merging competencies and knowledge from 5 highly recognized academic institutions and 9 large and small industries of several domains. Co-supervision from both sectors is a promising process to facilitate agility of our future professionals between the academic and industrial world.



“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813884”.



Table of contents

Project Abstract	3
Table of contents	4
Introduction	5
1. Data Summary	6
2. FAIR data	7
2.1. Making data findable, including provisions for metadata	7
2.2. Making data openly accessible	8
2.3. Making data interoperable	8
2.4. Increase data re-use (through clarifying licences)	9
3. Allocation of resources	9
4. Data security	9
5. Ethical aspects	9
6. Other issues	10



“This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813884”.



Introduction

The present Data Management Plan (DMP) is a deliverable of the Lowcomote project (Grant Agreement n°813884), funded by the European Commission Research Executive Agency (REA), under the Innovative Training Networks Programme of the Marie Skłodowska Curie Actions (H2020-MSCA-ITN-2018).

Carefully managing research data is an essential part of good research practice and starts with adequate planning. The DMP is a document specifying how data will be handled both during and after the project. The DMP details data collection, data storage, data security and data retrieval. For the dataset descriptions we used the official guidelines on Data Management Plan (DMP) from the Horizon 2020 portal.

The purpose of this document is to provide the general approach to the project data management and it will be updated regularly all along the project so as to address any data-related issues. The current deliverable cannot be considered as a fixed document; it will evolve and gain more precision and substance during the lifespan of the Lowcomote project. The document will be updated before the mid-term check review meeting (M14 - February 2019), at the end of the first reporting period (M26 - February 2021), and whenever needed.



1. Data Summary

Purpose of the data collection/generation. Lowcomote aims to train a generation of professionals in the design, development and operation of new Low-code Development Platforms (LCDP), by being scalable (i.e., supporting the development of large-scale applications, and using artefacts coming from a large number of users), open (i.e., based on interoperable and exchangeable programming models and standards), and heterogeneous (i.e., able to integrate with models coming from different engineering disciplines).

To reach its main scientific goal, Lowcomote integrates an interdisciplinary and intersectoral research program around three specific research objectives (RO). Lowcomote will train ESRs capable of:

- RO1: Enabling Low-code Engineering of Large-Scale Heterogeneous Systems, by smart development environments on the Cloud and precise integration of low-code languages with new domains.
- RO2: Developing a Large-scale Repository and Services for Low-Code Engineering, as a Cloud-based service able to handle a very large number of low-code artefacts, and automatically learn from them.
- RO3: Producing advancements in Scalable Low-Code Artefact Management, as new algorithms and reusable components.

Types and formats of data generated/collected. Within this framework, the data produced will be:

- Models, metamodels, model queries and transformations
- Experimental data (databases, xml, spreadsheets)
- Publications and reports (textual, document files).

The formats of the data depend on the type of the data collected/generated:

- For models and metamodels, Lowcomote will use standard open formats in the Eclipse Modeling Project. Metamodels will be produced in Ecore and models in XMI. Model queries and transformations will be produced using the open-source languages in the consortium (ATL, Epsilon, IncQuery). When modelling artefacts for specific ESR subjects are produced using proprietary formats, corresponding beneficiaries will be advised to provide also an equivalent version in the above-mentioned open formats.
- Experimental data (databases, xml, spreadsheets) will be the comprehensive result data sets of characterized samples that are used to create the figures and plots in scientific publications, such that other researchers can compare their results easier and such that further results including historic data can be produced quicker.

The data are published as value tables in Open Document Spreadsheet format (.ods) for limited amounts of data with typed columns or, UTF-8 encoded, comma separated value in textual format files (.csv) with column value and data format description. Proprietary raw data will not be published.

- Publications and reports are shared in PDF format. For technical reports, a document record and change track will be included (author contact information, status, version, change reason



"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813884".



and date, description of contents, title, origin of the data including a brief description of the measurement and/or experiment setup).

Expected size of data. Lowcomote will produce software handling low-code artefacts with millions of model elements (several GBs each). It is expected that the project will generate artefacts of similar size for experimentation.

The size of other experimental data is today not known. Initial experience with storing results from different kind of measurements will permit revising this initial data management plan.

Data utility. As far as the data utility is concerned, the Lowcomote data will be useful first of all to the consortium and to the project partners. The data sets will be shared within the consortium as the working baseline to produce the scientific publications, to verify and validate the results through repeated experiments at different locations and as a baseline for a comprehensive documentation.

The data can be used by independent researchers to understand better the contents and conclusions of the scientific publications, which base their findings on the data. Furthermore, independent researchers can use the files to produce figures and publications, showing comparisons of their own results and the Lowcomote results. Scientists can also use the data files to repeat the experiments and measurements to verify and validate the Lowcomote research. Finally, the data sets may also be used by scientific writers and the press to produce high-quality infographics, demonstrating the impact potentials of the technology.

2. FAIR data

Lowcomote project's consortium members fully commit to respect EU recommendations about research data management. To this end, measures will be taken by Lowcomote members to make research data findable, openly accessible, inter-operable and re-usable.

2. 1. Making data findable, including provisions for metadata

Discoverability of data. The Lowcomote consortium fully believes that, when possible, scientific knowledge should be openly shared. During the project implementation, data collected and generated by the project will be identified and documented according to the Data Management Plan.

The Zenodo platform¹ will be used to make the data openly accessible and discoverable. In any case, the data will be made available on the Lowcomote web site² together with metadata describing the data sets once they are released on Zenodo. Since the open data support the quality and credibility of the open publications, all data are discoverable through the scientific publications.

Metadata will be associated to all data files archived during the Lowcomote project. Because of the diversity of data types, it will be difficult to generate a universal metadata format. However, we will strive to define a common metadata standard, particularly for low-code artefacts.

Identifiability of data and standard identification mechanism. Each data set will carry a DOI as unique and persistent identifier. Data sets will be referenced in scientific publications and if the open data platform permits, scientific papers based on the data will be linked on the open data platform. The DOI is reserved when a Zenodo entry is created before any data are uploaded to the platform. At this point, the data set is not published and its visibility is classified as "Closed Access".

Approach for clear versioning. For archived objects that may exist in different versions (typically, software and models), all versions will be archived, including the metadata of the newer versions

¹ <https://zenodo.org/>

² <https://www.lowcomote.eu/>



indicating the location of the previous one and containing a summary of the changes. Versions will be numbered according to standard conventions.

2.2. Making data openly accessible

Data openly available. The publication of data is under the precondition that no legal, ethical or security constraints are identified and the Intellectual Propriety Rights (IPR) of all involved parties are considered appropriately. For each data that will be generated, the following points will be duly considered:

- Until such time as the DMP has been generated, all data generated by individual partners will belong to the individual partners;
- The M14 version of the DMP will delineate how data may be shared amongst partners, and who will maintain ownership and control of the data generated;
- A selection of data will be chosen for dissemination and exploitation, i.e. for publication and for intellectual property protection. Remaining data will be archived. A procedure for ensuring non-disclosure of sensitive data prior to patent protection will be agreed by the EB.

Whether the data will be made fully open immediately will be decided case-by-case. In any case, all data will be made fully accessible after publication of the corresponding article. In addition, the data will be openly shared among partners of the consortium, except in some specific cases. The DOI corresponding to the data will be provided in publications. For data made open prior to publications, the corresponding links will be found on the partners' institutional websites.

Methods or software tools needed to access the data. Whenever possible, the data will be converted into formats that can be opened using public domain software. If not, the specific tools required will be indicated in the metadata file (itself a PDF or markdown file, thus openly accessible). When a specific software will be created by consortium members to access/visualize the data, this software will be available as any other data file.

Placement of data, associated metadata, documentation and code. Lowcomote will use industry-standard collaborative content management services such as GitLab³. All open-source software will be published in the Lowcomote GitLab group⁴. Git repositories will be managed by each beneficiary and used to save, version, and share the data inside and outside the consortium. Each repository will provide links to the experimental datasets used in the validation of that particular research artefact.

Access in case of restrictions. For restricted datasets, access will be managed by the PI responsible for the data generation. In general, efforts will be made to curate the data file in order to provide the relevant source data for other researchers who wish to reproduce the analysis without compromising private information.

2.3. Making data interoperable

Interoperability of data (the use of data and metadata vocabularies, standards or methodologies in order to facilitate interoperability). Because of the diverse types of data and of tools, it will not be possible (nor necessarily desirable) to attain a full interoperability within the project. However, as specified above, efforts will be made, notably regarding low-code artefacts, to convert all modeling artefacts into a common format (readable with open source software) and have a common metadata format.

³ <https://about.gitlab.com/>

⁴ <https://gitlab.com/lowcomote>



2.4. Increase data re-use (through clarifying licenses)

Licensing of data in order to permit the widest reuse possible. In general, published data will be licensed according to a Creative Commons Attribution License⁵, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed. The open-source software produced during the project will follow the Eclipse Public License (EPL) when possible.

Availability of data for re-use. See the section "Availability of Data" above. Data will be fully available for reuse after publication of the corresponding article. Decisions on making the data immediately available will be made by PIs on a case-by-case basis.

Data quality assurance processes. Data quality will be the responsibility of each partner. However, a collegial evaluation of the archived data will be made during consortium meetings to ensure a shared minimal quality standard.

The length of time for which the data will remain re-usable. We will not impose limits on the length of reuse of the data.

3. Allocation of resources

Cost estimation. The estimated costs for making the Lowcomote data FAIR are mainly those of the time (estimated to be about 1 week/year for each partner) that the researchers will dedicate to the activities connected with this issue. These costs are thus covered by the project funds.

Regarding the costs of preservation of the digital data, Lowcomote will use a free source-code repository, GitLab, and a free data repository, Zenodo.

Responsibilities for data management. Each ESR and his/her main supervisor will be responsible for the management of the data that he/she will collect/generate. The Project Manager will make sure that the common agreed principles of management of the project data are respected. He will be contact person in case any of the project partners has a request concerning the management of the data.

4. Data security

The data generated all along the Lowcomote project will be first stored on the computers associated with the equipment where they were generated. Personnel working on the project will be permitted to have copies of the data on personal computers supplied by the partners.

Lowcomote will make sure that the all the data collected/generated is safely stored for long term preservation. As stated above, the general principle will be a central archival on the Zenodo platform for the whole consortium plus a storage of each partner's data on their institutional servers. This issue will be discussed later in the project and the proper measurements and actions will be taken toward this objective.

When Lowcomote partners will collect sensitive data, ensuring their secure storage and restricting access will be their responsibility.

5. Ethical aspects

All participants in this project will conform to current legislation and regulations in the countries where the research will be carried out. In relation to this project those most relevant are:

- The Charter of Fundamental Rights of the EU

⁵ <http://creativecommons.org/licenses/by/3.0>



"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813884".



- Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data
- The General Data Protection Regulation (GDPR) (EU) 2016/679.

For the first year of the project, the Executive Board will be accompanied by an Ethics Committee, composed by the Data Protection Officers (DPO) of (IMT Atlantique & Johannes Kepler University of Linz. All the physical meetings of the Supervisory Board in the first year will include one member of the Ethics Committee (on a rotating basis).

When one of the partners will collect personal data, consent from participants will be obtained by the partner. However, the data that will be shared with other members of the consortium or with the research community will include only curated data to remove any possibility to retrieve private information.

A specific WP in the project (WP7 "Ethics Requirements") will produce three deliverables at the end of the first year (M12):

- D7.1 "H - Requirement No. 1" will contain the procedures and criteria that will be used to identify/recruit research participants, the informed consent procedures that will be implemented for the participation, templates of the informed consent/assent forms and in-formation sheets.
- D7.2 "POPD - Requirement No. 2" will contain a description of the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants, detailed information on the informed consent procedures in regard to data processing must, relevant authorisations in case of further processing of previously collected personal data.
- D7.3 "GEN - Requirement No. 3" will contain a report concerning GDPR based requirements of the project and compliance of this project with the GDPR.

In the context of ESR7 "Mining Interaction Processes in Low-Code Engineering Platforms", hosted by Johannes Kepler University of Linz (JKUL), user studies are planned to verify the efficiency of the developed interaction mining framework. These users will be selected among the people directly involved in Lowcomote as well as students from the different academic partners. All users will receive an informed consent form explicitly stating the purpose of the study and the data which will be recorded. The data will be anonymized before usage. Furthermore, the participants will have the right to withdraw at any point in time and request the deletion of any kind of stored data about their participation. We will fully ensure compliance with GDPR, by applying the guidelines concerning data protection and privacy of JKUL. The whole process will be monitored by the DPO of JKUL appointed to the project.

6. Other issues

A few partner institutions require that articles be deposited into their own open archives (e.g. HAL). This will be one way to ensure fully open access in the cases where publications do not appear in Open Access journals.

In general, Open Access journals will be preferred for publications; however, this will have to be balanced with the need for high impact.

At the date of submitting this first version of Data Management Plan, no other specific procedures for data management have been identified by Lowcomote consortium's members. Should they arise during project implementation, they will be addressed in future versions of the Data Management Plan.



In particular, we provide a template for describing particular data management requirements for newly planned datasets that beneficiaries will be responsible of filling up. The following version of the DMP will consider these specific requirements.

Table 1. Template for data-management requirements for specific datasets

Project objectives	ESR number: title Project objectives: Submission date: dd/mm/yyyy
Data set reference and name	Identifier for the data set to be produced
Data set description	Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.
Data formats	Description of the format used for storing each digital data type and explanation on why certain (proprietary) formats could be selected over and above other open formats and why this would not be possible otherwise.
Risks and potential difficulties during data collection and processing	Description of factors posing a threat for the quality of the data while collecting and processing the data. If possible, description of how these risks will be dealt with and potential difficulties.
Standards and metadata	Reference to existing suitable standards for metadata creation. If there are no standards in this discipline, describe what type of metadata will be created and how.
Data sharing	Description of whether and how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups.
Archiving and preservation (including storage and backup)	Description of the data backup procedures that will be adopted to ensure the data and metadata are robustly stored during the lifetime of the project. Description of measures taken to ensure the security of data (especially for sensitive data). Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.
Ethical issues	Description of potential ethical issues of collecting, storing, processing and archiving data. Description of procedures of ethical approval related to the project.



"This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813884".



Responsibilities	Indication of who, within each team, will be responsible for data management, metadata production, dealing with quality issues and the final delivery of data for sharing or archiving. If several people will be responsible, specification of their roles and responsibilities. For collaborative projects explanation of the coordination of data management responsibilities across partners. Description of who will be responsible for the data once the researcher has left the institution.
Reported by	Name of the research fellow, supervisor, co-supervisor and mentor involved.