Symbiosis of smart objects across IoT environments

688156 - symbIoTe - H2020-ICT-2015

Initial Data Management Plan

The symbIoTe Consortium

Intracom SA Telecom Solutions, ICOM, Greece
Sveučiliste u Zagrebu Fakultet elektrotehnike i računarstva, UNIZG-FER, Croatia
AIT Austrian Institute of Technology GmbH, AIT, Austria
Nextworks Srl, NXW, Italy
Consorzio Nazionale Interuniversitario per le Telecomunicazioni, CNIT, Italy
ATOS Spain SA, ATOS, Spain
University of Vienna, Faculty of Computer Science, UNIVIE, Austria
Unidata S.p.A., UNIDATA, Italy
Sensing & Control System S.L., S&C, Spain
Fraunhofer IOSB, IOSB, Germany
Ubiwhere, Lda, UW, Portugal
VIPnet, d.o.o, VIP, Croatia
Instytut Chemii Bioorganicznej Polskiej Akademii Nauk, PSNC, Poland
NA.VI.GO. SCARL, NAVIGO, Italy
ATOS Wordline International, WLI, Spain

© Copyright 2016, the Members of the symbIoTe Consortium

For more information on this document or the symbIoTe project, please contact:
Sergios Soursos, INTRACOM TELECOM, souse@intracom-telecom.com

Document Control

Title: Initial Data Management Plan

Type: Public

Editor(s):   Sergios Soursos

E-mail: souse@intracom-telecom.com

Author(s):   Sergios Soursos

Doc ID: D8.1

Amendment History

Version	Date	Author	Description/Comments
v1	23/06/2016	Sergios Soursos	Initial version prepared
v2	30/06/2016	Sergios Soursos	Final version

Legal Notices

The information in this document is subject to change without notice.

The Members of the symbIoTe Consortium make no warranty of any kind with regard to this document, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The Members of the symbIoTe Consortium shall not be held liable for errors contained herein or direct, indirect, special, incidental or consequential damages in connection with the furnishing, performance, or use of this material.

Table of Contents

(This page is left blank intentionally.)

1 Introduction

symbIoTe participates in the pilots on Open Research Data in Horizon 2020. To this end, the project will identify the research data to be made available via open access. These data can fall under three main categories: i) dissemination material, ii) software source code and iii) pilot data. *Dissemination material* refers to all technical material that is published and may contain details on the project's work. *Software source code* refers to the source code of the middleware software that the project will design and implement, along with its documentation. Finally, the *pilot data* refer to all the raw or processed data that will be used as input in or produced by the envisioned pilots of the project. Since symbIoTe is about interconnecting Internet of Things (IoT) platforms, a number of data and meta-data will be handled by the symbIoTe middleware. Part of these data can be made available to the research community, especially in the case where such data will be used to (numerically or through experiments) evaluate the mechanisms employed by the symbIoTe system in several scientific papers.

This first deliverable on Data Management provides the plans for the collection, storage and use of open access research data. Apart from a selection of platforms that will host and provide access to the various data, it is also important to identify how all the various data will link to each other. Moreover, the proper description of the data will facilitate their discovery by external researchers.

1.1 Structure of the document

Chapter 2 describes the dissemination material that will be available as open access and provides the details of the different categories for such data.

Chapter 3 discusses how the software source code will be made available to the developers community.

Chapter 4 highlights how the pilot data will be collected, stored and accessed, while chapter 5 concludes this deliverable.

2 Dissemination Material

symbIoTe's dissemination material falls under the following main categories:

- Public deliverables
- Scientific papers
- Technical reports
- Public presentations, posters and flyers

This section provides a plan of where the aforementioned material will be stored and how it will be accessed. The particularities of each dissemination category, with respect to storage and access, are highlighted in the subsections below. Regarding the appropriate licensing schemes for certain material (e.g., presentations), the project will consider the *Creative Commons licence "Attribution", CC(by)* (creativecommons.org).

2.1 Public deliverables

The public deliverables of symbIoTe document the progress of the work conducted by the consortium and provide all the necessary details for a reader outside the consortium to understand the work in depth. Deliverables are expected to be accessed by researchers and stakeholders in the IoT market interested on IoT interoperability, IoT platform/application developers, as well as candidate applicants for the symbIoTe's Open Calls.

The primary entry point for accessing the public deliverables will be symbIoTe's web site (www.symbiote-h2020.eu), from the respective section (*"Resources / Deliverables")* (for more details, see Section 4.1 of D7.1 – Initial Dissemination Plans [1]). However, in order to make the deliverables discoverable in a wider scale (without having to visit symbIoTe's web site, all deliverables will be stored in *scribd* (www.scribd.com), an open publishing platform for documents of any kind. Hence, the symboIoTe web site will provide direct links to the respective scribd stored deliverables. This way, the deliverables will be archived and preserved in a popular archiving platform, benefiting from the storage and backup functionality of srcibd.

2.2 Scientific Papers

symbIoTe is aiming at publishing a number of high quality scientific papers, targeting the IoT research community as well as the IoT industry (for more details, see Section 4.6 of [1]). Access to the published articles will follow a combination of the 'gold' and 'green' open access models. The 'gold' and 'green' (after the embargo period, if any) open access will be realized by the symbIoTe web site (www.symbiote-h2020.eu), from the respective section (*"Resources / Publications")*. Additionally, source text may be also available from renowned research article repositories, like *ResearchGate* (www.researchgate.net) or *Acedemia* (www.academia.edu), from the authors personal accounts in those sites. This way, increased visibility of the papers will be achieved.

2.3 Technical reports

The symbIoTe consortium plans to make available technical reports, which will document the progress of the project in a simple and brief but also practical way. Such reports will be provided in the form of blog entries, in the technical blog section (*"Tech Blog")* of the symbIoTe's web site (www.symbiote-h2020.eu). It is envisioned that such blog entries will be the entry points for interested readers to the works of symbIoTe. For more details on specific aspects, links to the respective deliverables (if available) will be also provided. However, blog entries will not strictly follow the deliverables' schedule; they will rather mark certain internal milestones, always with respect to the planned Open Calls and the information required to be available to applicants.

2.4 Public presentations, posters and flyers

Apart from deliverables, scientific papers and technical blog entries, additional dissemination material will be produced throughout the lifetime of the project. Such material includes public presentations in various events, posters and flyers, all promoting the work done by symbIoTe. Such material, though indexed in the symbIoTe web site (www.symbiote-h2020.eu), from the respective section (*"Resources / Outreach material"*), it will be hosted in external platforms, for higher visibility. More specifically, flyers and posters will be stored in scribd (www.scribd.com), while public presentation will be offered by symbIoTe's *Slideshare* (www.slideshare.net/symbiote-h2020) account.

3 Source code

One of the main decisions of the symbIoTe project, is that the generated software and its source code will be available to the Open Source Community from the very beginning of the implementation phase. To this end, the entire source code will be available from the symbIoTe's *github* account (github.com/symbiote-h2020). The Open Source license of the source code is still under investigation and will be documented in the upcoming deliverable D5.1 – symbIoTe Implementation Framework. However, initial discussions consider the following licenses: the Apache License 2.0, the European Union Public Licence (EUPL), and the GNU General Public License (GPL). The decision will be also based on the need to involve the Open Source community as early and smooth as possible. Finally, all required documentation for installation instructions, developers' guide, etc., will be provided in *github* and its respective wiki pages, as well in technical blog entries on the symbIoTe web site (see Section 2.3).

4 Pilot data

symbIoTe will showcase the innovations designed and implemented within the project through a number of small scale proof-of-concept pilots, following the identified Use Cases D1.1 – Initial Report on Use Cases [2]. These pilots will collect a number of data as specified in D9.2 – POPD Requirement No.1 [3], mainly related to sensor readings. The project will consider which of these data can be made available for open access, considering any sensitive information involved, and applying the appropriate curation approaches to remove such information. In this process, the respective procedures identified in D9.1 – POPD Requirement No.2 [4] will be considered and, in fact, the respective consent will have to be acquired before making the data publicly available. Hence, prior to the pilot phase of symbIoTe, the identification of the appropriate data will take place, followed by the acquisition of the informed consent and the ethical approvals. Finally, any sensitive information will be filtered out before being made publicly accessible.

Besides the input data for the pilots, possible data generated by the pilots may also be considered for open access. At this point, it's too early to identify such data; this will happen after the prototype has been implemented and the pilots have been thoroughly designed. However, it is expected that data on which certain analysis is based for the efficiency and performance evaluation of the symbIoTe prototype, possibly also published in scientific papers, will be made available for open access, after following the same procedures for the input data, described above.

4.1 Data Description Template

As specified in the guidelines of the European Commission on Data Management [5], the data to be made available for open access will have to be described using the following template. These descriptions will be stored in the project's internal repository and team collaboration tool (see Section 3.1 of [1]) and will be provided to the EC in the next update of this deliverable (D8.3 – Final Data Management Plan).

Table 1: Data Description Template

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.
Standards and metadata	Reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.
Data sharing	Description of 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. Identification of the repository where data will be

	stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.). In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).
Archiving and preservation	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.

4.2 Data Management Platforms

As the main entry point for the symbIoTe open access data, the platform OpenAIRE (www.openaire.eu) will be used. From there, links to external data repositories will be provided. Candidate repositories for this purpose are OpenDOAR (www.opendoar.org), Zenodo (zenodo.org) and re3data (www.re3data.org). Moreover, links to other sources, especially for 'gold' and 'green' open access scientific papers, will be also provided, so as to provide the link between published papers and deliverables with the respective data used for certain studies that may be included there.

5 Conclusions and Next Steps

This deliverable provides the initial Data Management Plan of symbIoTe. Three categories of data that can be offered as open access has been identified; namely, the dissemination material, the software source code and the pilot data. For these categories, the respective platforms that will host the data have been identified.

Regarding the pilot data, as the project progresses, it will identified what possible data (or even meta-data) can be opened to other parties. Since symbIoTe focuses on the interconnection of IoT platforms and does not offer yet another IoT platform, it is expected that not all data stemming from the interconnected platforms will be accessed by the symbIoTe middleware. However, all possible sensing data have already been identified and it remains to be checked prior to the conduction of the trials, which kind of data can support the evaluation of the symbIoTe offerings so as to be made publicly available.

In this context, symbIoTe will provide an updated and more concrete Data Management Plan, including a first set of pilot data, on M22 (D8.3 – Final Data Management Plan).

6 References

[1] symbIoTe project Deliverable D7.1 – Initial Dissemination Plans; June 2016.

[2] symbIoTe project Deliverable D1.1 – Initial Report on Use Cases; June 2016.

[3] symbIoTe project Deliverable D9.2 – POPD Requirement No.1; February 2016.

[4] symbIoTe project Deliverable D9.1 – POPD Requirement No.2; February 2016.

[5] European Commission, Directorate-General for Research & Innovation, Guidelines on Data Management in Horizon 2020, version 2.1; 15 February 2016.

7 Abbreviations

CC   Creative Commons

EUPL   European Union Public Licence

GNU   GNU's Not Unix!

GPL   General Public License

IoT   Internet of Things

POPD   Protection of Personal Data