Deliverable D6.1

Open Data Use Plan

WP 6

Project Acronym & Number:	SmartCLIDE – GA 871177
Project Title:	Smart Cloud Integrated Development Environment supporting the full-stack implementation, composition and deployment of data-centred services and applications in the cloud
Status:	Final
Dissemination Level:	Public
Authors:	ATB
Contributors:	All Partners
Document Identifier:	SmartCLIDE-D6.1 Open Data Use Plan_v1.0.docx
Date:	30.06.2020
Revision:	1.0
Project website address:	www.smartclide.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No *871177*

Project Partners

Institut für angewandte Systemtechnik Bremen GmbH (ATB), Germany
INTRASOFT International SA (INTRA), Luxembourg
Fundacion Instituto Internacionale de Investigacion en Intelligencia Artificial y Ciencias de la Computacion (AIR), Spain
University of Macedonia (UoM), Greece
Ethniko Kentro Erevnas Kai Technologikis Anaptyxis (CERTH), Greece
X/OPEN Company Limited (TOG), United Kingdom
Eclipse Foundation Europe GmbH (ECLIPSE), Germany
Wellness Telecom SL (WT), Spain
Unparallel Innovation LDA (UNP), Portugal
CONTACT Software GmbH (CONTACT), Germany
Kairos Digital, Analytics and Big Data Solutions SL (KAIROS DS), Spain

Dissemination Level

PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Document Control

Version	Notes	Date
0.1	Creation of the document	16.04.2020
0.2	Add Sections 1, 2 and 3	07.05.2020
0.3	Final contributions	25.06.2020
0.9	Final draft version for review	30.06.2020
1.0	Final version	13.07.2020

Abbreviations

API	Application Programming Interface
BPM	Business Process Management
CSV	Comma Separated Values
DOI	Digital Object Identifier
e.g.	exempli gratia = for example
EC	European Commission
etc.	et cetera
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
GA	Grant Agreement
H2020	Horizon 2020
i.e.	id est = that is to say
IDE	Integrated Development Environment
IoT	Internet of Things
IP	Intellectual Property
IPR	Intellectual Property Rights
JSON	JavaScript Object Notation
REST	Representational State Transfer
UC	Use Case

Executive Summary

This document describes the initial Open Data Use Plan of the SmartCLIDE project and the initial data sets that have been identified to be utilised or to be generated by the four Use Case evaluations for industrial validation of the project technologies. This deliverable also outlines how the research data collected, or generated, will be handled during and after the SmartCLIDE project, describes which methodology for data collection and generation will be followed, and whether and how data will be shared. The data sets are described in accordance with the European Commission guidelines of the Open Research Data Pilot and include the key attributes of data type, format, metadata, use of standards and sharing modalities.

Table of Contents

List of Figures

List of Tables

1 Introduction

The SmartCLIDE project participates in the Open Research Data pilot under the Horizon 2020 Programme and, in the interest of supporting the SmartCLIDE ecosystem, seeks whenever feasible to share open data by making data sets used and created within the project publicly available. This will apply both to data sets used for the Use Case evaluations later in the project (if they are not proprietary), as well as data used in the development or laboratory validation of the work carried out under the research and development tasks, which may not be included as part of the project's formal reporting.

This deliverable outlines the initial Open Data Use Plan for SmartCLIDE, in line with the H2020 guidelines for Open Data Use Plan creation[1]. It identifies the initial classes of datasets that the project foresees to utilise and create primarily with respect to the Use Case evaluations, providing an outline of the type of data, format, metadata and sharing modalities for the data sets identified in the early stages of the project.

The purpose of this Open Data Use Plan is to provide a description of the main elements of the data management policies that will be used by the consortium with regard to all the data sets that will be generated or adopted by the project. The Open Data Use Plan is not a fixed document and is intended to evolve during the operation of the SmartCLIDE project. This initial version of the Open Data Use Plan includes an overview of the data sets to be produced by the project, and the specific conditions that are attached to them.

1.1 Data Lifecycle

The data management plan for the SmartCLIDE project is intended to evolve over the project operation to eventually cover the entire data lifecycle for the data created or adopted by the project. Figure 1 depicts the typical lifecycle for a given data set.

Data lifecycle support is an important aspect related to creating a sustainable SmartCLIDE ecosystem where evolving data sets can help sustain the ecosystems and create opportunities for new applications and services that further exploit the SmartCLIDE technologies for an integrated development environment supporting the full-stack implementation, composition and deployment of data-centred services and applications in the cloud. Supporting policies and procedures concerning SmartCLIDE related data sets will be further analysed and outlined in later deliverables addressing the SmartCLIDE ecosystems, exploitation strategies and associated business models.

[1] See European Commission "Guidelines on Data Management in Horizon 2020 available at: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

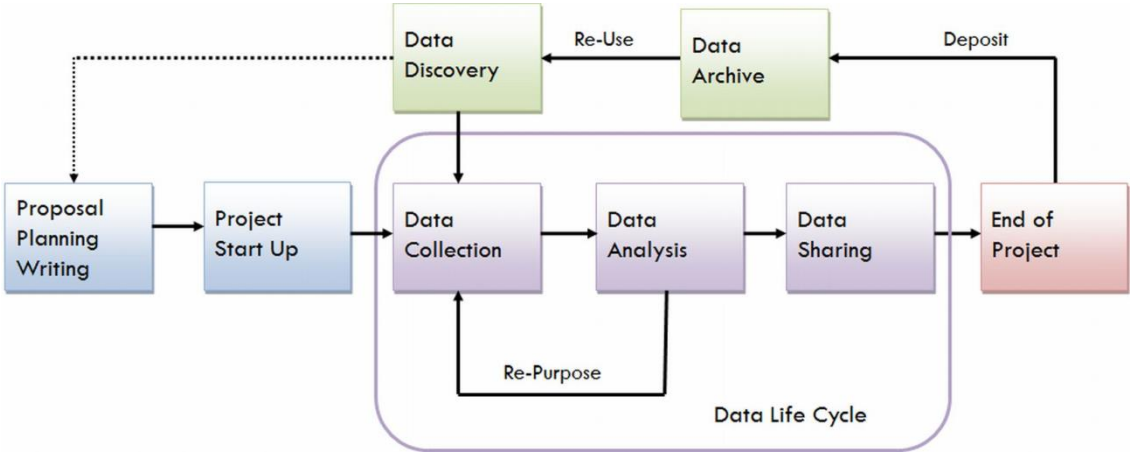

Figure 1: Data Lifecycle Elements[2]

Figure 1 depicts a typical data set lifecycle, although there can be many variations. For example, some Use Case providers may utilise or share raw data from Data Collection sources without any Data Analysis.

Key elements to be considered in managing data within the data lifecycle are the following:

- File formats
- Organisation and naming conventions
- Quality control
- Access modalities
- Persistence and recovery
- Metadata and data conversion
- Sharing and preservation

These elements are addressed for the initial data sets described in this plan in accordance with the European Commission guidelines. At this early stage of the SmartCLIDE project some of the data set information provided should be considered preliminary and subject to change.

1.2 Intended Audience

This deliverable is intended to establish a view of the initial data sets that have been identified for use in the project technology evaluations and development tasks and to document the planning that is already taking place within the project concerning access and preservation of data. This report provides initial guidance on data management to the project partners and is particularly relevant for partners responsible for data collection and Use Case evaluations. It should be considered a first snapshot, which will evolve throughout the project as further details concerning the technologies and Use Case evaluations are specified, and additional procedures and infrastructures are created or adapted for storing and managing project related data.

[2] Source: University of Virginia Library, Research Data Services

1.3 Structure of this Deliverable

This deliverable has been structured with three main sections addressing the different aspects of data management within the SmartCLIDE project as follows:

- Section 2 provides an overview of the initial data policies established for the project
- Section 3 gives an overview about the general Findable, Accessible, Interoperable, Reusable (FAIR) provisions that will be applied by the project for managing data sets
- Section 4 describes the data sets already identified for use in the Use Case evaluations and development tasks, most of which exist, and others will be created during the project operation

The final section provides the conclusions.

2 Project Data Policies

2.1 Participation in the Pilot on Open Research Data

The SmartCLIDE project participates in the Pilot on Open Research Data launched by the European Commission in conjunction with the Horizon 2020 Programme. The consortium strongly believes in the concepts of open research and development, and in the benefits that the European software development community can draw from allowing reuse of data on much larger scales across Europe. Therefore, whenever feasible, data produced by the project can potentially be published under open access procedures – though this objective may be constrained in view of other principles related to IPR and security as described below, as well as partner exploitation interests.

2.2 IPR Management and Security

Many project partners have or will have Intellectual Property Rights (IPR) on the project technologies and data, which for some partners are essential for economic sustainability. The SmartCLIDE project consortium will therefore have an obligation to protect these data and to publish data only if the concerned partner(s) have granted explicit authorisation. Another effect of IPR management is that some of the data collected through the SmartCLIDE project may be of high value for application and service providers and therefore due consideration of the business models and ecosystem should be taken in advance of open access decisions for project data sets. All measures should therefore be taken to prevent data from being leaked or hacked, which could potentially undermine the ecosystem planning for the project or the commercial opportunities for SmartCLIDE project partners. Repositories used by the project for data that have potential commercial value will be secured until decisions are taken concerning open access by the respective partner(s), and in view of the planned business models within the SmartCLIDE ecosystem.

For sensitive data, a holistic security approach will be undertaken to protect the three main pillars of information security: confidentiality, integrity, and availability. The security approach will consist of an assessment of security risks for each data set followed by an impact analysis. This analysis will be performed on the information and data processed by the SmartCLIDE system, their flows and any risk associated with their processing. Particular assessment attention will be placed on any data sets containing personally identifiable information.

2.3 Personal Data Protection

For some of the activities to be carried out by the project, it may be necessary to collect basic personal data (e.g. name, e-mail address, activities, role) for use in the Use Case evaluations, even though the project will avoid collecting such data unless deemed necessary. Such data will be handled in compliance with the EU's Data Protection Directive 95/46/EC1, and more recently the General Data Protection Regulation that came into effect in May 2018, aiming at protecting personal data. National legislations in Germany, Greece, Luxembourg, Portugal, Spain, and the United Kingdom applicable to the project will also be strictly followed.

All personal data collected during the project will be done after giving test subjects full details on the evaluation experiments to be conducted, and after providing options to opt-out of the collection of any personal data. Any existing data sets that contain personal data will be anonymised by the data set provider for use in the project to prevent inadvertent access to personal data by project partners.

Further details concerning the arrangements the consortium partners have put in place addressing specific Ethics topics such as informed consent and protection of personal data will be provided in deliverable *D8.1 – H - Requirement No. 1* (due on month 9 of the project).

3 General Provisions

A set of General Provisions have been established that provide the rules to ensure that the project datasets are **F**indable, **A**ccessible, **I**nteroperable, **R**eusable (FAIR). These are described in the following sections.

3.1 Making Data Findable

3.1.1 Metadata

Metadata are data on the research data themselves. They enable other researchers to find data in an online repository and, as such, are essential for the reusability of the dataset. By adding rich and detailed metadata, other researchers can better determine whether the dataset is relevant and useful for their own research. As described in the project GA (Article 29.2) for scientific publications, the bibliographic metadata should include all of the following:

- The terms "European Union (EU)" and "Horizon 2020"
- The name of the action, acronym and grant number
- The publication date, and length of the embargo period, if applicable
- A persistent identifier

These same attributes have been adopted by the project as the metadata elements to also be used for the project datasets in fulfilment of the requirements related to open access to research data in the GA (Article 29.3).

Note: All publications resulting from the SmartCLIDE project must acknowledge the financial support of the EU by the inclusion of the statement: "SmartCLIDE project has received funding from the European Union's Horizon 2020 research and innovation programme under GA No. 871177."

3.1.2 Persistent and unique identifier

DOI and Creative Common's license numbers will be used as persistent identifiers on open data repositories.

3.1.3 Data discovery

All datasets will be referenced and indexed in the most common internet search engines. Datasets will be therefore immediately discoverable using simple web search capabilities. In addition, Open and Embargoed datasets will be discoverable on the designated open repositories for the project (see Section 3.2).

3.1.4 Data identification

All Open and Embargoed datasets shall have a DOI. This DOI might be the same as the publication for which the data have been used for (e.g. technical publication), if the publisher requires data be uploaded contextually to the publication.

Restricted and Closed datasets do not have a DOI, however, their metadata may have. The metadata shall clearly reference the dataset to which they refer.

3.1.5 Keywords

All Open and Embargoed datasets shall be tagged with the keywords “H2020”, “SmartCLIDE”, “871177”, and at least one additional keyword indicative of the content of the dataset.

3.2 Making Data Accessible

3.2.1 Projects datasets

The SmartCLIDE project datasets will be first stored and organized in a database by the data owners (personal computer or on the organisation’s secure server) and, where appropriate, on the project database. Some datasets, for which the Consortium declares no confidentiality or IPR issues, will be also stored in Zenodo[3], the open access repository of the Open Access Infrastructure for Research in Europe (OpenAIRE). In such cases, the data access policy will be unrestricted. An embargo period may be incurred if collected datasets are linked to green open access publication. Access levels to datasets are specified as:

- O = Open
- E = Embargo, followed by expiry date
- R = Restricted
- C = Closed

3.2.2 Accessibility of files

Most data files are accessible with common and free software:

- .png, .tif, .jpg, .raw: Any image viewer such as XnView, IrfanView, GIMP etc.
- .pdf: Acrobat reader
- .xls, .xlsx: Google sheets, LibreOffice Calc
- .doc, .docx, .odt:, txt Google docs, LibreOffice Writer
- .xml, .csv, .sql: any text editor, or database import tools/viewers.

Additional file types that may be agreed during the progress of the project will be described accordingly within the SmartCLIDE services which produce or use them.

3.2.3 Repositories

All datasets that will be produced within the SmartCLIDE project, and for which the Consortium declares no confidentiality or IPR issues, will be uploaded to Zenodo, with their respective access rights.

3.3 Making Data Interoperable

3.3.1 Data and metadata vocabularies

The Consortium will seek to use standard vocabularies within the datasets that are widely accepted by the respective communities. Controlled vocabularies should be

[3] https://zenodo.org/

used in descriptive metadata fields to support consistent, accurate, and quick indexing and retrieval of relevant data. Keywords (see section 3.1.5) and their synonyms will be used for indexing and subject headings of the data and metadata.

3.3.2 Mapping of non-common ontologies

Non-common, non-obvious ontologies should be explained in the metadata of the individual datasets.

3.4 Making Data Reusable

3.4.1 Data licensing

Creative Common Licensing will be used to protect the ownership of the datasets. Both Share-Alike and Non-Commercial-Share-Alike licenses will be considered for the parts of datasets that can be made publicly available.

3.4.2 Third party access

Third parties must request access rights to Restricted and Closed data according to the procedure indicated in the respective metadata. Access rights to third parties will be evaluated by the data owner on a case-by-case basis. In general, third parties will have to state the purpose concerning why access rights are needed and sign a Non-Disclosure Agreement.

3.4.3 Reasons for restriction

Restricted and closed data are raw experimental results which still need to undergo rigorous evaluation, data which might undermine the beneficiaries' IP protection and commercialisation strategies if published, or data containing sensitive information.

3.4.4 Long-term reusability

Unless otherwise indicated, all Open and Embargoed data will be available for an indefinite time and curated until two years after project conclusion. Restricted and Closed data will be available and curated until two years after project conclusion.

4 Initial SmartCLIDE Data Sets

The different data sets that will be gathered and processed by the SmartCLIDE project are described in the following subsections. The descriptions follow the guidelines provided by the European Commission with respect to data set characteristics. These initial data sets will be updated and extended with additional data sets by the project partners responsible for the different Use Case evaluations to be conducted later in the project, as well as by partners involved in technology development of the core technology components where benchmark and other relevant data sets may be generated or adopted during development. Table 1 provides an overview of the initial datasets identified for the project related to the evaluations to be carried out for the project Use Cases.

Table 1: Summary of initial SmartCLIDE Use Case Data Sets

No.	Data set name	Description
1	BPM Notations of PERSEUS Fundamental Business Flows	BPM notations of PERSEUS fundamental business flows, which may be re-used or adapted to similar needed ones for the further development of the PERSEUS functionalities.
2	PERSEUS Micro-Services Data Set	The data set includes PERSEUS micro-services for generic and thus re-usable functionalities, along with relevant annotations/metadata
3	Static Code Analysis Tests of PERSEUS Software/Micro-Services Code	Static code analysis data of Perseus modules, that includes Perseus Core system, Revenue Management System, Perseus Web-Services, Legal System, Consumer Modules and more.
4	Information from Schedules	Data collected from schedule applications.
5	Voice/Video Sessions	Data collected from finished voice/video sessions.
6	Real-time Suggestions	Data collected from finished voice/video sessions.
7	IoT Catalogue Components	JSON structures for the identification and description of software components in IoT-Catalogue.
8	Enhanced Components' Metadata	Metadata about the software components produced by SmartCLIDE to allow them to be used in development of SmartCLIDE applications.
9	Static Code Analysis	Static code analysis data of relevant applications such as CONTACT Elements for IoT and Platform
10	Productivity Power Report	Data collected from daily work through automatically collecting metrics from SCMs and the in-house installation of CONTACT Elements
11	SmartCLIDE Context Models	The SmartCLIDE context models will describe context information and relations between this information. The context models will form the basis for the context sensitivity modules.

The descriptions provided below of each data set follow the European Commission provided template and includes the following elements:

- **Data Set Name** – used to keep track of different data sets.
- **Contributor(s)** – organisations that are responsible for the data. This can be a project partner, but also external organisations can be contributors.
- **Description** – briefly summarises the type of data elements that exist, or will be created, and the format.
- **Standards** – any standards the data set might follow in the way elements are described or structured.
- **Quality Assurance** – procedures that might be in place to ensure quality is maintained such as consistency or reliability of the data.
- **Access** – indication if the existing data set or data to be created in the SmartCLIDE project will be publicly accessible or restricted.
- **Archiving and preservation** – indicates if facilities for preserving the data are provided by the project partner, a third-party, or the SmartCLIDE project will need to create facilities (e.g. project website).

The characteristics of each of the identified data sets are summarised in the following sections. The following sections summarise the data sets that have been identified with respect to the Use Case evaluation tasks to be carried out in the project.

4.1 Use Case 1: INTRASOFT

4.1.1 UC1.1 – BPM Notations of PERSEUS Fundamental Business Flows Data Set

Data set name or reference	BPM Notations of PERSEUS Fundamental Business Flows
Contributor(s)	INTRASOFT International
Data set description and format	BPM notations of PERSEUS fundamental business flows, which may be re-used or adapted to similar needed ones for the further development of the PERSEUS functionalities. The aim is to support Programming by Example and Low Code paradigm
Standards (if any)	BPMN 2.0
Quality assurance (if any)	-
Access	☐Public access ☒Restricted access ☐Other (please describe):
Archiving and preservation responsibility	☒Project partner ☐External party ☐SmartCLIDE project ☐Other (please describe):

4.1.2 UC1.2 – PERSEUS Microservices Data Set

Data set name or reference	PERSEUS Microservices Data Set
Contributor(s)	INTRASOFT International
Data set description and format	The data set includes PERSEUS micro-services for generic and thus re-usable functionalities, along with relevant annotations/metadata
Standards (if any)	JSON, or any data format when the API Gateway pattern is used when exposing the PERSEUS services to external parties, using Apache Camel.
Quality assurance (if any)	-
Access	☐Public access ☒Restricted access ☐Other (please describe):
Archiving and preservation responsibility	☒Project partner ☐External party ☐SmartCLIDE project ☐Other (please describe):

4.1.3 UC1.3 – PERSEUS Static Code Analysis Tests Data Set

Data set name or reference	Static Code Analysis Tests of PERSEUS Software/Micro-Services Code
Contributor(s)	INTRASOFT International
Data set description and format	Static code analysis data of Perseus modules, that includes Perseus Core system, Revenue Management System, Perseus Web-Services, Legal System, Consumer Modules and more. Analysis includes both overall and per period of time: • Bugs and Vulnerabilities • Code Smells • Overall Coverage • Duplications of code blocks. • Issue analysis • Statistics per Reliability, Security and Maintainability.
Standards (if any)	SonarLint / SonarQube format
Quality assurance (if any)	-
Access	☐Public access ☒Restricted access ☐Other (please describe):
Archiving and preservation responsibility	☒Project partner ☐External party ☐SmartCLIDE project ☐Other (please describe):

4.2 Use Case 2: Wellness Tech

1.1.1 UC2.1 – Information from Schedule Applications

Data set name or reference	Information from Schedules
Contributor(s)	Wellness TechGroup
Data set description and format	Data collected from schedule applications: • Expected duration • Invited participants • Confirmed participants • Scheduled Date
Standards (if any)	CSV
Quality assurance (if any)	-
Access	☐Public access ☐Restricted access ☒Other (please describe): Information is restricted to some users. In case data are accessible via REST API, this is restricted to authorised users.
Archiving and preservation responsibility	☒Project partner ☐External party ☐SmartCLIDE project ☐Other (please describe):

1.1.2 UC2.2 – Quality of Experience Post-Voice/Video Session

Data set name or reference	Voice/Video Sessions
Contributor(s)	Wellness TechGroup
Data set description and format	Data collected from finished voice/video sessions: • Duration • Participants • Room type (audio/video) • Aggregated throughput • Avg. packet loss ratio • Avg. delay • Avg. jitter • avg. satisfaction (from 1 to 5)
Standards (if any)	CSV
Quality assurance (if any)	-
Access	☐Public access ☐Restricted access ☒Other (please describe): Information is restricted to some users. In case data are accessible via REST API, this is restricted to authorised users.
Archiving and preservation responsibility	☒Project partner ☐External party ☐SmartCLIDE project ☐Other (please describe):

1.1.3 UC2.3 – Real-time Suggestions

Data set name or reference	Real-time Suggestions
Contributor(s)	Wellness TechGroup
Data set description and format	Data collected from finished voice/video sessions: • Camera recommendations • Microphone recommendations • Bandwidth consumption recommendation • Camera events • Microphone events
Standards (if any)	CSV
Quality assurance (if any)	-
Access	☐Public access ☐Restricted access ☒Other (please describe): Information is restricted to some users. In case data are accessible via REST API, this is restricted to authorised users.
Archiving and preservation responsibility	☒Project partner ☐External party ☐SmartCLIDE project ☐Other (please describe):

4.3 Use Case 3: Unparallel

4.3.1 UC3.1 – Usage of Components described in IoT-Catalogue

Data set name or reference	IoT Catalogue Components
Contributor(s)	UNPARALLEL, External Partners
Data set description and format	JSON structures for the identification and description of software components in IoT-Catalogue. Each structure provides several fields, including: • Name of Component • Legal owner • Licence • Link for repository (if open source) • Website
Standards (if any)	-
Quality assurance (if any)	-
Access	☐Public access ☐Restricted access ☒Other (please describe): Some information is publicly available in IoT-Catalogue website. Information about some components is restricted to some users. Data access via REST API is also restricted to authorised users.
Archiving and	☒Project partner ☐External party

preservation responsibility	☐SmartCLIDE project ☐Other (please describe):

4.3.2 UC3.2 – Enhanced Components' Metadata described in IoT-Catalogue

Data set name or reference	Enhanced Components' Metadata
Contributor(s)	SmartCLIDE project
Data set description and format	Metadata about the software components produced by SmartCLIDE to allow them to be used in development of SmartCLIDE applications.
Standards (if any)	-
Quality assurance (if any)	-
Access	☐Public access ☒Restricted access ☐Other (please describe):
Archiving and preservation responsibility	☒Project partner ☐External party ☐SmartCLIDE project ☐Other (please describe):

4.4 Use Case 4: CONTACT

4.4.1 UC4.1 – Static Code Analysis

Data set name or reference	Static Code Analysis
Contributor(s)	CONTACT
Data set description and format	Static code analysis data of relevant applications such as CONTACT Elements for IoT and Platform: • Overall Test Coverage • Lines of Code • Coverage on new Code • Issues • Code Smells
Standards (if any)	xml, json
Quality assurance (if any)	-
Access	☐Public access ☒Restricted access ☐Other (please describe):
Archiving and preservation responsibility	☒Project partner ☐External party ☐SmartCLIDE project ☐Other (please describe):

4.4.2 UC4.2 – Productivity Power Report

Data set name or reference	Productivity Power Report
Contributor(s)	CONTACT
Data set description and format	Data collected from daily work through automatically collecting metrics from SCMs and the in-house installation of CONTACT Elements such as: – new and changed documents – new, reviewed and closed change requests – new releases – number of commits from Subversion – number of commits from Git – number of Demo Its! (demonstration work item from our software development process)
Standards (if any)	-
Quality assurance (if any)	-
Access	☐Public access ☒Restricted access ☐Other (please describe):
Archiving and preservation responsibility	☒Project partner ☐External party ☐SmartCLIDE project ☐Other (please describe):

4.5 R&D Data Sets

4.5.1 SmartCLIDE Context Models

Data set name or reference	SmartCLIDE Context Models
Contributor(s)	ATB
Data set description and format	The SmartCLIDE context models will describe context information for cloud development services deployed, developer devices being used, workflow status and history, development teams and collaborations, IDE usage as well as relations between this information. The context models will form the basis for the context sensitivity modules.
Standards (if any)	RDF/XML, OWL/XML
Quality assurance (if any)	-
Access	☒Public access ☐Restricted access ☐Other (please describe):
Archiving and preservation responsibility	☒Project partner ☐External party ☐SmartCLIDE project ☐Other (please describe):

5 Conclusion

This plan provides an overview of the data that the SmartCLIDE project will produce or adopt together with related challenges and constraints that need to be taken into consideration. The analysis contained in this report supports the procedures and infrastructures to be implemented by the SmartCLIDE project to efficiently manage the existing data that will be used, as well as data that will be produced within the project.

It is too early in the project to have a complete identification of the data sets that will be used or created by the project as the initial user requirements are still in progress and substantial design of the SmartCLIDE technologies and identification of required data sets is underway. Some of the data that will need to be collected is not yet sufficiently clear to be detailed with the required level of specification to be included in this preliminary plan, and others will be identified later in the project. This first version of the plan should therefore be considered an initial view that will be updated periodically as the project tasks progress.

By project completion, many SmartCLIDE project partners will be owners or producers of relevant data, in particular those associated with the Use Case evaluations. This implies specific responsibilities and this initial version of the plan is intended to create awareness among the project partners regarding the importance of appropriate procedures with respect to collection, publication in the case of open access data sets, and use of metadata to increase the value of data, as well as persistence of all the information necessary for the optimal use and reuse of SmartCLIDE related data sets to support the SmartCLIDE ecosystem.

Specific attention will be given to ensuring that the data made public breaks neither partner IPR rules, nor regulations and good practices related to personal data protection. For this latter point of view, procedures such as systematic anonymisation of personal data have been anticipated whenever data created within the SmartCLIDE project has potential for misuse or disclosure of personally identifiable information.