Cloud Orchestration at the Level of Application

Project Acronym: **COLA**

Project Number: **731574**

Programme: **Information and Communication Technologies**
Advanced Computing and Cloud Computing

Topic: **ICT-06-2016 Cloud Computing**

Call Identifier: **H2020-ICT-2016-1**
Funding Scheme: **Innovation Action**

Start date of project: 01/01/2017 Duration: 30 months

Deliverable:

D1.3 Data Management Plan

Due date of deliverable: 30/06/17 Actual submission date: 30/06/17

WPL: Gabor Terstyanszky

Dissemination Level: PU

Version: final

1 Table of Contents

2 List of Figures and Tables

Figures

no figures

Tables

3 Status, Change History and Glossary

Status:	Name:	Date:	Signature:
Draft:	Gabor Terstyanszky	24/05/17	Gabor Terstyanszky
Reviewed:	Marcos Rubio Redondo	28/06/17	Marcos Rubio Redondo
Approved:	Tamas Kiss	29/06/17	Tamas Kiss

Table 1 - Status Change History

Version	Date	Pages	Author	Modification
v1	24/05/17	7	G. Terstyanszky	Creating D1.3 template
v1.1	05/06/17	12	G. Terstyanszky	Adding Section 4 + Section 5
	06/06/17	13	N. Paladi	Section 6: WP7 inputs for the Technology DMP
	07/06/17	14	G. Terstyanszky	Adding Section 7
	07/06/17	16	J. M. Martin Rapun	Section 6: WP8 – Inycom + Sarga inputs for User DMP
v1.2	08/06/17	17	S. Budweg	Section 6: WP2 inputs for the Dissemination + Marketing DMP
	10/06/17	18	N. Fantini	Section 6: WP3 inputs for the Exploitation DMP
	10/07/17	20	G. Pattison	Section 6: WP8 – Saker + Brunel inputs for User DMP
v1.3	12/06/07	21	B. Despotov	Section 6: WP4 inputs for the Technology DMP
	13/06/17	22	A Worrad-Andrews	Section 6: WP8 – Outlandish + TAA inputs for User DMP
	15/06/17	24	J. Kovacs	Section 6: WP6 inputs for the Technology DMP
	15/06/17	25	G. Pierantoni	Section 6: WP5 inputs for the Technology DMP
v2	22/06/17	26	G. Terstyanszky	Revising inputs submitted by WP2-WP8 and creating new version of Section 6
v3	26/06/17	27	G. Terstyanszky	Adding Section 6.2
v4	27/06/17	27	G. Terstyanszky	Creating a version for internal review
final	29/06/17	29	G. Terstyanszky	Finalizing D1.3 considering comments of the internal review

Table 2 - Deliverable Change History

Glossary

COLA	Cloud Orchestration at the Level of Application
DMP	Data Management Plan
EC	European Commission
FAIR	Findable, Accessible, Interoperable and Reusable
ICO	Information Confirmation's Office
ISO	International Organisation for Standardisation
PCI-DSS	Payment Card Industry Data Security Standard
TAA	The Audience Agency
VM	Virtual Machine
WP	Work package

Table 3 – Glossary

4 Introduction

In COLA WP1 is responsible for creating DMP and for monitoring its implementation. This work package will coordinate and supervise how COLA data will be collected and/or produced by project partners and users, according to restrictions and rules described in this deliverable. There are two DMP levels: activity- and project-level. Activity-level DMPs are as follows: user, technology, exploitation, dissemination & marketing and project management. The project-level DMP integrates these activity-level DMPs and presents major aspects of the COLA data management strategy. This strategy will enable making COLA data available considering its access level and in a format defined in DMP for usage.

This deliverable presents the COLA Data Management Plan that outlines how COLA will collect, process, publish and store data at project and work package level. DMP defines a framework for managing COLA data to assure full lifecycle data management both during and beyond the project's lifetime. COLA started working on DMP at the very beginning of the project and this work will not end with submission of this report. DMP will evolve as COLA progresses. WP1 will monitor any activities that might affect DMP and upgrade it accordingly. In Section 5 the report lists DMP criteria used to define the COLA DMP. It was developed considering the requirements of activity-level DMPs: user, technology, dissemination & marketing, exploitation and project management. Section 6 first, describes the two DMP levels: activity- and project-level. Next, it presents these activity-level DMPs in a table format with a short explanation of each criterion. Finally, it outlines the project-level DMP based on activity-level DMPs. The report concludes with Section 7 that summarizes major issues considered in the COLA DMP.

5 Data Management Plant's Criteria

This section lists the DMP criteria that will be used in the COLA DMP. These have been selected considering the EC DMP guidelines and the Open Research Data Pilot. Particularly, considering recommendations for full life cycle management through the implementation of the FAIR principles, which state that the data produced shall be **Findable, Accessible, Interoperable and Reusable** (**FAIR**). COLA DMP will implement the FAIR principles at conceptual integration rather than at technical integration.

5.0.1 Data in COLA

Existing and New Data. It must provide a brief description of existing and new data i.e. nature, scope, and scale of the data that will be generated or collected. Good description of the data will help users to understand the characteristics of the data, their relationship to existing data, and any disclosure risks that may apply.

Data Format. The data format must specify the anticipated submission, distribution, and preservation formats for the data and related files. Using a pre-defined format for publishing and sharing data will make the processing and usage of data faster and more efficient.

Producers and Consumers It must describe who will produce, manage and consume the data throughout the data life cycle.

Metadata. This sub-section must explain how data will be described by metadata to enable that data can be effectively used. Metadata must provide all of the needed information for accurate and proper usage. Metadata is preferred to be structured or tagged metadata, like the XML format of the Data Documentation Initiative (DDI) format as standard.

5.0.2 Data Management in COLA

Data Management. It must explain how the data will be managed during the project, with information about version control, naming conventions, etc.

Storage and Backup. It must explain how and where data will be stored to ensure its safety. It must also outline how many copies will be maintained, where these copies will be stored, and how these copies will be synchronized.

Access to Data and Data Sharing. DMP must indicate how COLA intends to archive and share data and why particular options have been selected. Data sharing may include as technical solutions:

- repository such as community, national, international repository, etc.
- web site that COLA will create and maintain.

Data sharing can be either

- self-dissemination - the data producer must arrange for eventual archiving of the data after the self-dissemination period terminates and specify the schedule for data sharing in the grant application.
- delayed dissemination – the data producer must have an arrangement with a public data repository for archival preservation of the data with dissemination to occur.

Archiving and Preservation. It must ensure that data are preserved for the long term. Archiving and preservation will enable active management of digital data over time making it available and usable. COLA is considering depositing data with a trusted digital archive to ensure that they are curated and handled according to good practices. This sub-section must also Indicate how data will be selected for archiving, how long the data will be held,

and what COLA plans for eventual transition or termination of the data collection in the future.

Quality Assurance. It must specify how COLA will ensure that the data meet quality assurance standards because producing data of high quality is essential to the advancement of the project, and every effort should be taken to be transparent with respect to data quality measures undertaken across the data life cycle.

5.0.3 Security, IPR and Ethics in COLA

Security. It must ensure that data is secured over its life cycle. The security plan must outline how raw and processed data will be secured. Raw research data may include direct identifiers or links to direct identifiers and should be well-protected during collection, cleaning, and editing. Processed data may or may not contain disclosure risk and should be secured in keeping with the level of disclosure risk inherent in the data. Secure work and storage environments may include access restrictions (e.g., passwords), encryption, power supply backup, and virus and intruder protection.

Ethics and Privacy. If there are any ethics and/or privacy issues of any data there must be a written consent from the data producers that the information they provide will remain confidential when data are shared.

Intellectual Property Rights. It must describe who will hold intellectual property rights for the data and other information created by the project, i.e. the project consortium or the project partner that produced data. It must also outline whether these rights will be transferred to another organization for data distribution and archiving. Further, it must specify whether any copyrighted material will be used and the project or project partners will obtain permission to use the materials and disseminate them. The project will get a statement from the data producer of who owns the data to enable its dissemination.

Legal Requirements. It must indicate whether any legal requirements apply to archiving and sharing data. It is important to define how the project will manage legal requirements if there are any because some data may have legal restrictions that impact data sharing. This sub-section must describe these issues that might impact data sharing.

6 COLA Data Management Plan

COLA produces and manages the following major data types:

- use case data,
- technology data
- exploitation data
- dissemination and marketing data, and
- project management data.

Considering this wide range of data types the COLA Data Management Plan incorporates activity-level DMPs that address specific aspects and requirements of these data types.

WP8 elaborated the User DMP that outlines how the three COLA use cases: Social media data analytics for public sector organisations use case (Inycom + Sarga), Scalable hosting, testing and automation for SMEs and public sector organisations use case (Audience Agency + Outlandish), and Evaluation Planning Service use case (Saker + Brunel University) handles data. The first use case will handle social media data, the second one both company and social media data while the third one only company data. As a result, they have to manage different data types available in different data formats, stored and backed up at different locations and following different approaches and implementing multiple security measures to protect data. This heterogeneity will further increase when WP8 will implement another 20 proof of concept use cases using the COLA infrastructure and the MiCADO platform. It will require extending the User DMP.

The Technology DMP describes how data is handled in the COLA infrastructure and in the MiCADO platform. Key contributors to this DMP are WP4 (COLA infrastructure) and WP5-WP7 (MiCADO platform). The COLA infrastructure is IaaS while the MiCADO platform is PaaS that enables running use cases as SaaS. The COLA infrastructure incorporates one commercial cloud (CloudSigma) and three academic clouds (SICS, SZTAKI and UoW). As a result, similarly to the User DMP the Technology DMP also has a wide range of data management requirements.

There will be further three activity-based DMPs: Exploitation DMP, Dissemination and Marketing DMP and Project Management DMP.

To describe these activity-based DMPs WP1 developed a table (See Annex I) based on DMP criteria listed in Section 5 to define and describe DMPs listed above.

6.1 Activity-level DMPs

6.1.1 User DMP

Work package: WP8
Person: Jose Manuel Martin Rapun

Data
Existing and New Data
Social media data analytics for public sector organisations use case (Inycom + Sarga) This use case will produce new data processing data that will be collected from Twitter using Twitter APIs. There are two existing data types that will be used in this use case: Tweets: Data posted by Twitter users matching a list of keywords of interest for the Regional Government (tourism in the region, employment, etc.). This data will contain attributes, such as tweet id, user who posted the tweet, creation date, tweet content, number of retweets. There will be also calculated attributes, such as category among those defined in the keywords and sentiment analysis. Users: Data about Twitter users, for example about users who posted tweets related to issues of interest for the Government of Aragón. This data will contain attributes, such as user id, user name, location, date of birth, number of followers, number of tweets, users following. There will also be calculated attributes, such as activity of the user in twitter, influence of the user and sentiment analysis. **Scalable hosting, testing and automation for SMEs and public sector organisations use case (Audience Agency + Outlandish)** This use case will use the following existing data types: Ticketing Data: TAA has data processing agreements with all its clients in order to collect and process ticket sales data. This data is anonymised during the data transformation process before it reaches any part of the infrastructure which will be developed within the use case. From this form it cannot be traced back to the individual customer. Customers are identified by a key number which itself does not provide any information on the person. Certain tags derived from demographic data, such as Experian Mosaic codes and bespoke segmentation tags are also stored, along with customers' postcodes. Customer Surveys: Separately to ticketing data, responses to customer surveys run by arts organisations are stored in a different part of the system. These data might contain some basic demographic information along with a postcode. While TAA does not collect personal details, survey responses might contain pieces of socio-demographic information which might be considered as sensitive. These surveys collect the explicit consent of the user to process the requested sensitive data. Business information: Some of the data processed by TAA can also be regarded as business sensitive to its clients, e.g. ticket sales information or website analytics metrics. There will be agreements between use case partners that will explicitly state that any customer or organisation in any output of analysis or reporting will never be identified. Publicly available data: The collection and processing of other data will be from sources that are open to the public i.e. social media platforms such as Twitter. The types of data that are collected vary from user names, (reported) gender, age, sex, posts content and particular hashtags. In addition, the data collected from open social networks (i.e. Twitter) will be subject to the same laws governing voluntary disclosure detailed above in the Inycom case study on public social networks and be handled by the terms of the Twitter Privacy Policy and related third party development licenses. This use case will produce new data, for example summaries produced from the data

types mentioned above.

Evaluation Planning Service (Saker + Brunel University)

This use case will generate two data types:

Simulations models: A given model will be run with a defined scenario. The model will contain the definition of the logic and data structures; a database will store all input data and results.

Simulation data: Simulation data will be produced by the simulation runs.

Data Format

Social media data analytics for public sector organisations use case (Inycom + Sarga)

The collected data of this use case will be exchanged and distributed between the different system components in JSON format using web services.

Scalable hosting, testing and automation for SMEs and public sector organisations use case (Audience Agency + Outlandish)

The data of this use case will not have a standardised format. It will be simple formatted and transparent text. It will have the processing of the data encoded in code. In addition raw text and CSV files might be used in data pipelines. This data will be stored in relational and non-relational databases.

Evaluation Planning Service (Saker + Brunel University)

Simulation models will be in proprietary Flexsim format (.FSM). These will contain the definition of the simulation model which encompasses logic, data structures and objects (processes, queues, etc.). Simulation data, such as input parameters and key performance indicators (results data) will be stored in SQL database. This data will be held in tables that cross reference each other. Further, data will be organised into scenarios and datasets where a scenario being defined as a base dataset that is overridden by the data in an ordered set of subsequent datasets for a sub-set of the data points. All data points belong to datasets that will be hosted on a server residing on the web.

Data Producers and Consumers

Social media data analytics for public sector organisations use case (Inycom + Sarga)

Data producers: Twitter users posting information about topics of interest to the Aragon Regional Government.

Data consumers: In the COLA project the data consumers will be civil servants in the Aragon Regional Government eAdministration.

Scalable hosting, testing and automation for SMEs and public sector organisations use case (Audience Agency + Outlandish)

Data producers: Arts organisations, TAA (for data summaries), public social media companies as producers of social media data. The two organisations processing data will be the Audience Agency and Outlandish. No third parties will be used in order to process data.

Data consumers: The Audience Agency staff, selected clients (for suitably selected subsets of data and data summaries), public (for non-sensitive data summaries and open data)

Evaluation Planning Service (Saker + Brunel University)

Data producers and the Data consumers will be the same entity. This will be the end user of the simulation model in the client organisation. The end user will create a scenario to be simulated. The simulation model will run which to generate a series of KPIs. These will be reviewed by the end user and where appropriate communicated to any interested parties within the client organisation.

Metadata

Social media data analytics for public sector organisations use case (Inycom + Sarga)

Data in this use case will be stored in a structure defined using xml format in SOLr.

Scalable hosting, testing and automation for SMEs and public sector organisations use case (Audience Agency + Outlandish)
A set of tags specifying data range and type, e.g. genre, art form, venue, year, data source. In general metadata itself is the result of the relational databases being relational. If metadata is stored it will be done within databases themselves.
Evaluation Planning Service (Saker + Brunel University)
This use case will not use any metadata to describe use case data.

Data Management

Storage and Backup

Social media data analytics for public sector organisations use case (Inycom + Sarga)
The data will be stored in databases, namely SOLr for the tweets, and probably MySQL for users (a NoSQL option such as JENA TDB or MongoDB will be explored). There will be weekly backup but daily incremental backups will be also considered. The Twitter data will be reprocessed using the Twitter API at least for one month. As a result, the above outlined backup policy is considered proper and safe. The backup copies will be stored in servers located in in different locations and managed by different providers.
Scalable hosting, testing and automation for SMEs and public sector organisations use case (Audience Agency + Outlandish)
Use case data will be stored in a series of databases managed by Outlandish. The majority of these systems are immutable and able to be restored without requiring backups from either configuration as code or machine images. Backups will be taken every day and will be kept for a week, unless a different schedule is identified as most suitable in the course of business requirements gathering. There will be also regular restoration rehearsals. Use case documentation and information will be stored using SharePoint with continuous backups.
Evaluation Planning Service (Saker + Brunel University)
All simulation models will be stored by simulation users. This could be Saker Solutions or Saker's client. These models will typically be held on a local server that is backed up outside of the COLA project. The SQL Server Database will reside on a server that is regularly backed up by the company that is responsible for hosting the database. For example, at Saker Solutions the database servers are backed up daily.

Access to Data and Data Sharing

Social media data analytics for public sector organisations use case (Inycom + Sarga)
The dissemination and exploitation of the data will be done by means of a dedicated website interface (self-dissemination), where the users will be able to filter and analyse the data. They will also have the ability to print the charts and export the data in the tables to csv format.
Scalable hosting, testing and automation for SMEs and public sector organisations use case (Audience Agency + Outlandish)
Confidentiality Agreement will be signed separately with each organisation joining the warehouse. Any publicly available data (e.g. from Twitter, Open Geography portal) that will be processed will be always obtained through legal means and in accordance to relevant terms and conditions agreements. Only anonymised and aggregated information will available in the publicly accessible parts of the system (i.e., the dashboards). All non-publicly available data processed by TAA is provided by their clients (arts organisations).
Certain data summaries might be made publicly available, although in each case this will have to undergo a review process to ensure that no sensitive information is disclosed.
Evaluation Planning Service (Saker + Brunel University)
All simulation models, input data and results are confidential. Access to data is dependent upon the sensitivity required by the client and may be subject to prior formal authorisation. The data for each project that is run is specific to that project and may only be accessed by persons who are authorised to do so.

Archiving and Preservation
Social media data analytics for public sector organisations use case (Inycom + Sarga) Long term archiving (> 1 year) of Twitter data could be a challenging task because of its would be huge volume. Thus, SOLr data older than 1 year will be removed from the production servers and archived at least for one year. User data will be much smaller in volume and it will be more important in long term than Twitter data. As a result, it will be kept in production environment for a longer time (3-5 years) and archived every year. **Scalable hosting, testing and automation for SMEs and public sector organisations use case (Audience Agency + Outlandish)** Each data backup will be archived for the period of one week **Evaluation Planning Service (Saker + Brunel University)** Data is to be backed up every 24 hours but use case data will not be archived.
Quality Assurance
Social media data analytics for public sector organisations use case (Inycom + Sarga) The quality and value of the data will be monitored by the users. Wrong data will be either corrected or removed using tools. The main risk will be that fake data can be loaded in the system. For example, a Twitter search may provide data that is not relevant because the search filters (keywords) were not properly set. To sort out this issue the search filters must be improved when irrelevant data is detected. **Scalable hosting, testing and automation for SMEs and public sector organisations use case (Audience Agency + Outlandish)** Each data backup will be archived for the period of one week Data will be regularly checked for consistency and quality, through automatic and (occasionally) manual checks. **Evaluation Planning Service (Saker + Brunel University)** This use case will use the Saker Solution's Quality system. It will be updated to cover any quality related matters pertaining to data uploaded to the Cloud for running this use case.
Security, IPR and Ethics
Security
Social media data analytics for public sector organisations use case (Inycom + Sarga) Both data types are not sensitive and have been publicly disclosed by the Twitter users themselves. The access to the web interface will require user/password with different access levels. Two major user groups will need access: administrators (can change configuration such as keywords, crawlers) and end users (only enabled to look up the data). Regarding the infrastructure servers, at least firewalls and access restricted by IP will be used, apart from those security strategies/features provided by the COLA platform. **Scalable hosting, testing and automation for SMEs and public sector organisations use case (Audience Agency + Outlandish)** It is company standard at Outlandish that developers must encrypt their hard drives as well as all S3 buckets they use to store sensitive data. They also dispose of any Amazon Web Server (AWS) instance following all AWS protocols and standards. If necessary these AWS servers can be encrypted at rest. When AWS servers are shut down they are safely destroyed to standards that are acceptable to government agencies. TAA uses multi-factor authentication for access to their systems and only selected staff members have access to full data sets. Full access to databases and servers will be granted on a need-to-know bases to selected staff members. **Evaluation Planning Service (Saker + Brunel University)** All simulation models, input data and results of this use case are confidential. Access to this data will depend upon the sensitivity as defined by the client and may be subject to prior formal authorisation. The data for each project is specific to that particular project and may only be accessed by users who are authorised to do so. It will at least need to be protected from 3rd party access and in some cases will be such that it can only run on

private networks. Considering these security requirements Saker will run this use case on the Cloud, for example G-Cloud, and on a private infrastructure, for example hosted at Saker.

Ethics and Privacy

Social media data analytics for public sector organisations use case (Inycom + Sarga)

This use case involves the collection and processing of personal data and information that are not sensitive. The use case will follow a comprehensive approach based on the "Data protection by design" principle. This approach will protect data from the first stage up to the final stage. It is important to highlight that the use case does not collect data directly from individuals only from second sources, such as social networks, for example Twitter, and tools owned by the end users, for example tool of the Aragon public administration. Additionally, it must be noted that the information gathered is just information that individuals post voluntarily and disclose publicly, so there is no need in obtaining the prior consent of individuals. Therefore, this use case will sign an agreement with the data provider (the so called Developer Agreement, which will grant the license to use Twitter API and Content). This agreement will be based on the Privacy Policy that individuals accept when joining the social network. On the other hand, concerning data provided by the public administration, they will be the responsible of collecting the information by their own means. This structure of the information flow will be configured through another file and the corresponding agreement entitling us to process such data. As a result, this use case will comply with applying rules on data protection and will ensure the rights of the individuals within the framework of this use case.

Scalable hosting, testing and automation for SMEs and public sector organisations use case (Audience Agency + Outlandish)

The data that will be accessible or processed in the course of the project is not personal information and by ensuring the proper care and attention in processing and output, this data does not become personally identifiable. Some data is personal sensitive data or business sensitive data. This use case will only collect and process this data in line with appropriate regulations and signed agreements with clients and data subjects. This data is subject to strict collection, storage, retention and destruction protocols and TAA is registered with the ICO. All Outlandish employees are well versed in data handling and relevant legislation. All Outlandish employees sign agreements to this effect and abide by a strict non-disclosure agreement.

Evaluation Planning Service (Saker + Brunel University)

All data is confidential and belongs to a client company (or Saker Solutions in the case of prototypes / demonstrators / internal projects). There are no specific ethical issues related to this use case.

Intellectual Property Rights

Social media data analytics for public sector organisations use case (Inycom + Sarga)

According to the Spanish law based on the European law, databases are protected under intellectual property rights as long as they are intellectual creations with the requirements of originality and creativeness; and it is structure of the database, the "container" what is protected, but data themselves, the "content", are outside the scope of such protection. This use case involves the creation of databases, but they are not intellectual creations according to the terms mentioned before, so there are no intellectual property rights concerns.

Scalable hosting, testing and automation for SMEs and public sector organisations use case (Audience Agency + Outlandish)

TAA has agreements in place with all data suppliers for the perpetual use of any non-public data. Outlandish will endeavour to make as much software as is possible during the course of the COLA project open source under permissive or share-alike GPL licenses.

Evaluation Planning Service (Saker + Brunel University)

All data will belong to Saker Solutions and / or the client (end user) company.

Legal Requirements

Social media data analytics for public sector organisations use case (Inycom + Sarga)

This use case is to be implemented in Spain, so the Spanish and European laws will be taken into account. Currently, the *Organic Law 15/1999 on Protection of Personal Data,* which transposed the *Directive 95/46/CE on the Protection of Personal Data* into the Spanish legislation, and the additional rules implementing these two ones. From May 2018, Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (directly effective and applying in all EU countries). For the moment, it will be considered as an inspiring framework. The framework created by these laws under which this use case will be developed is based on the following principles and guidelines

- transparency, legitimate purpose, and proportionality that means that data will be:
  - processed fairly and lawfully
  - collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes
  - adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed
  - accurate and, where necessary, kept up to date
  - kept in a form which does not permit the identification of data subjects
- data protection by design, to provide adequate protection from the mere design of the system
- ensure the exercise of the rights of the subjects (which will be expanded with the new Regulation)

Finally, it must be noted that this use case does not involve the processing of sensitive data, so the specific and tougher requirements about them do not apply.

Scalable hosting, testing and automation for SMEs and public sector organisations use case (Audience Agency + Outlandish)

TAA serves as a data processor providing research services to their client organisations under the research exemption of the Data Protection Act. The Audience Agency does not collect data directly from data subjects. TAA obtain data such as ticket sales records or audience/visitors survey response from their client organisations and process this data as part of research services. It is a requirement that before any client organisation is allowed to commit data, a Data Use and Confidentiality agreement is signed, where the client organisations warrant that correct notifications have been given and consent have been obtained from data subjects. Processing of all these data is absolutely critical to TAA's mission to provide arts organisations with insight on demographics of their customers and their ticket sales patterns within various socio-demographic groups, as well as information on how well the arts organisations serve different parts of the community. This data will not be made publicly available. The only data that might be made publicly available will be non-sensitive summaries of data and open data. Any data obtained from third parties (through TAA's services that process client's ticketing data, customer surveys, business information and commercially licensed datasets – see also below) are obtained with their full consent and conforms to EU wide standards of transparency, information, access, erasure and so on. TAA are registered as a data controller with the Information Commissioner Office, with registration number ZA009719

Evaluation Planning Service (Saker + Brunel University)

Data usage and confidentiality agreement must be signed by developers and users.

6.1.2 Technology DMP

Work package: WP4-WP7
Persons: WP4 Bogdan Despotov
WP5 Gab Pierantoni
WP6 Jozsef Kovacs
WP7 Nicolae Paladi

Data
Existing and New Data
Technology specific data will be IaaS, PaaS and Saas type data. The COLA infrastructure incorporates one commercial IaaS platform (CloudSigma) and 3 research IaaS platforms (SICS, SZTAKI and UoW). The infrastructure owners will create and publish infrastructure related data as WP4 partners. COLA developers and application users will generate and use Paas and SaaS data produced in WP5-WP7. WP6 will elaborate the MiCADO platform to support dynamic and secure deployment and run-time orchestration of cloud applications. This work package will develop and manage source codes and Docker/Virtual Machine images of the MiCADO services but it will not handle data generated by COLA applications. WP5 in cooperation with WP8 will describe applications creating TOSCA based Application Description Templates to specify the cloud applications' service topologies and their policies. WP4, WP6 and WP7 in collaboration with WP5 will define deployment, performance, scalability and security policies.
Data Format
At one side there will be no specific data format to manage the COLA infrastructure data. WP4 will manage a wide variety of infrastructure specific data. At the other side WP5 and WP6 will specific data formats. WP5 will describe applications using TOSCA based on YAML. It will also manage deployment and implementation artifacts of COLA applications. WP6 will handle data of the MiCADO platform data, such as binaries of MiCADO services, scripts, Docker and Virtual Machine images using Occopus and TOSCA descriptors.
Data Producers and Consumers
Data producers in COLA will be system administrators of the COLA infrastructure in WP4, COLA application developers in WP5 and WP8, and MiCADO platform developers in WP6-WP7. Developers in WP5-WP7 will be data consumers of the COLA infrastructure data while users in WP8 will be consumers of the TOSCA based application descriptions and MiCADO specific data.
Metadata
WP4 will not use metadata to manage COLA infrastructure. Data. In contrast, WP5-WP7 will widely use metadata to support access to the MiCADO platform and re-usability of COLA applications. WP5 and WP8 will describe each COLA application in TOSCA. These descriptions will also contain descriptive metadata defined in the TOSCA specification. Additional metadata can be added to TOSCA Application Descriptions to support their sharing in digital markets. WP6 will add metadata to Docker and Virtual Machine images to describe MiCADO services and COLA applications they contain. This metadata may contain metadata such as version controller code, previous versions, release tags and commit times, etc.
Data Management
Storage and Backup
WP4 will use a Storpool storage, located in in Zurich, to store data about the COLA

infrastructure. It protects data and guarantees data integrity via a 64-bit checksum and version for each sector maintained by the storage system. WP4 will provide several existing CloudSigma backup solutions, for example the snapshot functionality, among those service owners can select the required backup solutions.

WP5 will upload TOSCA based applications descriptions to GitHub. Backup copies will be also available on COLA Pydio (https://cola.fst.westminster.ac.uk) and gdrive. After installing the COLA repository and digital market product-quality application descriptions will be also stored in these facilities. WP6 will also store binaries, source codes and documentation of MiCADO services in GitHub. Documentation of MiCADO services will be also uploaded to the COLA Pydio and website. WP6 will publish Docker and Virtual Machine images on the Docker hub.

Access to Data and Data Sharing

WP4 will provide access to data of the COLA infrastructure from within the COLA infrastructure.

Both applications descriptions including their artifacts such as deployment and implementation artifacts (WP5) and binaries, images and source code of MiCADO services (WP6) will follow the Open Access policy. As a result, they will be publicly available, no restriction is planned regarding their accessibility.

Archiving and Preservation

In WP4 CloudSigma uses a block storage system in the COLA infrastructure to archive and store VMs. This block storage solution is able to provide implicit non-disruptive backups at the storage block level for all user data. This includes any data contained within virtual drives including application data, databases, all operating system information etc. It provides full drive level backup of customer data. It backs up all end-user computing data each night and retains seven days of rolling snapshots. In addition to the automatic backup system, users are able to create point--in--time snapshots of their drives, which can later be cloned and upgraded to create stand-alone drives. A snapshot can be created on-demand while the server is running, thus in no way affecting the performance or availability of the systems. By using snapshots, customers can protect themselves from data corruption or use them for auditing purposes.

WP5 and WP6 uses archiving and versioning services of GitHub and Docker hub to create back-ups of application descriptions, binaries and source codes of MiCADO services and images applications and MiCADO services.

Quality Assurance

The integrity of the COLA infrastructure data is guaranteed by Storpool's storage solution. Redundancy is provided by multiple copies (replicas) of the data written synchronously across the cluster. Users can set the number of replication copies, with the CloudSigma cloud configured to store three copies of all data. This technology is superior to RAID in both reliability and performance. Unlike RAID, the system replication distributes copies across different servers. As such, in the case of a server or component failure, data that is stored on this affected server is not lost.

The integrity of application description and MiCADO service will be guaranteed by GitHub where these types of data are stored. The integrity of the image data is managed and provided by Docker hub.

Security, IPR and Ethics

Security

Each COLA partner data security policies must comply with the criteria set out in the COLA DMP strategy document (D1.3 deliverable). CloudSigma's cloud solution being the only commercial infrastructure made available for use in the project is ISO-27001 (2015) certified and PCI-DSS compliant. In addition, CloudSigma's data centre in Zurich is

covered by the following certifications:

- SAS 70 compliant data centre
- ISO-9001:2008 for quality management systems
- ISO-27001:2005 for information security management systems
- Gold LEED certification for environmental sustainability
- FACT certification – key data protection certification for the European film and broadcasting industry

Application descriptions (WP5) and MiCADO platform data (WP6) will require access control mechanisms to be used for data confidentiality and integrity protection. These data types do not have any further security issues because they are available and its accessibility is not restricted.

Ethics and Privacy

WP4-WP7 do not foreseen any ethics and/or privacy issues in the project.

Intellectual Property Rights

Intellectual property rights related to WP4-WP7 will be managed in accordance with the IPR management plan outlined in the DoW (Section 3.2.6 Management of Knowledge and Intellectual Property), the strategy for addressing issues formalised in the Consortium Agreement, as well as monitored and controlled by T3.2. The Consortium Agreement lists all background included and excluded from excess right.

Legal Requirements

One of the key factors regarding data protection relates to the physical location of stored data and the implications of the differences in legislation and regulation between jurisdictions. The four COLA infrastructure providers (SICS, CloudSigma, SZTAKI and UoW) in WP4 reside in four European countries. CloudSigma provides its commercial IaaS platform in Zurich (CH), but also makes available its resources located in Frankfurt (DE) if required. SIC academic cloud is located in Lulea (SE), SZTAKI academic cloud in Budapest (HU) and UoW academic cloud in London (UK). Data protection laws are consistent across EU countries, due to the EU Data Protection Directive (Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data) adopted in 1995. However, there are some differences in legislation and regulation between Switzerland and the EU, as Switzerland only partially implemented the EU Directive on the Protection of Personal Data in 2006. One of the main differences is that, unlike the data protection of many other countries, the Swiss Federal Data Protection Act (DPA) protects both personal data pertaining to both natural persons and legal persons. Special requirements apply to the transfer of personal data outside of Switzerland. Depending on the circumstances, the Swiss Federal Data Protection and Information Commissioner must be informed before personal data is transferred outside of Switzerland. This is an important factor for companies or individuals storing sensitive information if they want to circumvent the US Patriot Act or the US Safe Harbour or Data Protection acts. CloudSigma has set-up its corporate structure in such a way that each cloud location is managed by a local entity and therefore subject to that jurisdiction. This allows customer data to be treated in accordance with the country where it is physically residing, essentially enabling customers with sensitive information to circumvent the Patriot Act. CloudSigma's holding company is Swiss, which means it has no concept of extra-territorial jurisdiction (unlike US holding companies). This means CloudSigma's US entity is subject to US law only and their Swiss cloud location is subject to Swiss law only. New operational companies are opened for each new location. This way, CloudSigma's customers can make informed decisions about where they store their data according to the data protection laws relative to jurisdiction.

WP5-WP7 has no legal requirements that are foreseen.

6.1.3 Exploitation DMP

Work package: WP3
Person: Nicola Fantini

Data
Existing and New Data
WP3 will generate three types of data: exploitation, IPR and sustainability data. Sustainability-related data. To validate the economic feasibility of the implemented COLA use cases WP3 will collect the sustainability-related data. COLA use case owners will provide this data based on different business models e.g., detailed use case description, partner resources, customer data, revenue structure, cost structure, value proposition details, etc. Exploitation-related data. To contribute to the commercial exploitation planning WP3 will collect exploitation-related data from COLA partners. They will forward their exploitation plan WP3, including specific metrics that will be used to measure the economic impact of COLA use cases implemented as cloud-based solutions on the MiCADO platform. IPR-related data. To investigate and handle IPR management issues, WP3 will request IPR-related data from the partners, such as data related to any IP brought to the project, IP generated in and out of the project. WP3 will process and describe all three types of data in D3.1-D3.3. The reports will be used by other COLA WPs and also submitted to EC.
Data Format
Most of data will be collected and produced in .doc and .pdf formats. .pptx format will be also used for better data presentation and visualization.
Producers and Consumers
Data producers. COLA project partners will provide sustainability, exploitation and IPR data about the COLA infrastructure, the MiCADO platform and COLA use cases. WP3 will process and describe this data in D3.1-D3.3. Data consumers. COLA work packages will be the consumers of exploitation, IPR and sustainability data presented in D3.1-D3.3. The key consumers will COLA use case owners and WP2. COLA use case owners will use this data to improve exploitation ad sustainability of their applications. WP2 will use public exploitation and sustainability data in dissemination activities to promote COLA, particularly, how SMEs can use a cloud-based platform to run their applications.
Metadata
WP3 will not use metadata to describe sustainability, exploitation and IPR data.
Data Management
Storage and Backup
All WP3-related data, such as sustainability data provided by partners, exploitation data and IPR data collected, corresponding deliverables and reports, agendas and minutes of meetings, etc. will be uploaded to the COLA storage - a Pydio repository - available at https://cola.fst.westminster.ac.uk/.
Access to Data and Data Sharing
The Pydio based COLA storage has access rights based access control. Since exploitation and sustainability data is private the repository's access control enables only COLA project partners to manage (upload, search, select, edit, etc.) these data types.
Archiving and Preservation
There is no specific archiving policy on data collected and/or processed by WP3. The data is stored in the format it was originally collected or provided in the COLA storage.
Quality Assurance
There are three phases of quality control of exploitation, IPR and sustainability data. In the first phase WP3 will process the collected sustainability, exploitation and IPR data checking whether there is any missing and wrong information. This processing includes

correcting, excluding and finalizing this data. In phase 2, data producers, such as COLA infrastructure provider, MiCADO platform developers and COLA use case owners will check the finalized data. This data will be used to compile WP3. In phase 3 D3.1-D3.3 data will be examined whether it accurate, whether it has the needed quality, etc. by the internal COLA review process and modified as required.
Security, IPR and Ethics
Security
All WP3 data is stored in the COLA repository. This storage facility has access rights based control that enables/disables access to the data and documents uploaded and stored in the repository. As a result, only Pydio users with the proper access right can reach WP3 data.
Ethics and Privacy
The exploitation, IPR and sustainability data does not raise any specific ethical issue. See report D9.1-D9.2. Exploitation and sustainability data of COLA use cases is confidential. This data can be shared among project partners only i.e. neither data nor WP3 reports will be available to the general public. In contrast these data types of the MiCADO platform are not confidential and COLA will make it available to the general public.
Intellectual Property Rights
COLA use case owners will hold IPRs for exploitation and sustainability data and other information they produced. Since this data is confidential the IPRs will not be transferred even to project partner for data distribution and archiving but they partners will be allowed to use this data as long as they follow the confidentiality requirements. To disseminate this data WP2 should get a statement from the use case owner who owns the data to allow its dissemination.
Legal Requirements
Confidentiality of exploitation and sustainability data of COLA use cases raises the issue of how archive and share this data. WP3 does not archive these data types. Their sharing is implemented through the COLA storage at technical level. As a result, there is no any specific legal issue that must be addressed in WP3

6.1.4 Dissemination and Marketing DMP

Work package: WP2
Person: Andreas Ocklenburg / Steffen Budweg (CloudSME UG)

Data
Existing and New Data
WP2 will develop two types of data: dissemination and marketing data and administrative data. Dissemination and marketing data will incorporate academic and commercial publications, such as COLA leaflets, posters, etc. web content on COLA and other website, images, etc. WP2 will produce project deliverables and interim dissemination and marketing reports as administration data. There will be the following deliverables: D2.1 Dissemination Plan (M03), D2.2 First periodic dissemination report (M12), D2.3 User community feedback (M24), D2.4 Final dissemination report (M30), D2.5 Report on standardisation activities (M30). This data type will also include data relevant to the work in WP2, such as agendas, minutes, interim documents, publications, etc.
Data Format
Dissemination and marketing data will be mostly produced in doc and pdf format. This data might include different data types generated by WP2 and project partners (e.g. marketing material with images, academic and commercial publications, web content, etc.)
Data Producers and Consumers
Data producers: All project partners will contribute to the dissemination activities under coordination of WP2. They will collect and forward their dissemination data WP2 that will produce dissemination and marketing data, and lead dissemination and marketing activities, plus compiling and submitting periodic dissemination and marketing reports and WP2 deliverables. Data consumers: They will be public services and SMEs targeted by WP2 dissemination and marketing activities.
Metadata-
WP2 will use metadata to describe data (e.g. metadata for publications, web content and images etc. WP2 will use the following metadata protocols and standards: • ISO 19005-1:2005 standard with compliance to PDF/A1 for long term archival of documents • ISO 16684-1:2012 standard (XMP) for metadata of documents and images • ISO 15836 Dublin Core for Metadata Element Set
Data Management
Storage and Backup
All WP2 data including dissemination/marketing materials (flyers, images, posters, etc.) and administrative data (WP2 deliverables and reports) will be uploaded to the COLA storage - Pydio repository - available at https://cola.fst.westminster.ac.uk/
Access to Data and Data Sharing
The COLA storage has access rights based access control. All WP2 data but administrative data such as agendas and minutes of WP2 meetings is public. Access to non-public data is restricted to project partners using access right control of the COLA storage. All public dissemination and marketing data will be shared through the COLA website.
Archiving and Preservation
WP2 will archive dissemination and marketing documents and materials using the ISO 19005-1:2005 standard (Document management — Electronic document file format for long-term preservation - Part 1) with compliance to PDF/A1.
Quality Assurance
There will be two phases of quality control of dissemination and marketing materials. In the first phase WP2 will develop and circulate dissemination and marketing materials

among COLA project partners. In phase 2, they will review these materials and provide feedback WP2. Having this feedback WP2 will finalize and publish dissemination and marketing materials. Quality assurance of WP2 deliverables will be provided by the internal COLA review process and modified as required.
Security, IPR and Ethics
Security
As described in “Data Storage and Data Sharing” section the COLA storage facility has access rights based control that enables/disables access to the documents uploaded and stored in the repository.
Ethics and Privacy
Dissemination and marketing does not raise any extra ethical issue other than describe in D9.1-D9.2. These reports outline how COLA will address these issues. There are no specific privacy requirements for dissemination and marketing data.
Intellectual Property Rights
WP2 will follow the IPR management policy of the COLA project. This policy is included in the Grant Agreement.
3.4 Legal Requirements
There are no specific legal requirements for dissemination and marketing data.

6.1.5 Project Management DMP

Work package: WP1
Person: Gabor Terstyanszky

Data
Existing and New Data
Administrative data: Existing administrative data contains the Project Proposal, the Grant Agreement and the Consortium Agreement. New data includes project partners' interim progress reports (every sixth month), the annual (M18) and final (M30) project reports. It also contains any data relevant to COLA, the Project Management Board (PMB), Technical Task Force (TTF) and Application Task Force (ATF) for example: agendas, minutes, etc. Even if most of the project deliverables are either dissemination or technical reports WP1 coordinates their writing, reviewing and publishing. Financial data: Existing data is the project budget included in the Project Proposal and finalized in the Grant Agreement. Project partners' new financial data consists of researchers' timesheets and receipts of project related costs and expenses. Project partners produce an informal short financial report every sixth month based on research staff's timesheets and project related costs and expenses for the COLA Financial Officer. Having these reports he/she first, checks how project partners spend and use their budget. Secondly, he/she also produces the financial report to be submitted to EC.
Data Format
Administrative data is produced in doc and pdf format while financial data is generated in doc and xls format.
Data Producers and Consumers
Data producers: Local Project Officers collects and forward project partner specific administrative data to the COLA Project Officer who creates the interim project reports. He coordinates annual and final report writing collecting and integrating contributions from Local Project Officers. They also collect timesheets and receipts from researchers and forward these documents to the Local Financial Officers who will compile the local financial reports. Having these the COLA Financial Officer produces the interim, annual and final financial reports. WPs compile project deliverables under the COLA Project Officer's coordination of involving the project partners. Data consumers: The EC Project Officers are the consumers of the administrative and financial data and project deliverables.
Metadata
COLA does not use any metadata to describe administrative and financial project data.
Data Management
Storage and Backup
Administrative data: All project-level administrative data, such as interim, annual and final project reports, project deliverables, agendas and minutes of meetings, etc. is uploaded to the COLA storage - a Pydio repository - available at https://cola.fst.westminster.ac.uk/. Financial data: Project partners collect and store all local financial data and the relevant documents, such as timesheets, receipts, tickets, etc. Several partners scan these documents and store their e-version with other financial data on their storage facilities. Local Project Officers forward project partners' financial information directly to the COLA Financial Officer who uploads and stores this data to the Central Finance storage of the University of Westminster. See backup of financial data in 2.3 Archiving and Preservation sub-section.
Access to Data and Data Sharing
Administrative data: All reports but D3.1-D3, D4.4, D8.1-D8.4 and D9.1-D9.2 are publicly available through the COLA storage. Access to further administrative data such as agendas and minutes of project meetings is restricted to project partners using access right control.

Financial data: Only COLA and Local Financial Officers plus EC Project Officer can access project partners' financial data. Summary of their financial data is included in the annual and final project reports. This summary is publicly available on the COLA storage.
Archiving and Preservation
Administrative data: COLA will use the Pydio solution to archive data. Financial data: Each project partner's Central Finance storage is archived at every 24 hours.
Quality Assurance
Administrative data: COLA set up a quality control procedure to check all project reports. There is a well-defined review process to check the quality of each deliverables. Financial data: COLA and Local Financial Officers check financial data and monitor how project partners spend and use their budget.
Security, IPR and Ethics
Security
Administrative data: The COLA storage facility has access rights based control that enables/disables access to the documents uploaded and stored in the repository. Financial data: It is uploaded and stored on storages of Central Finance of project partners that have sophisticated access control.
Ethics and Privacy
Not relevant to administrative and financial data
Intellectual Property Rights
Not relevant to administrative and financial data
3.4 Legal Requirements
Not relevant to administrative and financial data

6.2 Project-level Data Management Guidelines

Considering the activity-level DMPs WP1 developed Data Management Guidelines for the COLA project. As a result, we collected the data, data management, IPR. legal and security requirements of the activity-level DMPs and compiled three tables (See Table 6.1-6.3). These tables will be considered as Data Management Guidelines and will be recommended both COLA developers and users.

Data aspects of activity-level DMP

	Sarga	TAA	Saker	COLA infrastructure	MiCADO platform	Exploitation	Dissemination & marketing	Project management
data types	• tweets • user data	• tickets • customer surveys • business info • public data	• simulation model • simulation data	• infra data	• TOSCA templates • source codes + binaries • application images	• sustainability data • exploitation data • IPR data	• publications • presentations • leaflets • posters • PR images	• admin data • financial data
data formats	• JSON	• raw data • formatted data • .CSV	• models in .FSM • data in SQL	• none specific format	• YAML descriptions • application artifacts • application images • service images	• .doc • .pdf	• .doc • .jpg • .pdf	• .doc • .pdf • .xls
data producers	• Twitter users	• art organisations • social media companies	• simulation users	• infra sysadmins	• application developers, • MiCADO developers	• project partners	• project partners	• project partners
data consumers	• civil servants	• TAA employee • TAA clients • public	• simulation users	• infra sysadmins	• application users • MiCADO platform users	• use case owners • WP2 partners	• SMEs • public services	• project partners • EC Project Officers
metadata	• XML comments in SQLr	• tags used in RDBM	• none	• none	• TOSCA metadata • Image metadata	• none	• metadata on WP2 data	• none

Table 6.1: Data aspects of activity-level DMPs

Data management aspects of activity-level DMP

	Sarga	TAA	Saker	COLA infrastructure	MiCADO platform	Exploitation	Dissemination & marketing	Project management
data storage	• tweets in SQLr • users data in MySQL	• SQL & noSQL database	• SQL database	• Storpool storage	• GitHub • Docker hub • gdrive • Pydio	• Pydio	• Pydio	• Pydio • financial data -> partner servers
data backup	• daily • weekly	• daily kept for a week	• daily	• CloudSigma solution	• hub + Pydio backups	• Pydio backups	• Pydio backups	• Pydio backups • server backups
data access	• dedicated web interface	• public -> anonymized • non-public -> dedicated clients	• access right based access	• access right based access via COLA infra	• open access	• access right based access to COLA Pydio	• access right based access to COLA Pydio	• access right based access to COLA Pydio + servers
data archivation	• tweets: up to one year • user data: 3-5 years	• one week	• none	• Storpool block storage	• GitHub + Docker hub archivation	• Pydio archivation	• Pydio archivation	• Pydio archivation + server archivation
data Quality Control	• user filters	• automatic + manual QC	• Saker quality system	• Storpool QC	• GitHub + Docker QC	• 3 phase QC	• 2 phase QC	• COLA QC procedure

Table 6.2: Data management aspects of activity-level DMPs

Security, IPR and Ethics aspects of activity-level DMP

	Sarga	TAA	Saker	COLA infrastructure	MiCADO platform	Exploitation	Dissemination & marketing	Project management
data security	• username + password for non-sensitive & public data	• data encryption + multi-factor authentication for sensitive data	• access right based access	• access right based access	• access right based access	• access right based access	• access right based access	• access right based access
ethics + privacy	• no relevant -> no sensitive personal data posted by individual	• no personal data: not relevant • personal data -> EC ethical regulations	• COLA ethics policy (see details in COLA D9.1 + D9.2)	COLA ethics policy (see details in COLA D9.1 + D9.2)	COLA ethics policy (see details in COLA D9.1 + D9.2)	COLA ethics policy (see details in COLA D9.1 + D9.2)	COLA ethics policy (see details in COLA D9.1 + D9.2)	COLA ethics policy (see details in COLA D9.1 + D9.2)
IPRS	• no specific requirements	• non-public data is protected by IPR	• Saker clients own data	• COLA IPR policy (see in the Grant Agreement)	• COLA IPR policy (see in the Grant Agreement)	• COLA IPR policy (see in the Grant Agreement)	• COLA IPR policy (see in the Grant Agreement)	• COLA IPR policy (see in the Grant Agreement)
legal requirements	• data usage + confidentiality agreement must be signed	• data usage + confidentiality agreement must be signed	• data usage + confidentiality agreement must be signed	• EU + Swiss laws	• none	• none	• none	• none

Table 6.3: Security, IPR and Ethics aspects of activity-level DMPs

7 Conclusion

The COLA work packages elaborated activity oriented DMPs considering the diversity and heterogeneity of data to be produced and managed in COLA. These DMPs implement the FAIR principles to support the full-lifecycle data management:

Findability

- DMP defines data formats and gives recommendation on metadata used to describe data needed and produced in COLA use cases and the MiCADO platform. COLA will use Digital Object Identifiers (DOI) to identify data.

Accessibility

- DMP specifies access type of data items and objects produced in COLA as public (or open) and private. Technology specific data, such as MiCADO platform data, will be public. Companies and public services (users of the MiCADO platform) define in the User DMP which data will be public and which data will be private and who can access it.

Interoperability

- COLA will use standard data format and metadata will follow metadata standardization. These details are outlined in the COLA DMP.

Re-use

- COLA public data will be available to third parties free of charge for scientific purposes but restrictions may apply for commercial use in compliance with open access regulations.
- DMP gives recommendations for quality control measures for data produced in the MiCADO platform considering technology and user data.

COLA will use the FAIR principles and follow the full lifecycle data management to allow the best possible dissemination, sharing and usage of COLA data. However, as the COLA consortium incorporates data providers and data users with different expertise and data resources to be managed to create a single Data Management Plan is not a realistic objective. Section 6.2 presents a summary of the activity-level DMPs. This summary will guide developers and users to manage data in the COLA infrastructure.

WP1 will monitor how COLA work packages follow and use the activity-level COLA DMPs. This work package will extend/upgrade these DMPs based on this monitoring and considering new data requirements focusing on FAIR principles.

Annex I Data Management Plan Table

Work package:
Person:

Data in COLA
Existing and New Data
Data Format
Metadata
Data Management
Storage and Backup
Access to Data and Data Sharing
Archiving and Preservation
Quality Assurance
Security, IPR and Ethics
Security
Ethics and Privacy
Intellectual Property Rights
Legal Requirements