

Deliverable 5.11

Platform and Data Management software 3

Date:	June 2018
Authors:	Christos Mitatakis, Maria Pontiki (ATHENA RC)
Contributors	Nikos Larios, Panagiota Koltsida Haris Papageorgiou, Athanasios Balis (ATHENA RC)
Quality assuror(s):	Panagiotis Zervas (Agroknow)
Dissemination level:	PU
Work package	WP5
Version:	1.0
Keywords:	Data Management, Metadata Management, Data Analytics, Software, Platform
Description:	This deliverable is a software deliverable describing the third software release of the CAPSELLA platform and its services



ICT-10-2015 Collective Awareness Platforms for Sustainability and Social Innovation

CAPSELLA (Collective Awareness PlatformS for Environmentally-sound Land management based on data technoLogies and Agrobiodiversity)

Project No. 688813

Project Runtime: January 2016 – June 2018

Copyright © CAPSELLA Consortium 2016-2018

Version History

Version	Date	Description
0.1	01 June 2018	First version
0.2	13 June 2018	Updates wiki pages
1.0	03 July 2018	Final version

Glossary

ABBREVIATION	DEFINITION
ABSA	Aspect Based Sentiment Analysis
API	Application Programming Interface
LDAP	Lightweight Directory Access Protocol
OAI-PMH	The Open Archives Initiative Protocol for Metadata Harvesting
REST	Representational state transfer
SWAGGER API	API interaction in a sandbox user interface
UI	User Interface

Disclaimer

This document contains description of the CAPSELLA project findings, work and products. Certain parts of it might be under partner Intellectual Property Right (IPR) rules so, prior to using its content please contact the consortium head for approval.

In case you believe that this document harms in any way IPR held by you as a person or as a representative of an entity, please do notify us immediately.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of CAPSELLA consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 28 Member States of the Union. It is based on the European Communities and the member states cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors. (<http://europa.eu.int/>)



CAPSELLA is a project partially funded by the European Union

Executive Summary

Deliverable D5.11 “Platform and Data Management software 3” is a software deliverable that presents the third and last public release and deployment of the full set of CAPSELLA platform services.

CAPSELLA platform is the core component of the systems and pilot applications that have been developed during the project’s lifetime. Aiming on supporting them and based on a co-design process from top to bottom, ending to the CAPSELLA platform, a rich set of functionalities and services are offered, including data and metadata storage, access and search, social media data extraction and data analytics (i.e. sentiment analysis on social data sources).

Interoperability with other repositories is of high importance and a set of services has been implemented based on the well-known and used OAI-PMH ¹protocol standard.

¹ OAI-PMH protocol: <https://www.openarchives.org/pmh/>

Table of Contents

Glossary.....	3
Executive Summary.....	5
Introduction.....	7
1. Software.....	8
1.1 Platform Overall Architecture	8
1.2 Authentication Service	9
1.3 Data Management Service	10
1.4 Data Catalogue Portal	10
1.4.1 Accessing API.....	11
1.5 Data Analytics and Management System	12
1.6 OAI-PMH Publisher Service	12
1.7 OAI-PMH Harvester Service	12
2. Software Verification	14
3. Future Steps	15

List of Figures

Figure 1: CAPSELLA Platform Architecture.....	9
Figure 2: CAPSELLA Data Catalogue Portal.....	11

Introduction

The intention of this report is to gather in one place all links related to the CAPSELLA Platform and Data Management Software. The actual descriptions and the working URLs can be retrieved via the links provided.

Following the CAPSELLA work plan, platform services have been ready to a large extent already within the second year of the project, so as to be in the position to serve the technical needs that would arise from the applications developed within the pilots. In the last months of the project the work indeed focused on supporting technically the pilot applications. Through this work the suitability of the platform services developed and deployed could be confirmed, as minor changes and additions have been needed. In fact, platform services and functionalities that have been added, are in all cases a respond to requirements set by the pilot applications.

Furthermore, the scalability of the CAPSELLA platform could be tested, as in the last period a big number of datasets have been generated by the CAPSELLA pilot applications and made available through the data infrastructure. This process was also particularly helpful to expose problems and bugs, fix them, and thereby improve particular services. A related example is the CAPSELLA authentication service, which was extended in order to allow CAPSELLA users to create unique application tokens for each application that uses CAPSELLA platform. Using this service, the users of CAPSELLA pilot apps have a controlled indirect access to CAPSELLA platform.

1. Software

In this third, final release of the software, all services designed and planned have been made available and are already used by the seven pilots (for details about the pilots you may refer to D4.2 and D4.3), while some fine-tunings took in place in parallel. In the next sections a brief description of each available service is provided.

This last release includes bug fixes, performance improvements in various services and additional functionality that had been requested by the pilot applications as they evolved. In particular, the new services that have been developed are:

- **Application token:** a user has the option to generate an application token with long range expiration date. This functionality is useful for applications that want to use the platform (store, get data) without a login process. This functionality can be enabled by the End User Interface when an active user logs in. The produced token may have access to the groups of the user who created it. This was a requirement from the seeds app pilot.
- **Authentication Service change password functionality:** A generic service for registered users to change their passwords. This service is implemented for the Public Food Pilot and especially for the administration application.
- **Publish pilot-generated datasets:** all datasets produced by CAPSELLA pilots and stored to the platform as “public” are automatically published in the Data Catalogue Portal. Specifically, from the Public Food pilot, the Compost Calculator pilot and the Soil Health pilot.

The Data Manager Service of the CAPSELLA platform supports the functionalities of storing, searching, updating and deleting of the datasets and allows the management of datasets with the following data types:

- csv files,
- JSON files,
- image files (jpeg, png),
- shapefiles (geospatial data standard),
- relational data that are stored along with a predefined database schema.

1.1 Platform Overall Architecture

The general architecture of the platform, including all the developed services and displaying how these are connected, is given in the following diagram. No further details are provided in this section, as each service is described in separate sections.

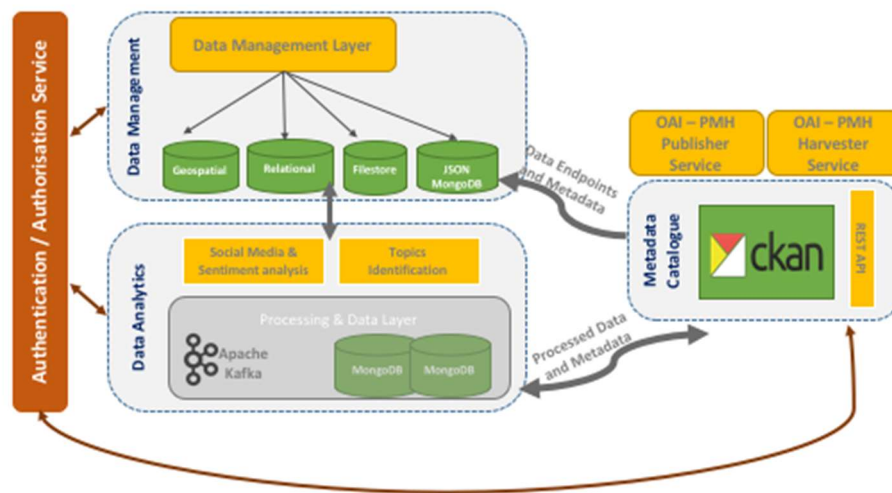


Figure 1: CAPSELLA Platform Architecture

1.2 Authentication Service

The CAPSELLA Authentication service is based on the Lightweight Directory Access Protocol (LDAP)² for registering, authenticating and assigning groups and roles to the users, which are then exploited by the service for accessing and authorisation purposes. The LDAP server is managed internally by the service, which on top of it offers a basic User Interface (UI) for registration, login and information display of the logged in user and a REST API for use by external applications (i.e. pilot applications). Moreover, a user has an option to generate an application token with long range expiration date. This functionality is useful for applications that want to use the platform (store, get data) without a login process. This functionality can be enabled by the End User Interface when an active user logs in. The produced token may have access to the groups of the user who created it. The administration users of Public Food Pilot (city officers) use this service.

The authentication service is available at the following links:

- End User Interface:
https://capsella-services.madgik.di.uoa.gr:8443/capsella_authentication_service/#/login
- Swagger³ Interface:
https://capsella-services.madgik.di.uoa.gr:8443/capsella_authentication_service/swagger-ui.html

Detailed documentation is available at the following URL:

https://ticketing.madgik.di.uoa.gr/redmine/projects/capsella-public/wiki/Capsella_Platform_Authentication_Service

² LDAP: <https://www.ldap.com/getting-started-with-ldap>

³ Swagger API framework: <https://swagger.io/>

1.3 Data Management Service

The Data Management System offers storage, retrieval and management of various data types, that are efficiently stored using the most appropriate mechanisms for each type. It consists of a set of services for each supported data type and the orchestrator service that interacts with them. The applications can only interact with the orchestrator service, named: Data Manager Service. The service offers a REST API that includes all the needed requests for interacting with it. The Swagger endpoint is available at:

<https://capsella-services.madgik.di.uoa.gr:8443/data-manager-service/swagger-ui.html>.

Detailed documentation is available at the following URL:

https://ticketing.madgik.di.uoa.gr/redmine/projects/capsella-public/wiki/Capsella_Platform_Data_Management_System

1.4 Data Catalogue Portal

The Data Catalogue portal is based on the open source CKAN⁴ technology offering browse and discovery facilities for all the available datasets and their resources. Currently, the portal offers a set of open datasets and repositories imported from existing ones covering agriculture, biodiversity, food and other related domains and all CAPSELLA generated public datasets, that are automatically published to the catalogue.

Every dataset is accompanied by a rich set of metadata for facilitating discovery and identification. The available datasets have been grouped into different Organisations, exploiting CKAN's data model. Each organisation can have a number of groups where datasets can belong. The organisation of datasets can be updated at any time in order to reflect pilot's needs and requirements. In addition, organisations and / or datasets can be either public or private and openly accessible or accessible only by authenticated users.

The data catalogue portal is available at the following URL: <http://capsella.madgik.di.uoa.gr/>ⁱ

⁴ CKAN open-source data portal platform: <http://ckan.org/>



Figure 2: CAPSELLA Data Catalogue Portal

1.4.1 Accessing API

In addition to the User Interface (UI), a programmatic interface (API⁵) based on the REST protocol is available for accessing the datasets in a machine-readable format. The API is offered by CKAN and is a powerful, RPC-style API that exposes all CKAN's core features to API clients by exploiting the JSON⁶ standard exchange format.

Detailed information on how to use and exploit the CKAN's API can be found here: <http://docs.ckan.org/en/latest/api/>

⁵ Application Programming Interface (API): https://en.wikipedia.org/wiki/Application_programming_interface

⁶ JSON: <https://en.wikipedia.org/wiki/JSON>

1.5 Data Analytics and Management System

The Data Analytics and Management System performs online Social Media knowledge extraction and offers analytics services (e.g. sentiment analysis) in order to promote candidate solutions and relevant information to the communities and pilot applications involved in the CAPSELLA project, as well as to enhance a given knowledge base of a community in general. The System is a semi-automated pipeline consisting of 5 main parts: Data collection, data ingestion, data management, data analysis and a set of APIs. The CAPSELLA pilots are interesting in the analysis part where topic Detection, polarity Detection, as well as Aspect Based Sentiment Analysis (ABSA) are performed on the data and the raw data along with results of the analysis are provided via dedicated REST APIs.

Due to the nature of this system, there is no default endpoint available, but a dedicated endpoint is becoming available for each client.

Detailed documentation is available at the following URL:

https://ticketing.madgik.di.uoa.gr/redmine/projects/capsella-public/wiki/Capsella_Platform_Data_Analytics_and_Management_System

1.6 OAI-PMH Publisher Service

Aiming on being interoperable with external catalogues and offering CAPSELLA's platform datasets through a standard API and protocol, a service compatible with the most known and adopted standard, the Open Archive Initiative Protocol (OAI-PMH) has been developed, released and deployed on top of the data catalogue.

The OAI-PMH is a protocol developed for harvesting metadata descriptions of records in an archive so that services can be built using metadata from many archives. An implementation of OAI-PMH must support representing metadata in the Dublin Core⁷ schema, but may also support additional representations depending on their nature and origin.

The released service is aligned to the protocol's specification and is available from the following URL: <http://capsella.madgik.di.uoa.gr:8081/oaipmh-publisher?verb=Identify>

1.7 OAI-PMH Harvester Service

The OAI-PMH Harvester service is a REST service which harvests metadata exposed by repositories using the OAI-PMH specification and inserts them into the CAPSELLA metadata catalogue. The harvester service is initiated on demand by inserting the URL of the repository and a periodic harvesting either per repository if set or a global one is done so as to update the catalogue with the latest datasets.

⁷ Dublin Core metadata schema: <http://dublincore.org/>

The harvester service may only be available for accessing to dedicated IPs for security reasons, as it inserts automatically metadata into the CAPSELLA catalogue and this can only be done by administrators, who access it through internal IPs.

2. Software Verification

The released and deployed systems and services are available in the mentioned URLs. In addition, the source code of the developed systems is available on project's GitHub repository at: <https://github.com/CAPSELLA/Platform>

The repository was continuously updated to contain all new developments and updates that consist and are part of the CAPSELLA platform.

As this is a static report and the URLs of the services may change during time, please always visit the CAPSELLA platform documentation wiki to find out the latest URLs:

https://ticketing.madgik.di.uoa.gr/redmine/projects/capsella-public/wiki/Software_URLs.

3. Future Steps

This software report is the third of a series of three deliverables, containing the final version of systems and services. During the last months, new functionalities, bug fixes and improvements have been made available and covered by the appropriate documentation.

The CAPSELLA platform will be kept alive even after the closing of the project, as several pilots have already concrete plans to continue and sustain part of their activities. ATHENA will dedicate human resources to support proper functionality of all pilots. For the time being, this cannot include however further development or addition of new functionalities, but rather any debugging or troubleshooting needed to sustain the aforementioned software services. In the case of interest by new communities, or by the current ones for further developments, ATHENA will discuss together with the CAPSELLA partners how to implement one of the exploitation solutions defined in the Community Engagement and Sustainability plan, or propose an adequate new one.
