H2020-SC5-01-2017

MED-GOLD

Turning climate-related information into added value for traditional **MED**iterranean **G**rape, **OL**ive and **D**urum wheat food systems

Report on the final release of the MED-GOLD ICT platform (Deliverable 1.8)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under gran agreement No 776467.

Document Status Sheet

Title	**Report on the Final release of the MED-GOLD ICT platform (Deliverable 1.8)**	
Brief Description	This report briefly describes the final release of the MED-GOLD ICT platform, documenting both the general technological infrastructure and the content of the associated code repositories.	
WP number	1	Setting the scene: appraising the MEDGOLD sectors, assessing existing climate information and development of a common ICT platform
Lead Beneficiary	LUTECH BEETOBIT	
Contributors		
Creation Date	28/05/2021	
Version Number	1.1	
Version Date	30/05/2021	
Deliverable Due Date	31/05/2021	
Actual Delivery Date	31/05/2021	
Nature of the Deliverable	D	*R - Report* *P - Prototype* *D - Demonstrator* *O - Other*
Dissemination Level/ Audience	PU	*PU - Public* *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*

Revision History Log

Version	Date	Created / Modified by	Pages	Comments
1.0	28-05-2021	BEETOBIT	11	Initial Draft
1.1	30-05-2021	BEETOBIT	15	Added executive summary; applied correct template; various corrections

All partners involved in the production/implementation of the deliverable should comment and report (if needed) in the above table. The above table should support the decisions made for the specific deliverable in order to include the agreement of all involved parties for the final version of the document.

Finally, after the peer review process, the deliverable should be modified according to the comments and the reflections to the comments should be reported in the above table.

Disclaimer

TABLE OF CONTENTS

EXECUTIVE SUMMARY

In this report, we describe the general architecture of the MED-GOLD ICT platform's infrastructure, with its subcomponents, and the content of the associated code GIT repositories. The repositories are the central location in which data and/or codes are stored and managed.

The ICT platform code has been published to several different public repositories, according to the following table:

GIT Repository URL	Content description
https://github.com/medgold-ict-platform/mgd-iac-live	live infrastructure definition files for the main ICT platform infrastructure
https://github.com/medgold-ict-platform/mgd-iac-mods	infrastructure definition modules (user management, security and permission definition, metadata storage, virtual networking) for the main ICT platform infrastructure used by live definition files .
https://github.com/medgold-ict-platform/mgd-cds-downloader	components and scripts used to retrieve climate data from the CDS-C3S
https://github.com/medgold-ict-platform/mgd-processing-compute	scripts and data fixtures for climate indices calculations using Amazon Athena, including data conversion scripts
https://github.com/medgold-ict-platform/mgd-processing-mods	live infrastructure definition files for the indices calculation
https://github.com/medgold-ict-platform/mgd-processing-live	infrastructure definition modules used by the indices calculation live infrastructure files
https://github.com/medgold-ict-platform/mgd-platform-frontend	frontend code for the web data storage application
https://github.com/medgold-ict-platform/mgd-general-api	Python code and OpenAPI documentation for the REST API
https://github.com/medgold-ict-platform/mgd-dashboard-backend	backend code for the dashboard visualisation tool
https://github.com/medgold-ict-platform/mgd-dashboard-frontend	frontend code for the dashboard visualisation tool

1. OBJECTIVES

With this deliverable, the project has contributed to the achievement of the following objectives (DOA, PartB Table1.1):

No.	Objective	Yes
1	To co-design, co-develop, test, and assess the added value of proof-of-concept climate services for olive, grape, and durum wheat	Y

2	To refine, validate, and upscale the three pilot services with the wider European and global user communities for olive, grape, and durum wheat	Y
3	To ensure replicability of MED-GOLD climate services in other crops/climates (e.g., coffee) and to establish links to policy making globally	Y
4	To implement a comprehensive communication and commercialization plan for MED-GOLD climate services to enhance market uptake	Y
5	To build better informed and connected end-user communities for the global olive oil, wine, and pasta food systems and related policy making	Y

2. IMPACT

No.	Expected impact	Yes
1	Providing added-value for the decision-making process addressed by the project, in terms of effectiveness, value creation, optimised opportunities and minimised risk	Y
2	Enhancing the potential for market uptake of climate services demonstrated by addressing the added value	Y
3	Ensuring the replicability of the methodological frameworks for value added climate services in potential end-user markets	Y
4	To implement a comprehensive communication and commercialization plan for MED-GOLD climate services to enhance market uptake	Y
5	To build better informed and connected en d-user communities for the global olive oil, wine, and pasta food systems and related policy making	Y

3. DEFINITIONS

Concepts and terms used in this document and needing a definition are included in the following table:

Concept / Term	Definition
Dashboard	Web-based application that was developed to apply climatic services to the agricultural sectors of MED-GOLD project
Bio-climatic indicator	Indicator describing a specific climate characteristic for the agricultural sector
Elasticsearch	A scalable no-sql DB and sta analytics system
Python	An interpreted high-level general-purpose programming language. It is supported by the Python Software Foundation
Repository	A central location in which a certain amount of data and/or code is stored and managed
Lambda	an AWS service that allows publishing code modules, which can be subsequently executed in response to various types of events, without having to manage the configuration and scaling or servers of containers.
S3 Bucket	a data storage location within S3 (see below)
Terraform	Terraform is an open source IaC tool (see below) which allows to programmatically define and version ICT infrastructure using a declarative language. The infrastructure Terraform can manage includes low-level components such as compute instances, storage, and networking, as well as high-level components such as DNS entries, SaaS features, etc.
Terragrunt	Terragrunt is a thin wrapper that provides extra tools for, working with multiple Terraform modules, and managing remote status.
Fargate	An AWS serverless computation engine, abstracting container orchestration systems such ECS and Kubernetes.
Container	Containers are lightweight, standalone, executable packages of software that includes everything needed to run an application: code, runtime, system tools, system libraries and settings.

4. ACRONYMS

Acronyms used in this document and needing a definition are included in the following table:

Acronym	Definition
MED-GOLD	Turning climate related information into added value for traditional Mediterranean Grape, OLive and Durum wheat food systems
CDS-C3S	Climate Data Store of the Copernicus Climate Change Service
AWS	Amazon Web Services, the Public Cloud platform the MED-GOLD ICT platform is built upon
REST API	Representational State Transfer Application Programming Interface, the machine-to-machine interface of the ICT platform
IaC	Infrastructure as Code: the process of managing and provisioning IT infrastructures (“datacenters”) through machine-readable definition files, rather than physical hardware configuration or interactive configuration tools
VPC	Virtual Private Cloud, the virtual networking subsystem of AWS
S3	Simple Storage Service, an object db-like durable storage on AWS

5. REFERENCES

The following documents, although not part of this document, amplify or clarify its contents. Reference documents are those not applicable and referenced within this document. They are referenced in this document in the form [RD.x]:

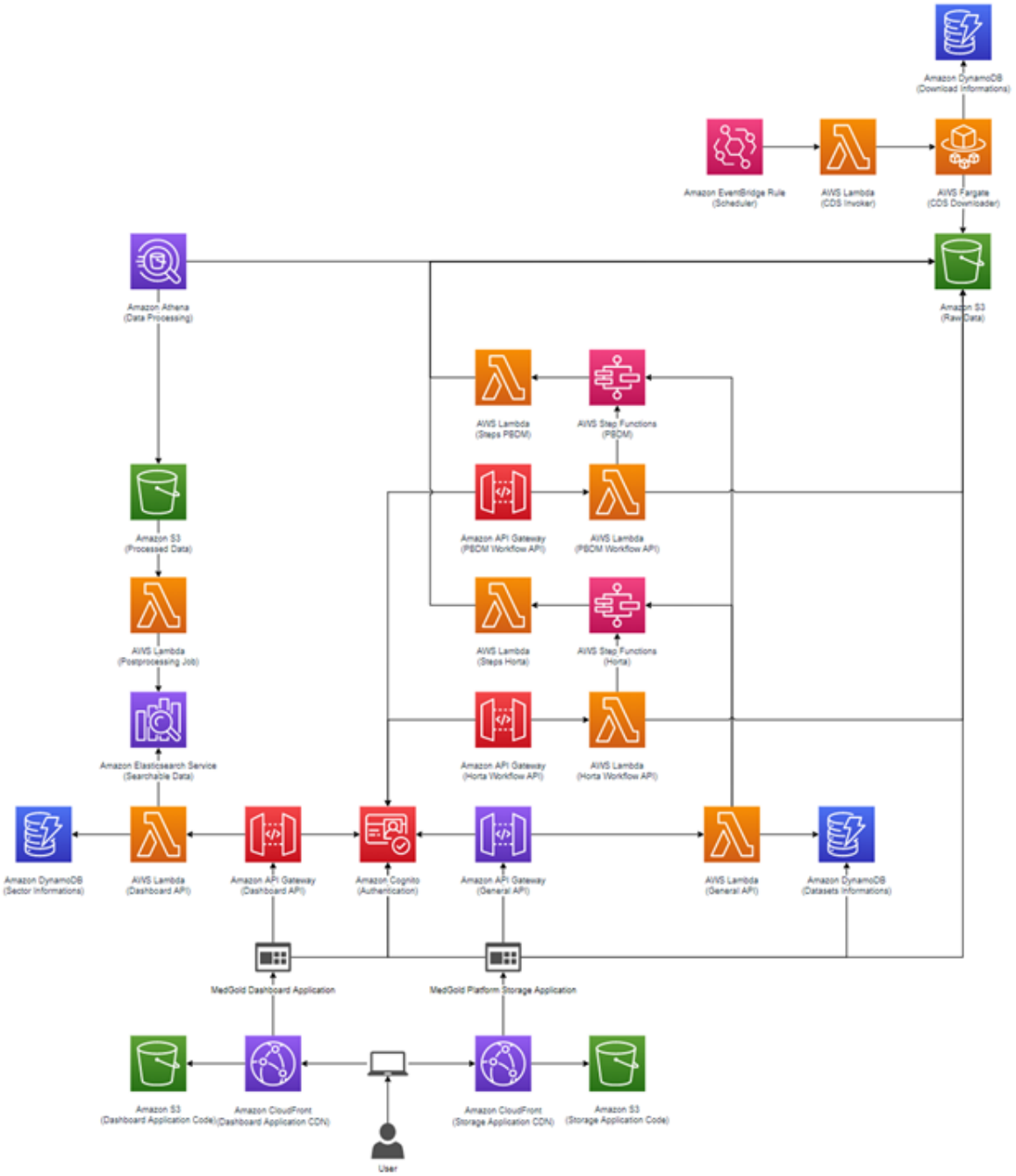
Ref.	Title	Code	Version	Date
[RD.1]	MED-GOLD deliverable 1.5: Deployment of the MED-GOLD ICT platform			2018
[RD.2]	MED-GOLD deliverable 2.8: Report on the Final release of the tool for the olive oil sector			2021
[RD.3]	MED-GOLD deliverable 3.8: Report on the Final release of the tool for the wine sector			2021
[RD.4]	MED-GOLD deliverable 4.8: Report on the Final release of the tool for the durum wheat sector			2021
[RD.5]	MED-GOLD deliverable 2.5:A handy easy-to-use manual for stakeholders and practitioners of the climate service tool. PART I: the olives/olive oil sector			2021
[RD.6]	MED-GOLD deliverable 3.5: A handy easy-to-use manual for stakeholders and practitioners of the climate service tool. PART II: the grape/wine sector			2021

6. ICT PLATFORM ARCHITECTURE AND CODE REPOSITORIES

6. 1 GENERAL ARCHITECTURE AND EXECUTION WORKFLOW

As originally stated in [RD.1], the role of MED-GOLD ICT platform is primarily that of providing a horizontal layer of data management and data processing tools, implementing a reproducible and scalable data pipeline, starting with the import and normalisation of climate data used as an input, supporting the execution of the pilot tools, calculating climate indicators, and serving their outputs both for machine-to-user interactions (mainly via the dashboard visualisation tool) and for machine-to-machine workflows (via the REST API), in a secure, efficient, economically sustainable and scalable way.

Here follows the architectural diagram of MED-GOLD ICT Platform:



The data processing workflow of the ICT platform typically begins on top of the diagram: a scheduled AWS Lambda Function periodically runs a task on Amazon Fargate (on ECS – Elastic Container Service) to download CDS-C3S data to a local storage location: an S3 Bucket; here the data is processed and forwarded as input of the subsequent step of the workflow. Moreover, information and metadata related to each downloaded file is stored on an Amazon DynamoDB table.
The data stored on S3 is then queried by a process running on Amazon Athena, a data analytics serves, based on Apache Hive, which processes properly structured files using a SQL-like syntax and it proceeds

with the computation of various indicators, whose output is stored as CSV files in a new, dedicated S3 Bucket.
When a new file is created, it triggers a Lambda function that produces an optimized object for a full-search engine, indexed within an Elasticsearch cluster: from here, all the data can be queried via the MED-GOLD Dashboard REST API.
The Dashboard API is then used as a backend by MED-GOLD Dashboard, a data visualisation tool which allows end users and stakeholders to query the climate information via a web-based graphical interface, including a geographical map, graphing and data export capabilities.
The users of the ICT Platform can also utilize a Storage Web Application that can be used to upload and share additional datasets; this application is also served by a Generic API which provides additional metadata information over the datasets.
Finally, the ICT platform supports the creation of custom-made workflows utilizing the data stored within the platform to perform further computation, and export their output via scalable REST API.
Two such workflows (PBDM and Horta) have been implemented so far; they're described in more detail in [RD.2] and [RD.4].

6.2 COMMON INFRASTRUCTURE

Most of the modules of the ICT platform share some common components: to keep track of these parts of the infrastructure we used state of art open-source IaC tools Terraform and Terragrunt.
These architectural key components include the major part of the storage layer, Amazon DynamoDB tables, Amazon S3 Buckets and Amazon Elasticsearch Managed Service, and the virtual network layer over which the key services are instantiated.

The IaC code for the common infrastructure, and its documentation can be found at:

- https://github.com/medgold-ict-platform/mgd-iac-live
- https://github.com/medgold-ict-platform/mgd-iac-mods

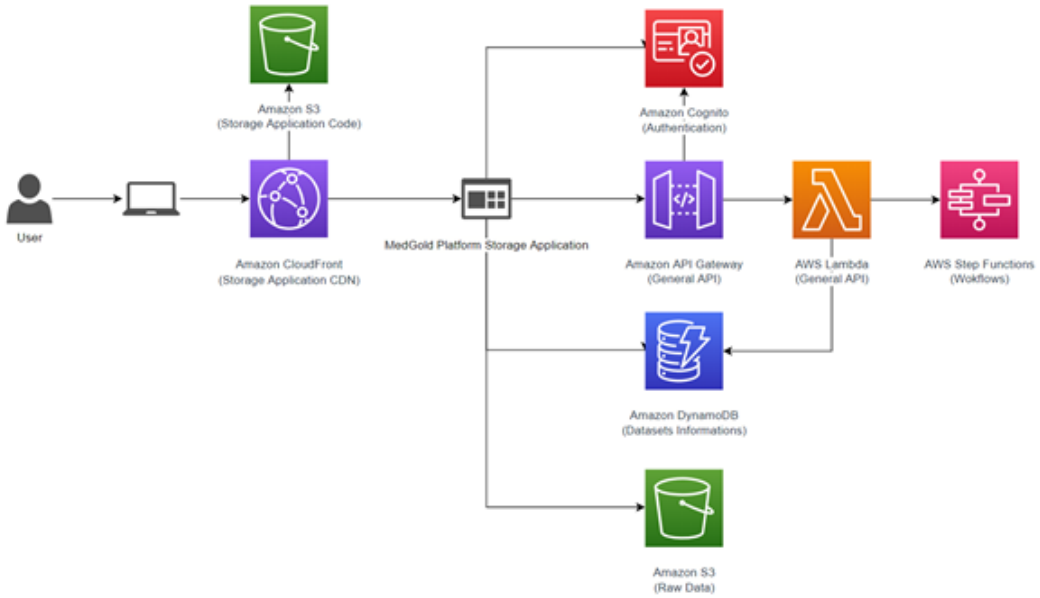
6.3 DATA PROCESSING

The data processing subsystem implements an automated pipeline of IT services that is periodically executed to retrieve climate data, process it and convert it into an easy-to search format, upon which the other platform services are built.
This pipeline is executed monthly and downloads the latest climate data from CDS-C3S and stores it within the storage layer: this is done thanks to an Amazon Event Bridge Schedule Rule configured with a crontab expression which, in turn, executes a Lambda Function.
The download task is typically, time-intensive, which is not a good match for Lambda, which is designed for code with very short execution times (typically up to 5 minutes): the Lambda function is then used to run a task on Amazon Fargate (on ECS) using customised containers. In this case, a suitable Docker image was created, supporting multi-thread processing to improve download and computation speed.
The data, converted to structured data file, is then stored on a dedicated S3 Bucket, and is then accessed c by Amazon Athena: the service is capable of data interpretation from a wide range of files, and it can construct relational tables over them, allowing SQL-like querying capabilities, while exploiting the optimized parallelized query engine.
A set of custom-developed queries on Amazon Athena is used to compute climatic indicators for the three sectors, and its outputs are stored, as CSV files, on another S3 bucket.
The last step of the data processing workflow is dedicated to the parsing of these files, to buildJSON objects that will be uploaded on a dedicated Elasticsearch cluster, which in turns allow fast, scalable searching capabilities over large-scale datasets.

The code and its documentation for the data processing subsystem can be found at:

- https://github.com/medgold-ict-platform/mgd-cds-downloader
- https://github.com/medgold-ict-platform/mgd-processing-compute
- https://github.com/medgold-ict-platform/mgd-processing-mods
- https://github.com/medgold-ict-platform/mgd-processing-live

6.4 STORAGE / API TESTBED WEB APPLICATION



The storage web application provides a sharing space between MED-GOLD users: it's designed to host and share a large set of data files.Moreover, it allows to store and retrieve metadata for the shared datasets (e.g.: descriptions, variables, etc.). The datasets can then be explored through a file system-like view.

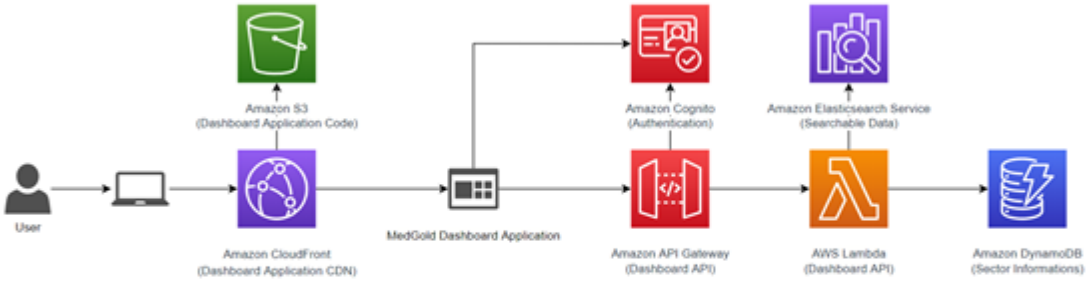
The application has also been integrated with an Open API 3 renderer, which allows users to browse the documentation of MED-GOLD REST APIs and to execute them on their browser, without needing any other programming tool.

The storage web application is accessible on: https://platform.med-gold.eu

Source code for the storage / API testbed web application can be found at:

- https://github.com/medgold-ict-platform/mgd-platform-frontend
- https://github.com/medgold-ict-platform/mgd-general-api

6.5 DASHBOARD WEB APPLICATION



The MED-GOLD dashboard is a data visualisation tool composed of two major components: the REST API and the web application itself.
The API provides a set of endpoints which are called by the frontend to retrieve data based on several parameters: temporal scope, indicator type, year, region, etc. The code running these APIs has been designed to exploit all the benefits of the background Elasticsearch cluster, so that the data can be retrieved efficiently and rapidly.
The web application, which uses the API as a backend, as described, provides a map over which data is displayed, using overlays.

The Dashboard Web Application has three main sections: Historical Climate (for past observations), Seasonal Forecast (for short-term forecasts) and Long Term Projection. For each one, the users can query the available variables and indicators, using its own set of parameters to adjust the retrieval: e.g.: on Historical Climate users can query by year and month.

Dashboard features are described in detail in [RD.5] and [RD.6].

The dashboard can be accessed at: https://dashboard.med-gold.eu

Source code for the dashboard can be found at:

- https://github.com/medgold-ict-platform/mgd-dashboard-backend
- https://github.com/medgold-ict-platform/mgd-dashboard-frontend

END OF DOCUMENT