
# D4.13 ChEESE code repository

**Version 1.2**

## Document Information

| | |
|---|---|
| Contract Number | 823844 |
| Project Website | https://cheese-coe.eu/ |
| Contractual Deadline | 31/03/2022 |
| Dissemination Level | Confidential |
| Nature | Report and associated Web Repository |
| Author | Tomaso Esposti-Ongaro (INGV) |
| Contributors | Stefano Lorito (INGV) |
| Reviewers | Josep de la Puente (BSC), Jacopo Selva (INGV) |

*The ChEESE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No 823844*

## Change Log

| Version | Description of Change |
|---|---|
| 1.0 | Initial draft |
| 1.1 | Version revised by all authors |
| 1.2 | Includes reviewer's comments |

# Table of Contents

## Introduction

This document reports about the activity carried out to improve the accessibility to computing applications developed in ChEESE. In particular:

1. the creation of a public repository to store software, workflows and simulation toolkits, and their documentation;
2. the effort to move towards future full integration of ChEESE products into the EPOS digital research infrastructure.

The two actions will enable researchers and IUB members to be more productive, leading to scientific excellence and economic and social benefit.

To implement the code repository, we had to consider the following constraints:

- Most of the ChEESE flagship codes already have a public repository (on GitHub/GitLab), in some cases they are accessible through a web portal, which is a consolidated reference for a scientific community.
- Some of the codes are not open-source. Some of the developments might be not yet publicly available.
- The repository should be flexible enough, and sustainable beyond the duration of the ChEESE project.
- It should satisfy FAIR (Findable, Accessible, Interoperable, Reusable) principles.

To integrate the ChEESE catalog in the future as a EPOS service, we had to define a Metadata scheme for software and inspect interoperability with the EPOS catalog. As a reference, we have started from EPOS VO-TCS software metadata model. The integration into EPOS will be the target for future developments.

## EPOS VO-TCS CIV Software metadata model

As a general scheme, we used the Software metadata model adopted by one of the Volcano Observatories EPOS TCS services, i.e. CIV, the Computational Infrastructure for Volcanology catalog (https://civ.pi.ingv.it). The CIV portal satisfies the minimal requirements to be interoperable with EPOS by providing a web service and exposing metadata.
The semantics of CIV metadata are described in Table 1.

| Element name | Description |
|---|---|
| ***Identification*** | All elements needed to identify the model (and its use in the framework of the ICS). |
| ***DataIN*** | Descriptors of the Input Data (if needed).<br>Multiple entries should be admitted. Includes configuration files. |
| ***DataOUT*** | Descriptors of the Output Data<br>Multiple entries should be admitted. |
| ***Categories*** | Solid Earth Sciences<br>Others? |
| ***Domains*** | Descriptor of the typical physical domain of application.<br>(Text String) Multiple choice among:<br>1. Magma Chamber<br>2. Volcanic Conduit<br>3. Volcanic Edifice<br>4. Topography |

| | |
|---|---|
| | 5. Atmosphere<br>6. Magmatic Fluids |
| ***Disciplines*** | Descriptor of the physical approach.<br>(Text String) Multiple choice among:<br>1. Fluid mechanics<br>2. Thermodynamics<br>3. Elastodynamics<br>4. Statistics |
| ***Information*** | Available information and documentation for the users. |
| ***Documentation*** | Links to available documentation |
| ***References*** | List of references |
| ***Copyright*** | Descriptors/disclaimers of intellectual property rights and ownership. |

Table 1. Child elements in EPOS VO-TCS CIV software metadata

The sub-child elements of VO-TCS CIV metadata and their format are described in Table 2.

| Element name | Sub Child name | Description |
|---|---|---|
| ***Identification*** | ***DOI*** | (String) Digital Object Identifier (if available) |
| | ***Name*** | (String) Model name or acronym |
| | ***Version*** | (String) Release number |
| | ***LastUpdate*** | (Date format) Date of last release |
| | ***Location*** | (String) Link to download or execution page |
| | ***isInteractive*** | (Boolean). Can be run online |
| | ***isPublished*** | (Boolean). Published on the web portal |
| ***DataIN/DataOUT*** | ***Type*** | (Text) Description of the datum |
| | ***Format*** | (String) Standard format type (if available) |
| | ***Extension*** | (String) File extension |
| ***Information*** | ***Language*** | (String) Programming Language |
| | ***Description*** | (Text) Model Description |
| | ***Algorithm*** | (Text) Algorithm Description |
| | ***Implementation*** | (Text) Implementation Description |
| ***Documentation*** | ***Web Page*** | (String) Link to web page (if available) |
| | ***UserGuide*** | (String) Link to User Guide (if available) |
| | ***ReferenceGuide*** | (String) Link to Reference Guide (if available) |

| | | |
|---|---|---|
| | ***Gallery*** | (String) Link to Image and Video gallery (if available) |
| ***Copyright*** | ***Author*** | (String) Name(s) of the author(s) |
| | ***Author ID Type*** | Multiple choice:<br>1. ORCID<br>2. ResearcherID<br>3. Scopus |
| | ***Author ID*** | (String) Author ID |
| | ***Affiliation*** | (String) Institution |
| | ***License*** | (String) Text or Link to license page |
| | ***Contacts*** | (String) contact email |

Table 2. Sub-child elements of EPOS VO-TCS software metadata

## Zenodo ChEESE-CoE Software repository

Zenodo (www.zenodo.org) is a repository for research data hosted at CERN, allowing users to share their data according to the EU policies and make them citable by their own DOI. Zenodo is open to all disciplines of science, and users can deposit any kinds of data, such as spreadsheets, figures, reports, posters, presentations, and software for free. The submission mask is very user-friendly and offers many options of customizing metadata.

Most of the metadata describing the ChEESE software is already implemented in Zenodo, so we decided to simplify the EPOS VO-TCS software descriptors to use the Zenodo repository. In particular, we consider the DataIn, DataOut, Categories, Domain, Disciplines elements of Table 2 as arbitrary, also because not the entire toolkit developed in ChEESE can have such specifications. The

Table 3 reports the Child/subChild elements describing the ChEESE software.

By using Zenodo, the requirements of FAIRness and long-term sustainability will be guaranteed (https://about.zenodo.org/principles/).

The ChEESE software repository can be browsed at:

https://zenodo.org/communities/cheese-coe/search?page=1&size=20&q=&type=software

| Child Element name | Sub-child name | Description |
|---|---|---|
| ***Basic Information*** | ***DOI*** | Digital Object Identifier |
| | ***Title*** | Model name or acronym |
| | ***Version*** | Release number |
| | ***Publication Date*** | Date of last release |
| | ***Repository*** | Link to download page |
| ***Information*** | ***Programming Language*** | Programming Language |
| | ***Libraries*** | Required ancillary software |
| | ***Requirements*** | Hardware requirements (e.g. minimum RAM, CPU/GPU architecture, storage) |
| | ***Description*** | Physical Model Description |
| | ***Algorithm*** | Mathematical Algorithm Description |

| | | |
|---|---|---|
| | ***Implementation*** | Description of the Software engineering (including parallelization model) |
| ***Documentation*** | ***Web Page*** | Link to web page (if available) |
| | ***Related/alternate identifiers*** | Publication(s) of reference |
| | ***UserGuide*** | Link to User Guide (if available) |
| | ***ReferenceGuide*** | Link to Reference Guide (if available) |
| | ***Gallery*** | Link to Image and Video gallery (if available) |
| ***References*** | | Related references |
| ***Copyright*** | ***Authorship*** | Name(s) of the author(s) |
| | ***Author ID*** | Author ID |
| | ***Affiliation*** | Institution |
| | ***Contacts*** | Contact email |
| | ***License*** | Text or Link to license page |

Table 3. Metadata scheme for ChEESE software. In **red**, the "mandatory" elements requested in Zenodo are highlighted. In **orange**, the "recommended" ones. All sub-child elements in **black** can be inserted in one of the red/orange categories.

## Future integration in EPOS: the EPOS D-CAT Metadata catalog

ChEESE candidates as a future provider of assets (software and workflows) to the EPOS community and it aims at integrating its code and toolkit repository in the EPOS catalog.

A major challenge in EPOS is the integration of multi-disciplinary, multi-organisational, distributed resources and community assets into a single overarching Research Infrastructure - the EPOS Integrated Core Services (ICS). ICS aggregate and harmonize descriptions of datasets, data products, software and services from different domain-specific services - the Thematic Core Services (TCS). TCS adopt heterogeneous formats, vocabularies, protocols and standards to represent and make their resources available.

One of the two main ingredients used by EPOS to achieve interoperability are metadata, that describes services and assets provided by the Thematic Communities.

The exchange of metadata between ICS and TCS is crucial to achieve integration and interoperability in EPOS. ChEESE Zenodo repository objects might be directly integrated in EPOS, however the optimal approach would require the implementation of a web-service and the storage of the ChEESE catalog in a specific format.

In order to capture, organize and harmonize information from different sources and to enable semantic interoperability, a data model has been developed and adopted, namely EPOS-DCAT-AP. It extends and builds on an established W3C standard - the Data Catalog Vocabulary (DCAT). EPOS-DCAT-AP is represented in RDF/turtle, its latest version is available on GitHub - it includes a UML diagram, ontology definition, examples and more details. The actual EPOS catalog uses as its format CERIF: Common European Research Information Format, an EU Recommendation to Member States for research information. CERIF is maintained by EuroCRIS and used widely (also outside Europe) and is embedded in products from Elsevier and Thomson-Reuters as well as being used in OpenAIRE. The metadata catalog is populated from the heterogeneous metadata formats used within the TCS via EPOS-DCAT-AP in an ingestion pipeline that includes dynamic update by specified TCS managers within a secure environment. The EPOS-DCAT-AP specifications already include concepts to

describe software, in particular Software Applications and Software Source Code, which are suited to describe HPC software (not run online). These will be adopted in future implementations of the ChEESE catalog.
The full transition from a repository to a service with the potential to be integrated in the EPOS delivery framework is being achieved also thanks to the momentum offered by the newly funded HE Geo-INQUIRE project, in which several ChEESE partners will offer Virtual and/or Trans National Access to softwares and workflows. Within Geo-INQUIRE, also the cross-TCS service integration within EPOS, but also towards a unified service delivery across different ERICs (e.g. EMSO), will be sought. The ChEESE-CoE repository will provide blue print for this transformation as far as software and workflow provisions as a service are concerned.

## Summary

The ChEESE-CoE community repository of flagship codes, workflow codes, data and products is now accessible at:
https://zenodo.org/communities/cheese-coe

Future integration of ChEESE products (including software) in EPOS will be possible through the development of a dedicated web service allowing the ingestion of metadata into the EPOS Integrated Core Service.