

ChEESE

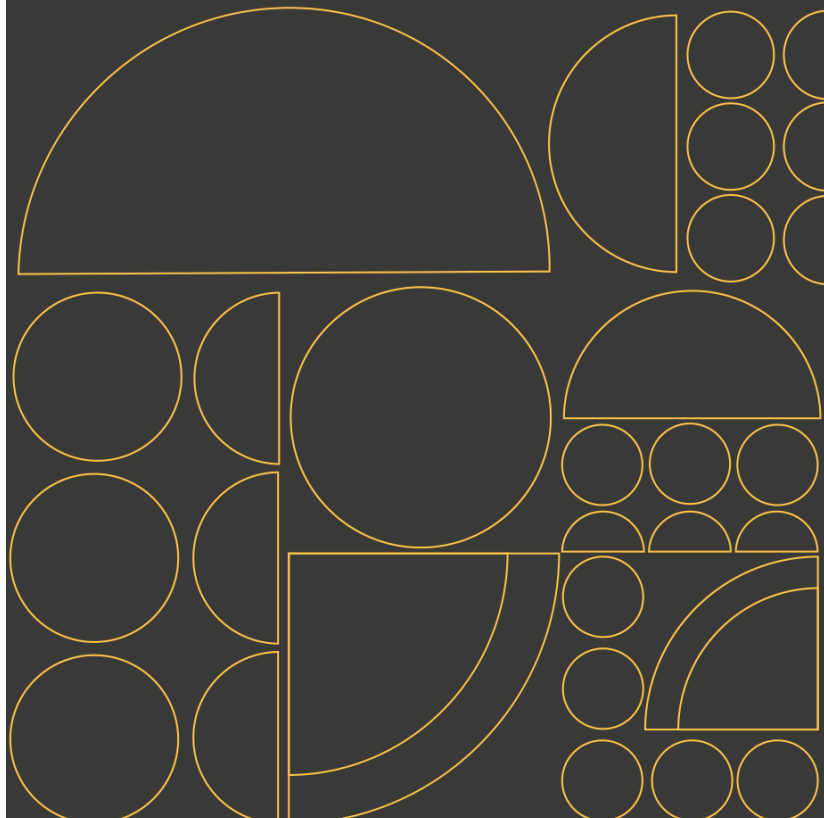
ChEESE centralised repository: proposal discussion

Second General Assembly
Reykjavík, 17-19 September 2024

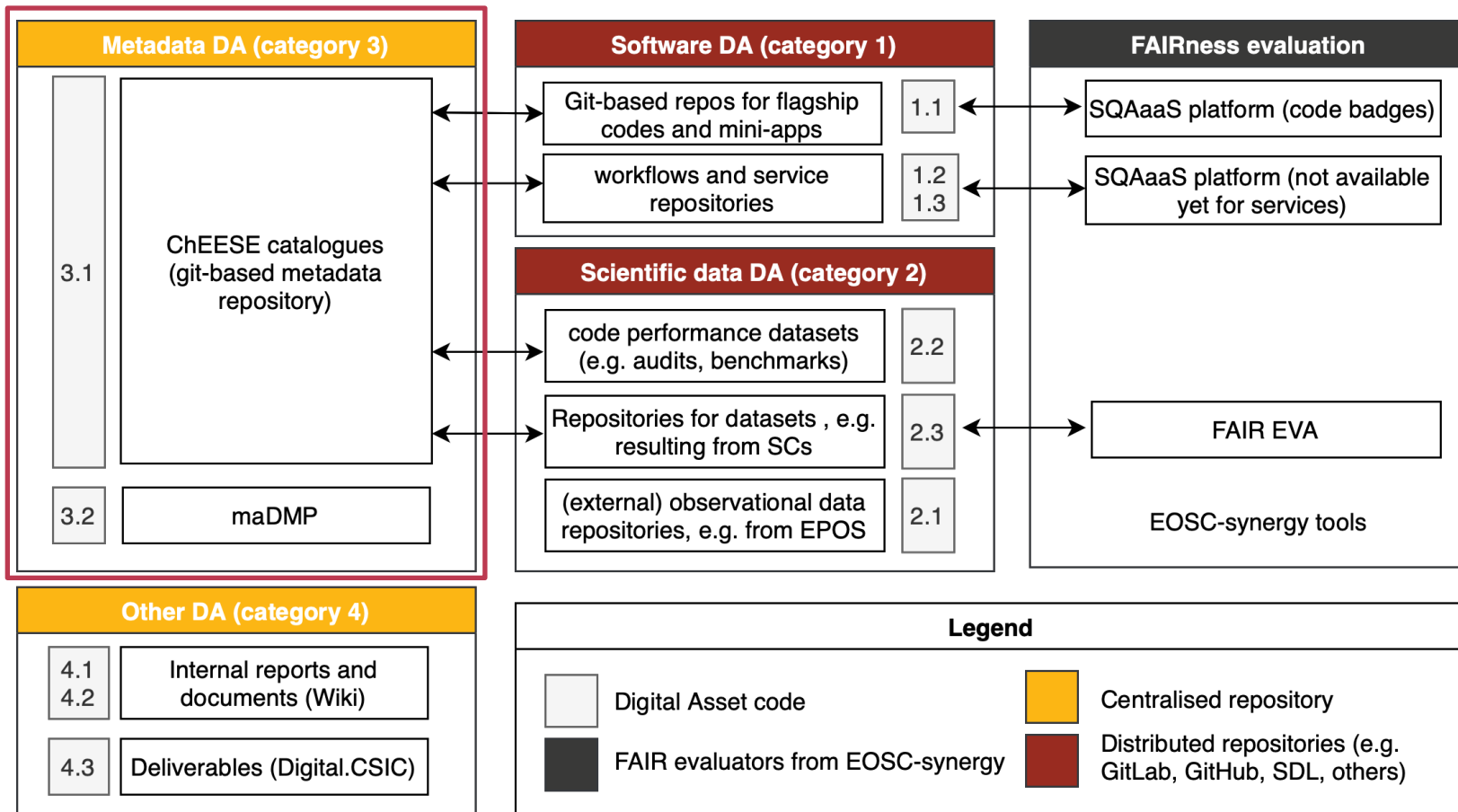
Arnau Folch
Geociencias Barcelona (GEO3BCN-CSIC)



Project funded by EuroHPC under the grant agreement No 101093038.



Types of repositories



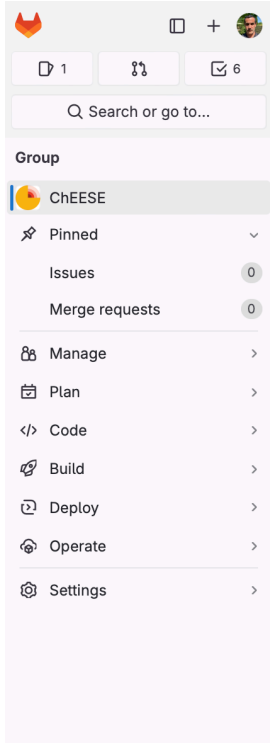
Purpose and objectives of the repo

1. To have a **single entry point** to access and to catalogue ALL the ChEESE Digital Assets
 - Codes and mini-apps
 - Performance tests and benchmarks
 - CI/CD runners (will depend on the CASTIEL-2 system readiness)
 - Workflows and services
 - Datasets from Pilot Demonstrators (PDs) and Simulation Cases (SCs)
 - Other (e.g. trainings, etc)
2. To gather basic **documentation** on the DAs above
3. To annotate the DAs with an **homogeneous metadata model** (if possible)
4. To comply with the **DMP** requirements and to provide a basic **ma-DMP**



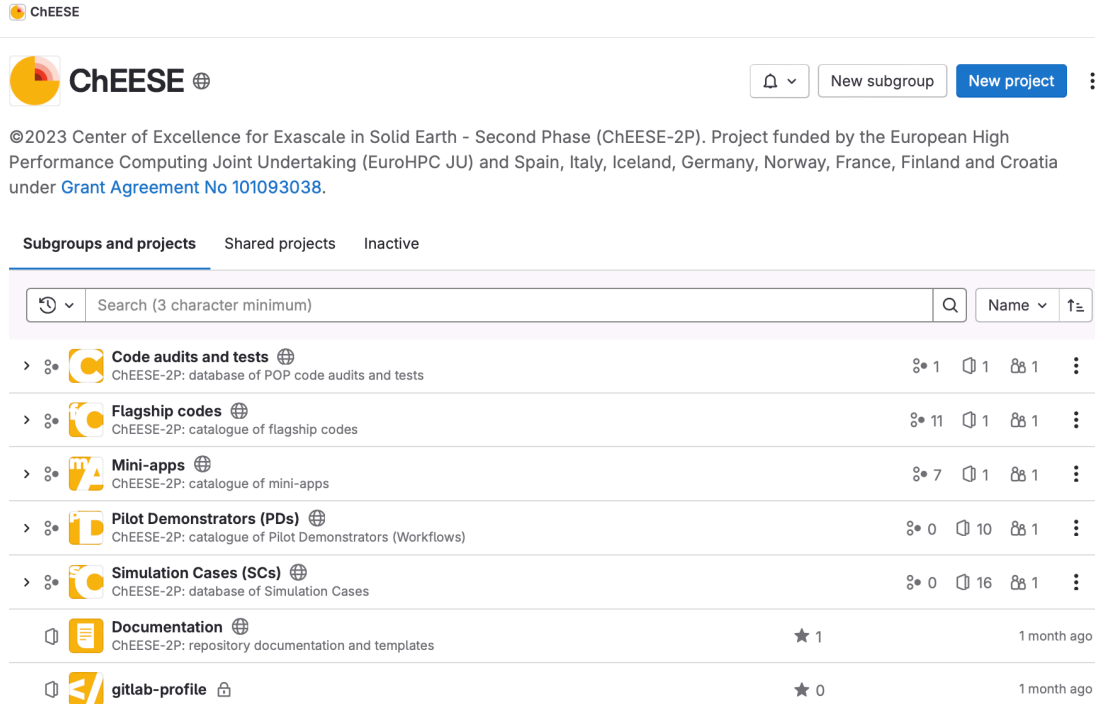
git-based metadata catalogue

- A proposal (incomplete and to be discussed) is available at: <https://gitlab.com/cheese5311126>



GitLab sidebar navigation menu:

- Group
 - ChEESE
 - Pinned
 - Issues
 - Merge requests
 - Manage
 - Plan
 - Code
 - Build
 - Deploy
 - Operate
 - Settings



ChEESE GitLab group page:

©2023 Center of Excellence for Exascale in Solid Earth - Second Phase (ChEESE-2P). Project funded by the European High Performance Computing Joint Undertaking (EuroHPC JU) and Spain, Italy, Iceland, Germany, Norway, France, Finland and Croatia under [Grant Agreement No 101093038](#).

Subgroups and projects Shared projects Inactive

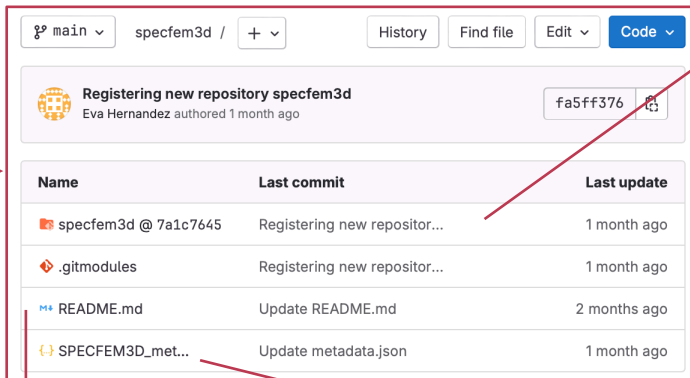
Search (3 character minimum)	Name	↑
> Code audits and tests	ChEESE-2P: database of POP code audits and tests	1 1 1
> Flagship codes	ChEESE-2P: catalogue of flagship codes	11 1 1
> Mini-apps	ChEESE-2P: catalogue of mini-apps	7 1 1
> Pilot Demonstrators (PDs)	ChEESE-2P: catalogue of Pilot Demonstrators (Workflows)	0 10 1
> Simulation Cases (SCs)	ChEESE-2P: database of Simulation Cases	0 16 1
Documentation	ChEESE-2P: repository documentation and templates	1 1 month ago
gitlab-profile		0 1 month ago

Collection of
repositories
(each belonging to
different users)

git-based catalogue: codes

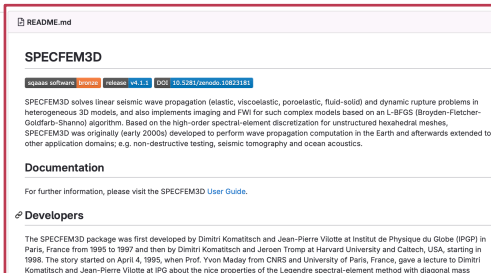
- >  **Code audits and tests** 
CHEESE-2P: database of POP code audits and tests
- >  **Flagship codes** 
CHEESE-2P: catalogue of flagship codes
- >  **Mini-apps** 
CHEESE-2P: catalogue of mini-apps
- >  **Pilot Demonstrators (PDs)** 
CHEESE-2P: catalogue of Pilot Demonstrators (Workflows)
- >  **Simulation Cases (SCs)** 
CHEESE-2P: database of Simulation Cases
- >  **Documentation** 
CHEESE-2P: repository documentation and templates
- >  **gitlab-profile** 

- >  **01-SeisSol** 
- >  **+2-SPECFEM** 
 -  **S SPECFEM1D** 
 -  **S SPECFEM2D** 
 -  **S SPECFEM3D** 
 -  **S SPECFEM3D_globe** 
- >  **03-Tandem** 
- >  **04-xSHELLS** 
- >  **05-HySEA** 
- >  **0 FALL3D** 
- >  **07-OpenPDAC** 
- >  **08-LaMEM** 
- >  **09-pTatin3D** 
- >  **10-ELMER_ICE** 
- >  **11-ExaHyPE** 
- >  **gitlab-profile** 



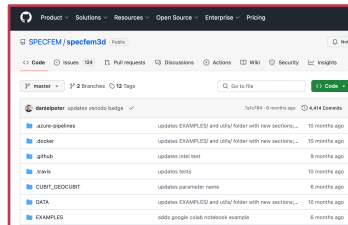
The screenshot shows the GitHub interface for the 'specfem3d' repository. The repository is owned by 'Eva Hernandez' and was created 1 month ago. The commit hash is 'fa5ff376'. The repository contains a README.md file and a SPECFEM3D_metadata.json file. The commit history shows that the README.md file was updated 2 months ago and the metadata file was updated 1 month ago.

Some basic
documentation
(developers,
licence, etc)



The screenshot shows the README.md file for the SPECFEM3D repository. It describes the software as a tool for solving linear seismic wave propagation problems in heterogeneous 3D models. It mentions that the software was originally developed in 2004 and is now maintained by the Institut de Physique du Globe (IPG) in Paris. The README also includes a section for developers and a link to the user guide.

git submodule
pointing to the native
repository



The screenshot shows the git submodule configuration page for the 'specfem3d' repository. It lists the submodules and their commit hashes. The submodules are 'specfem3d' (commit 'fa5ff376') and 'specfem3d_globe' (commit 'fa5ff376').

JSON file with basic metadata
(to comply with ma-DMP)

```
{
  "contact": {
    "name": "Eva Hernandez",
    "email": "ehernandez@geo3d.cnsc.es",
    "contact_id": {
      "identifier": "http://orcid.org/0000-0000-0000-0000",
      "type": "orcid"
    }
  },
  "contributor": {
    {
      "name": "Arnaud Folch",
      "email": "afolch@geo3d.cnsc.es",
      "contributor_id": {
        "identifier": "http://orcid.org/0000-0002-0008-0000",
        "type": "orcid"
      },
      "role": [
        "ProjectLeader"
      ]
    },
    {
      "name": "Heriberto Pascual",
      "email": "hpascual@geo3d.cnsc.es",
      "contributor_id": {
        "identifier": "https://orcid.org/0000-0002-4929-7875",
        "type": "orcid"
      },
      "role": [
        "Developer"
      ]
    },
    {
      "name": "Leonardo Ringari",
      "email": "lringari@geo3d.cnsc.es",
      "contributor_id": {
        "identifier": "https://orcid.org/0000-0002-4929-7875"
      }
    }
  ]
}
```

Example for FALL3D
using RDA-DMP
common standard
(this may change)

Catalogue summary

Name	SG4aaS	DOI	Version	License
SeisSol	 SeisSol software 	DOI: 10.5281/zenodo.1140512	release v1.14	License: BSD 3-Clause
SPECFEM	 SPECFEM software 	DOI: 10.5281/zenodo.1063191	release v4.1.1	License: GPL v3
Tandem	 Tandem software 	Reference	release v1.1.0	License: BSD 3-Clause
xSHELLS	 xSHELLS software 	Reference		License: GPL v3
HySEA	 HySEA software 	DOI: 10.5281/zenodo.1063465	release v1.1.1	License: GPL v3
FALL3D	 FALL3D software 	DOI: 10.5281/zenodo.1063191	release v8.3	License: GPL v3
OpenPDAC	 OpenPDAC software 	DOI: 10.5281/zenodo.1071203	release v1.0.2	License: GPL v3
LaMEM	 LaMEM software 	DOI: 10.5281/zenodo.1071203	release v2.1.4	License: GPL v3
pTatin3D	 pTatin3D software 	Reference		License: GPL v3
ELMER_ICE	 ELMER_ICE software 	DOI: 10.5281/zenodo.1071203	release v0.5	License: GPL v3
ExaHyPE	ExaHyPE software	DOI: 10.5281/zenodo.1071203		License: BSD 3-Clause

git-based catalogue: mini-apps

- >  **Code audits and tests** 
CHEESE-2P: database of POP code audits and tests
- >  **Flagship codes** 
CHEESE-2P: catalogue of flagship codes
- >  **Mini-apps** 
CHEESE-2P: catalogue of mini-apps
- >  **Pilot Demonstrators (PDs)** 
CHEESE-2P: database of Pilot Demonstrators (Workflows)
- >  **Simulation Cases (SCs)** 
CHEESE-2P: database of Simulation Cases
- >  **Documentation** 
CHEESE-2P: repository documentation and templates
- >  **gitlab-profile** 

- >  **Mini-FALL3D** 
 - >  **Mini-FALL3D CPU** 
 - >  **Mini-FALL3D GPU** 
- >  **Mini-HySEA** 
- >  **Mini-pTatin** 
- >  **Mini-SeisSol** 
- >  **Mini-SPECFEM** 
- >  **Mini-Tandem** 
- >  **Mini-XSHELLS** 
- >  **gitlab-profile** 

Some basic
documentation
(developers,
licence, etc)

Registering new repository mini-fall3d-cpu		
Eva Hernandez authored 1 month ago		
12d34b64		
Name	Last commit	Last update
mini-fall3d-cpu @ b8653dc8	Registering new repository mini-fall...	1 month ago
.gitmodules	Registering new repository mini-fall...	1 month ago
Mini-FALL3D-CPU_metad...	Update metadata.json	1 month ago
README.md	Update README.md	2 months ago

git submodule
pointing to the native
repository

Mini-FALL3D		
main	mini-fall3d	
History	Find file	Edit Code
Merge branch 'hotfix/source_uniform' into 'master'		
ARNAU FOLCH authored 1 month ago		
12d34b64		
Name	Last commit	Last update
Example	added other task	1 year ago
Sources	Fixed bug defining sources in the u...	2 months ago
md	First commit	1 year ago
gitignore	First commit	1 year ago
AUTHORS.md	First commit	1 year ago
CHANGELOG.md	First commit	1 year ago
LICENSE	First commit	1 year ago

JSON file with basic metadata
(to comply with ma-DMP)

! Catalogue summary

Name	SQAaaS	DOI	Version	License
Mini-FALL3D				
Mini-HySEA				
Mini-pTatin3D				
Mini-Seisol				
Mini-SPECFEM				
Mini-Tandem				
Mini-xSHELLS				

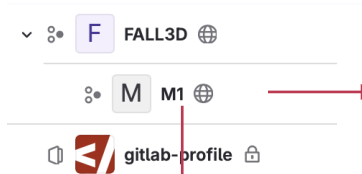
README.md
Mini-FALL3D CPU
A FALL3D mini-app was developed for both CPU & GPU using OpenACC directives in the latter case. This mini-app only solves the FALL3D task assuming a fixed problem, i.e., helicoidal velocity field. This FALL3D task involves fully solving the main computational kernel, i.e., solves a set of advection-diffusion-sedimentation (ADS) equations on a structured grid using a second order finite volume explicit scheme (Euler or RK 4th in time). As stated, the ADS kernel is fully computed through GPU cards, i.e., 100%
Documentation
For further information, please visit the FALL3D User Guide .
Developers
FALL3D is developed and maintained by the Consejo Superior de Investigaciones Científicas (CSIC) and the Istituto Nazionale di Geofisica e Vulcanologia (INGV)
- Arnaud Folch - Antonio Costa - Giovanni Macedonio - Leonardo Mingari
License and copyright

```
{
  "contact": {
    "name": "Eva Hernández",
    "email": "ehernandez@geo3bcn.csic.es",
    "contact_id": {
      "identifier": "https://orcid.org/0000-0000-0000-0000",
      "type": "orcid"
    }
  },
  "contributors": [
    {
      "name": "Arnaud Folch",
      "email": "afolch@geo3bcn.csic.es",
      "contributor_id": {
        "identifier": "https://orcid.org/0000-0002-0008-0000",
        "type": "orcid"
      },
      "role": [
        "ProjectLeader"
      ]
    },
    {
      "name": "Heriberto Pascual",
      "email": "hpascual@geo3bcn.csic.es",
      "contributor_id": {
        "identifier": "https://orcid.org/0000-0002-4929-7875",
        "type": "orcid"
      },
      "role": [
        "Developer"
      ]
    },
    {
      "name": "Leonardo Mingari",
      "email": "lmingari@geo3bcn.csic.es",
      "contributor_id": {
        "identifier": "https://orcid.org/0000-0002-4929-7875",
        "type": "orcid"
      }
    }
  ]
}
```

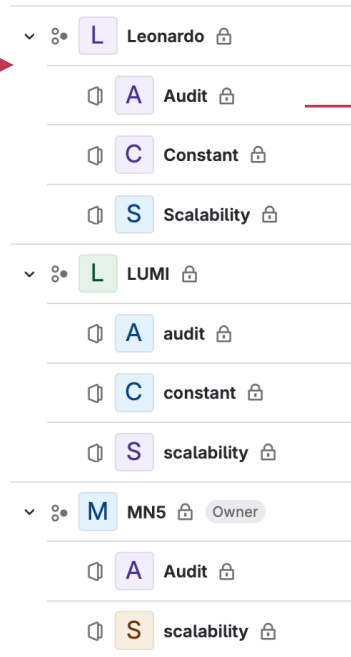
Example for FALL3D
using RDA-DMP
common standard
(this may change)

git-based catalogue: performance tests and audits

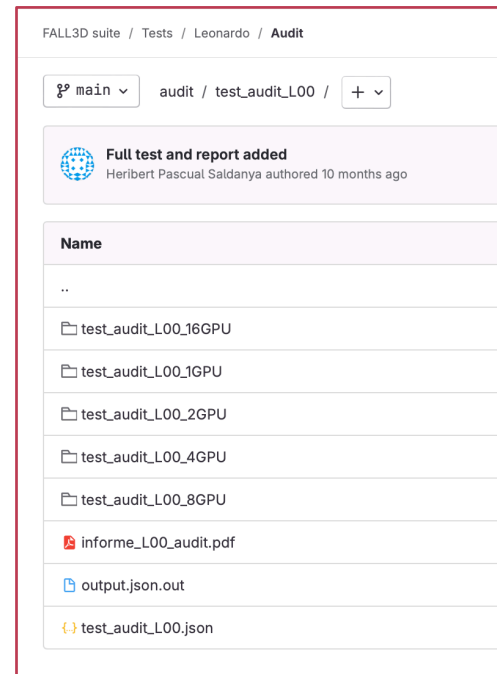
- >  **Code audits and tests** 
ChEESE-2P: database of POP code audits and tests
- >  **Flagship codes** 
ChEESE-2P: catalogue of flagship codes
- >  **Mini-apps** 
ChEESE-2P: catalogue of mini-apps
- >  **Pilot Demonstrators (PDs)** 
ChEESE-2P: catalogue of Pilot Demonstrators (Workflows)
- >  **Simulation Cases (SCs)** 
ChEESE-2P: database of Simulation Cases
-  **Documentation** 
ChEESE-2P: repository documentation and templates
-  **gitlab-profile** 



code baseline



git submodule
pointing to the native
repository



Screenshot of a GitLab commit page showing the commit history and file list.

Commit: **Full test and report added**
Heribert Pascual Saldanya authored 10 months ago

Files:

- test_audit_L00_16GPU
- test_audit_L00_1GPU
- test_audit_L00_2GPU
- test_audit_L00_4GPU
- test_audit_L00_8GPU
- informe_L00_audit.pdf
- output.json.out
- test_audit_L00.json



To be decided
(is it necessary?)



git-based catalogue: PDs (workflows)

- > **Code audits and tests**
CHEESE-2P: database of POP code audits and tests
- > **Flagship codes**
CHEESE-2P: catalogue of flagship codes
- > **Mini-apps**
CHEESE-2P: catalogue of mini-apps
- > **Pilot Demonstrators (PDs)**
CHEESE-2P: catalogue of Pilot Demonstrators (Workflows)
- > **Simulation Cases (SCs)**
CHEESE-2P: database of Simulation Cases
- Documentation**
CHEESE-2P: repository documentation and templates
- gitlab-profile**

- gitlab-profile**
- Pilot Demonstrator 1**
PD1: Extreme-scale modeling of seismic hazards
- Pilot Demonstrator 2**
PD2: High-resolution seismic inversion and tomography
- Pilot Demonstrator 3**
PD3: Global probabilistic tsunami hazard and uncertainty quantification
- Pilot Demonstrator 4**
PD4: Complex multi-source tsunami modeling
- Pilot Demonstrator 5**
PD5: Ensemble-based volcanic dispersal across multiple scales
- Pilot Demonstrator 6**
PD6: Multiphase 3D volcanic explosion modeling
- Pilot Demonstrator 7**
PD7: The Earth's dynamo model
- Pilot Demonstrator 8**
PD8: Geodynamics to geohazards
- Pilot Demonstrator 9**
PD9: Glacial outburst floods

Pilot Demonstrator 4

main pilot-demonstrator-4 / +

History Find file Edit Code

Update README.md
Cheese Admin authored 1 month ago 19c65c16

Name	Last commit	Last update
README.md	Update README.md	1 month ago

Some basic
documentation
(description, objectives)

README.md

Complex multi-source tsunami modeling

PD leader:

Partners:

Codes:

Derived SCs: [SC4.1](#)

Description

This PD will combine multiphysics processes from earthquakes, landslides, volcanoes, and acoustic pressure and atmospheric forcing to tsunami generation and propagation in a single workflow that allows coupling of different flagship codes. Particular attention is given to modeling processes that are not being tackled so far in other related projects (e.g. in the DT-GEO project towards DestinE). Several recent tsunami events worldwide have involved different geophysical processes that took place at the same time. For instance, the 2018 Palu tsunami involved both

Catalogue summary

JSON file with basic metadata
(problem: no standard metadata model for WFs exists...)

git-based catalogue: SCs

- >  **Code audits and tests** 
CHEESE-2P: database of POP code audits and tests
- >  **Flagship codes** 
CHEESE-2P: catalogue of flagship codes
- >  **Mini-apps** 
CHEESE-2P: catalogue of mini-apps
- >  **Pilot Demonstrators (PDs)** 
CHEESE-2P: catalogue of Pilot Demonstrators (Workflows)
- >  **Simulation Cases (SCs)** 
CHEESE-2P: database of Simulation Cases
-  **Documentation** 
CHEESE-2P: repository documentation and templates
-  **gitlab-profile** 

- SC1.1  SC1.1: Anthropogenic noise and seismicity, wave propagation at higher frequencies in Rade d'Hyères
- SC1.2  SC1.2: Next generation of earthquake simulators for the Hellenic arc and Iceland
- SC1.3  SC1.3: Physics-based PSHA and urgent UQ for Iceland: next generation hazard map
- SC2.1  SC2.1: Seismic tomography of the Adriatic-Ionian region
- SC3.1  SC3.1: The Global Tsunami Hazard Map
- SC4.1  SC4.1: Multi-source local high-resolution scenario with emphasis on the 2022 Tonga case
- SC5.1  SC5.1: Towards the European tephra hazard model
- SC5.2  SC5.2: Volcanic dispersal at local scale: a reanalysis of the 2021 La Palma case
- SC6.1  SC6.1: Urgent high-resolution 3D multiphase flow simulation of phreatic eruptions at Vulcano
- SC6.2  SC6.2: Long-term probabilistic hazard maps for phreatic eruptions at Vulcano Island
- SC7.1  SC7.1: MHD turbulence closure models in planetary cores using AI
- SC8.1  SC8.1: Magmatic systems and induced landslides: the Etna case
- SC8.2  SC8.2: 3D evolution of tectonic deformation as driver of seismic hazard in the Hellenic arc
- SC9.1  SC9.1: Eastern Skaffá cauldron GLOF simulation, Iceland
- SC9.2  SC9.2: Calving glacier hazard in Greenland

 **SC5.1** 

main sc5.1 / +

History Find file Edit Code

 **Update README.md**
Cheese Admin authored 1 month ago f3450513 

Name	Last commit	Last update
README.md	Update README.md	1 month ago
SC5.1.png	Upload New File	1 month ago
metadata.json	Add new file	2 months ago

Some basic
documentation
(description,
expected results)

JSON file with basic metadata
(problem: no standard metadata
model for simulations exists...)

 Catalogue summary

README.md

Towards the European tephra hazard model



SC leader:



Codes: [FALL3D](#)

Target EuroHPC Architectures: MNS, Leonardo

Type: Hazard assessment

Description

This capacity SC aims at deriving the first homogeneous multi-volcano probabilistic hazard model for volcanic tephra fallout in Europe. Due to data and knowledge heterogeneities, the main challenge relates to including multiple sources (i.e. volcanoes or volcanic districts) and, for each of them, capturing all possible eruptive scenarios relevant for tephra dispersal and fallout. Where the knowledge of volcanic sources is at sufficient detail (e.g. eruptive record

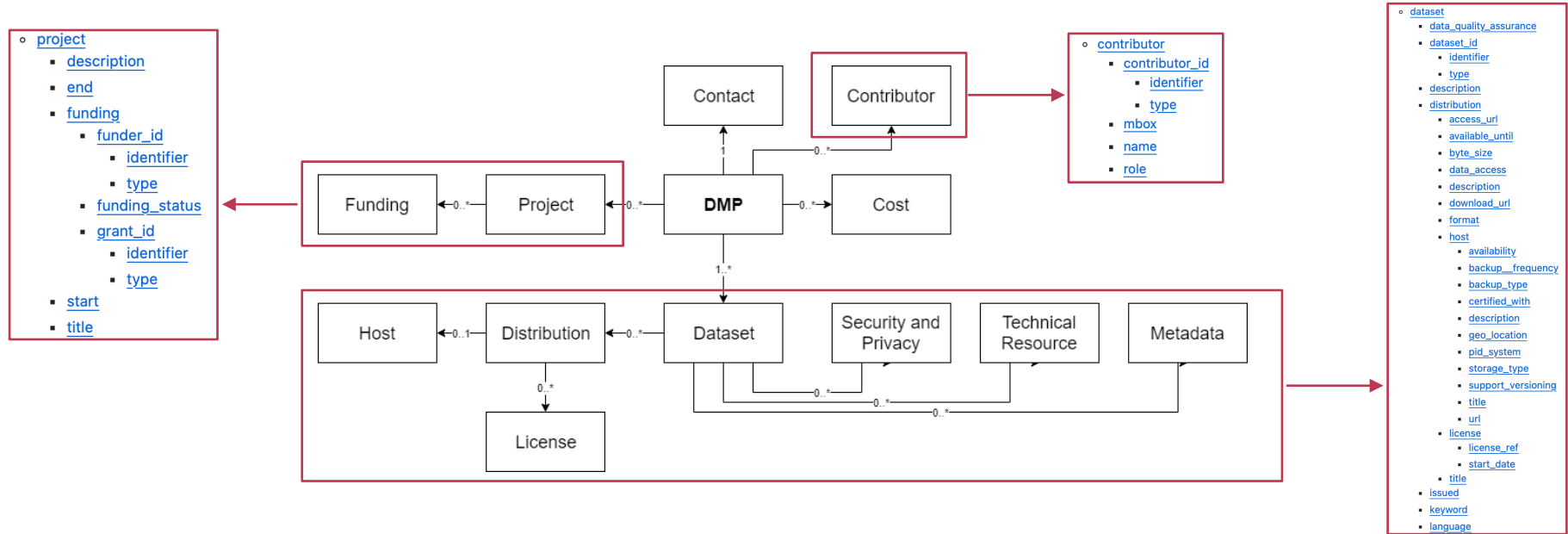
Metadata model(s): what to do?

- **Main problem:** a metadata standard model covering all typologies of DAs in ChEESE does not exist
- **Option 1:** Research Data Alliance (RDA) Data Management Plan Common Standard compiled in a JavaScript Object Notation (JSON) file
- **Option 2:** EPOS-DCAT-AP extensions (aligned with on-going DT-GEO and geo-INQUIRE developments)
- **Option 3:** Other? (open to suggestions)



Option 1: RDA-DMP common standard

- RDA-DMP: a **metadata application profile**(*) that provides basic interoperability between systems producing or consuming machine-actionable Data Management Plans (ma-DMPs)
- Further fields (extensions) can be added in specific deployments, but they do not guarantee interoperability
- Simple but very much oriented to **datasets** (may need our own extensions for codes, WFs and SCs)



(*) An application profile is a metadata design specification that uses a selection of terms from multiple metadata vocabularies

Option 2: EPOS-DCAT-AP extensions

DCAT (Data Catalog Vocabulary)

- A vocabulary for publishing and interoperate data catalogs on the Web
- Based on the Resource Description Framework (RDF), a World Wide Web Consortium (W3C) standard



DCAT-AP (Application Profile)

- An application profile of DCAT to be used in European data portals
- Universal metadata scheme, ready to be further profiled for specific domain needs



EPOS-DCAT-AP

- A an extension (profile) of DCAT-AP for Research Infrastructures in environmental and solid-Earth domains
- EPOS uses EPOS-DCAT-AP in metadata ingestion pipelines to populate TCS from the heterogeneous metadata formats
- The EPOS metadata catalogue (their DCAT-AP extension profile) is based on CERIF(*)



Extension of EPOS-DCAT-AP for **simulations**



Extension of EPOS-DCAT-AP for **workflows**

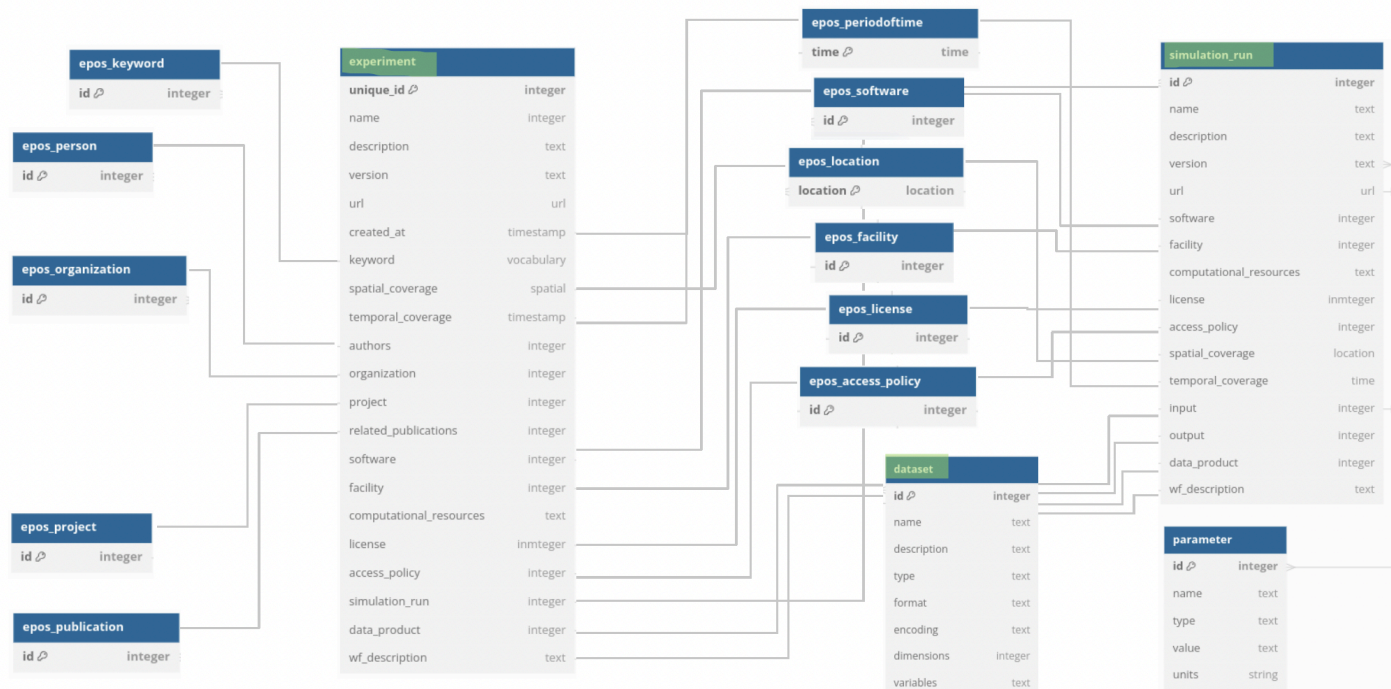
(*) CERIF (Common European Research Information Format): standard that the EU recommends to its member states for recording information about research activity. Since version 1.6 it has included specific support for recording metadata for datasets



EPOS-DCAT-AP extension (simulations)

Geo-INQUIRE Metadata Schema:

EPOS-DCAT-AP is a dialect of DCAT-AP



EPOS-DCAT-AP extension (simulations)

Simulation Data Lake Dataset (=Experiment)

EXPERIMENT

An experiment is made of 1- to-100K simulation runs
Up to tens of TB

SIMULATION

DATASETS

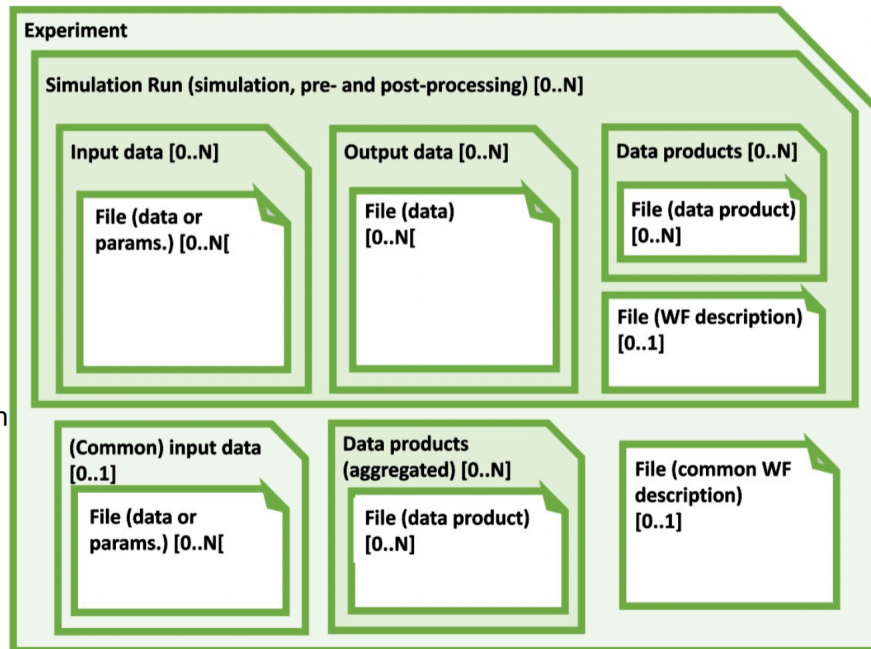
Input data:

- File of parameters (common to all or specific to each simulation run).
- Input data (common to all or specific to each simulation run).

Output data:

- Output data for each simulation run (e.g., netcdf, hdf5).
- Postprocessing
- Scripts
- Data products
- Figures

Workflow description



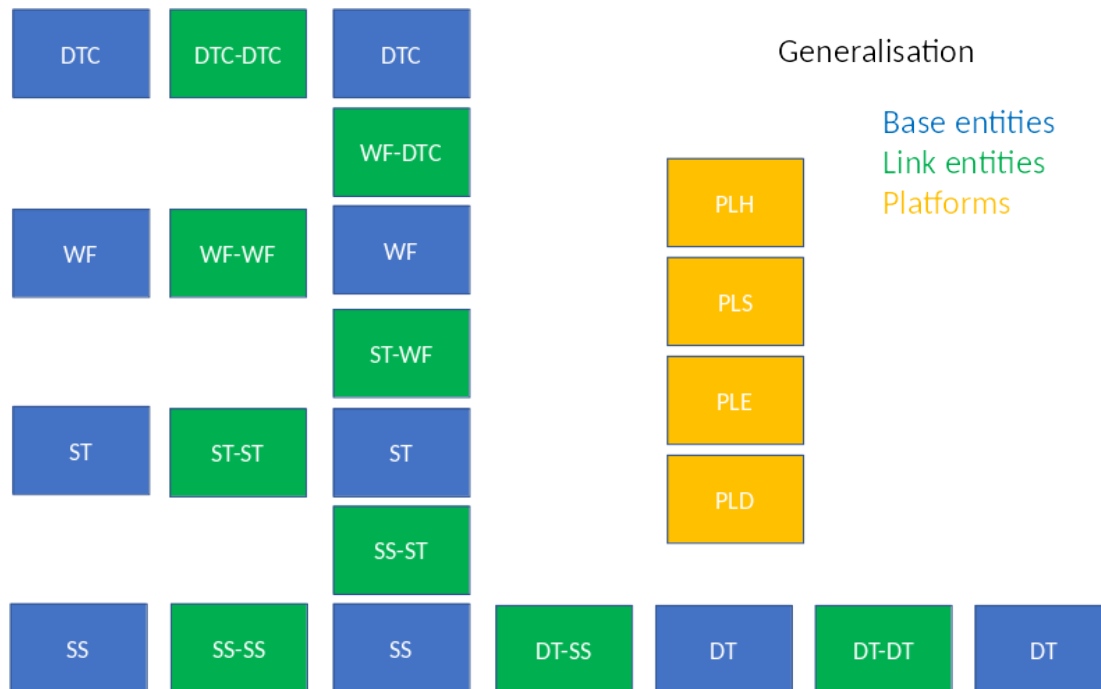
(Common) DATASETS

- Input
- Data product



EPOS-DCAT-AP extension (workflows)

- **DT-GEO**: further extension of EPOS-DCAT-AP to accommodate metadata for new entities in the DT-GEO schema such as Workflow and Step
- Adoption and extension of the EOSC-synergy SQAaaS platform for assuring quality on DAs (e.g. FAIR-EVA has been extended with a specific plugin to integrate with the EPOS-DCAT-AP catalog).



Summary (pros and cons)

RDA - DMP

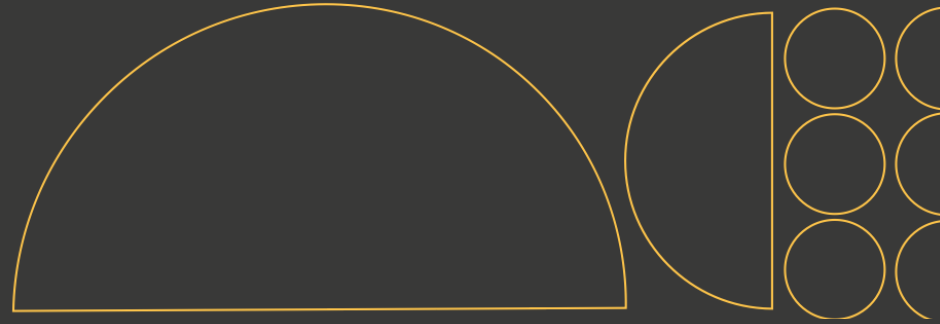
- ✓ Standard for DMPs
- ✓ Complies with the ChEESE ma-DMP requirements (e.g. interoperability)
- ✓ Easy to implement (less effort); e.g. basic JSON files
- ✗ Limited to datasets (would need basic extensions for codes, WFs, SCs)
- ✗ Extra effort will be needed to populate the SDL@CINECA

EPOS-DCAT-AP

- ✓ More general, aligned with EPOS
- ✓ Join efforts with parallel initiatives developing extensions for WFs and simulations
- ✓ Easier integration to the SDL@CINECA
- ✓ Services for FAIRness quantification will be available (e.g. FAIR-EVA)
- ✗ Harder to implement (do we have the **skills and man-power** for it?)
- ✗ **Risky**: depends on the progress (and success) of other projects



Thank you!



@cheese-coe



<http://cheese2.eu>



@cheese-coe



@cheese_coe@techhub.social

