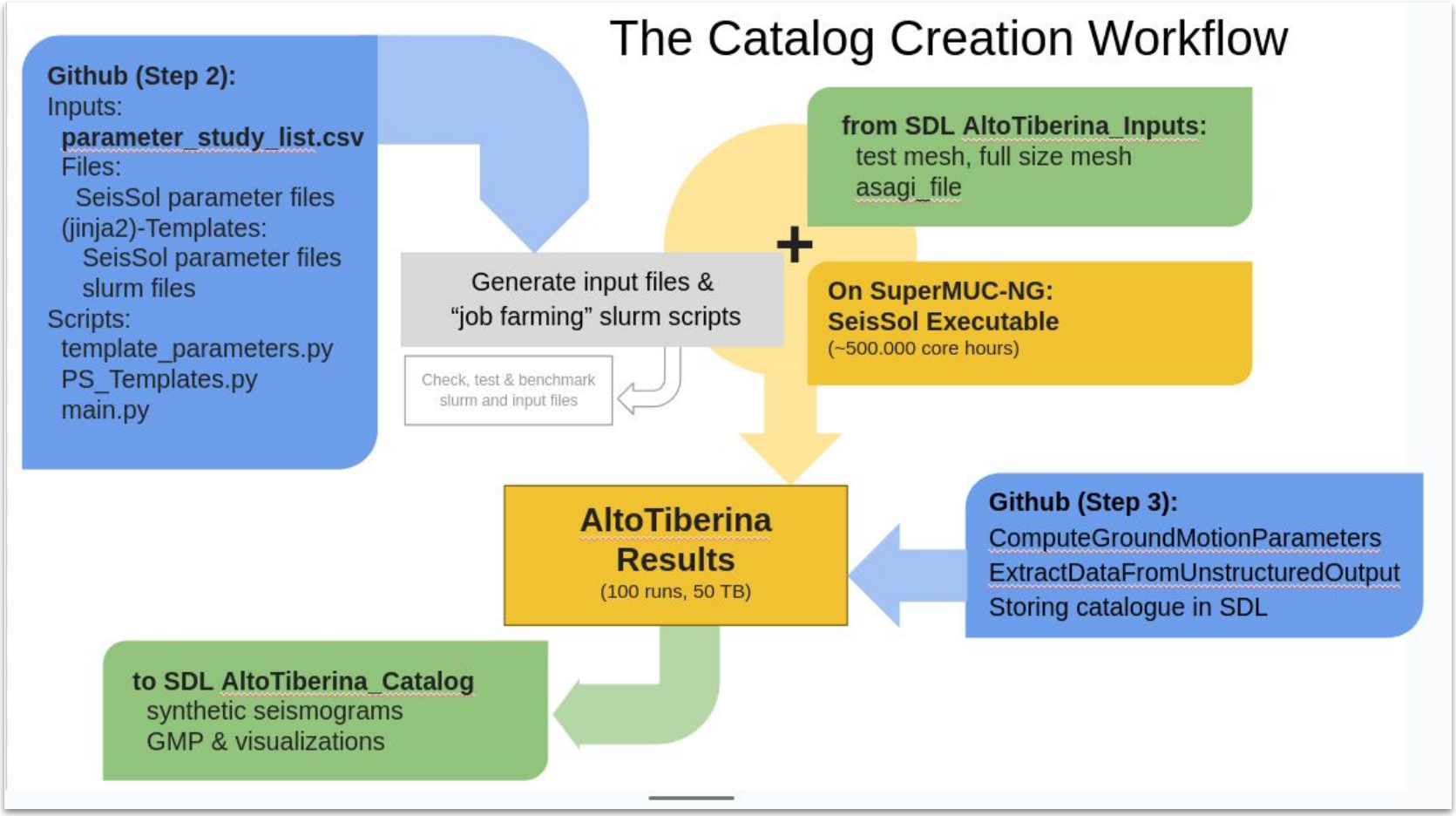
Using the SDL for one of LMU's DT-Geo Workflows

Iris Christadler and Mathilde Marchandon (LMU)

Contribution to "FAIR 7: FAIRy Tales & Forum"
May 20th, Geo-INQUIRE Online Training

Geo-INQUIRE *is funded by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.*

The Catalog Creation Workflow



The Rapid Response Workflow

README.md *(setup once)*

1. create venv with **requirements.yaml**
2. test venv with **cronjob.sh**
3. add crontab entry from **cronjob.txt**

Cronjob detect_event.py *(runs every hour)*

detect_event.py via obspy.clients.fdsn from https://www.seismicportal.eu

if no_earthquake_happened : exit; else:

download_waveforms.py from TABOO Near Fault Observatory

search_catalog.py for best fit with data from **AltoTiberina_catalog**

outputs: closest simulation from catalog

workflow schema

Github (Step 4):
README.md
environment.yml

WF_Step4_cronjob.sh
WF_Step4_cronjob.txt

detect_event.py
download_waveforms.py

Github (Step 5):
search_catalog.py

SDL:
AltoTiberina_catalog

input data

LMU's DT- Geo Workflow for AltoTiberina

WF7403

ST740301 (Input Data: Mesh, Asagi, CSV file):
https://sdl-dev.hpc.cineca.it/app/experiments/483/

ST740302 (Assemble Input Parameters and Slurm files):
https://github.com/christadler/SeisSol_ParamStudies

ST740303 (Generate Data Products: GMP, Visualization):
https://github.com/christadler/SeisSol_Postprocessing

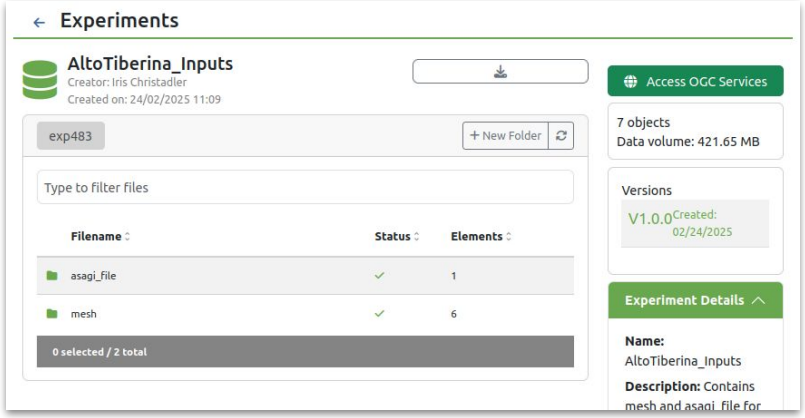
Intermediate Result: Catalog
https://sdl-dev.hpc.cineca.it/app/experiments/567/

ST740304 (Download Real Time Data):
https://github.com/marcmath/download_TABOO_waveforms

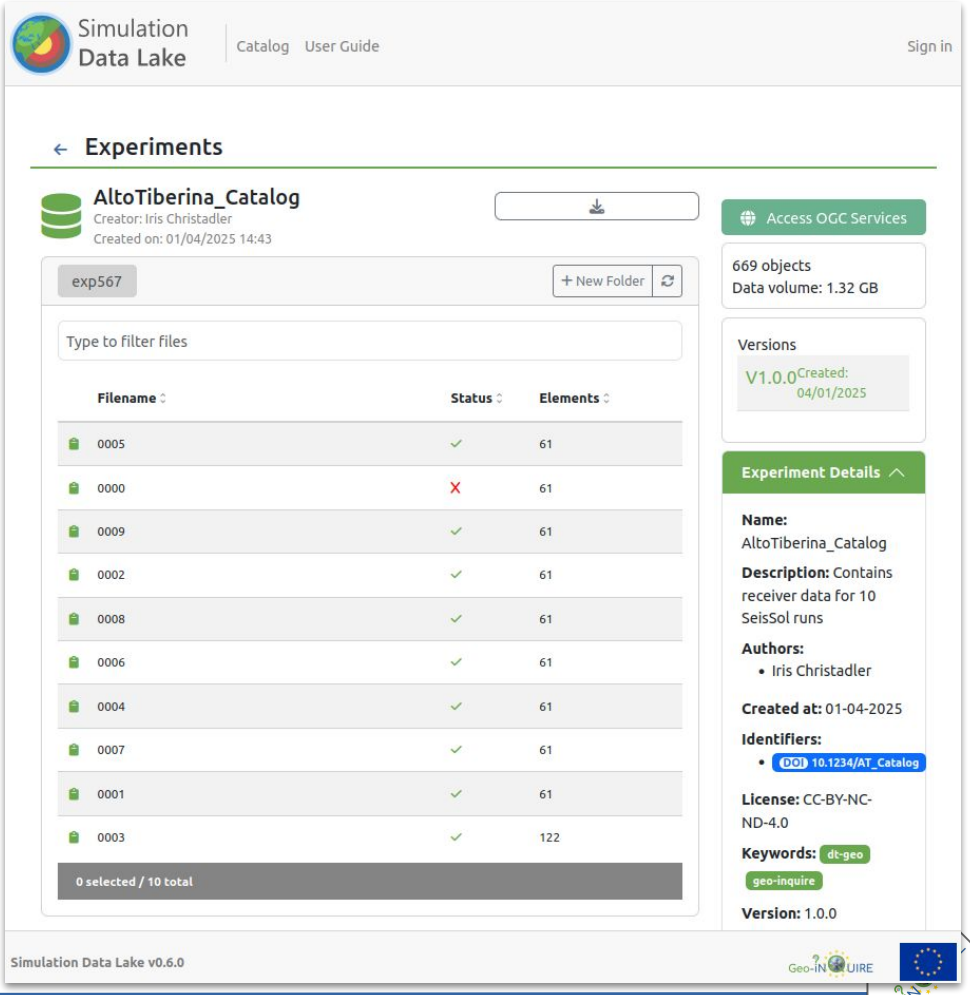
ST740305 (Find closest match):
https://github.com/NicoSchlw/search_SeisSol_ensemble

		high	Developed	6	This step includes the automatic creation of a large number of input files with different parameter settings and the automatic scheduling of the associated simulations
ST740302	Forward dynamic rupture simulation	high	Fully implemented	7	During this step, SeisSol runs the simulations prepared in step ST740301
ST740303	Ensemble scenarios	high	Under development	5	The scenarios computed in step ST740302 are post processed and the associated shakemaps are computed.
ST740304	Seismic and GPS data	high	Under development	2	In this step, seismic waveforms associated with a given earthquake are downloaded. For the time being, the ID and date of the earthquake has to be provided. In future
ST740305	Single scenario dynamics and shake maps	high	Under development	6	This step compares the data with synthetics of the earthquake catalog to find the scenario that best fits the data. It will also output the shakemap associated with the best-fitting scenario.

Input Files:



Catalog :

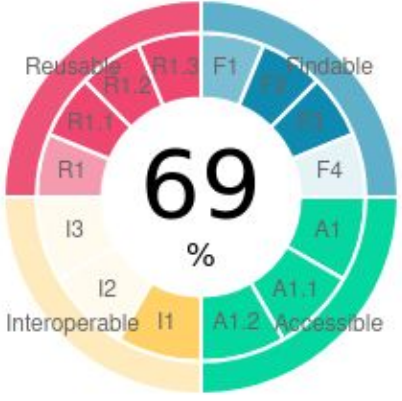


Zenodo File Size limitations:
50 GB per record, max. 100 files
one-time quote of 200GB possible (upon request), use ZIP if you have more files (already for 20+ files)
Simulation Data Lake:
Dev: in total 1 TB, Prod: 500 TB in total

Conclusions

Findable Accessible Interoperable Reusable

- Perfect, for all data from simulations
- Perfect, for big data and many files
- Use the SDL wizard to upload your data
- To get a FAIR score from the FUJI tool, you'll need to wrap your URL
- Very convenient for Reusability and Reproducibility:
  - For exchange between collaborators
  - For tutorials
  - For paper publishing
  - For "artifact descriptions"
  - …



Findable: yes
Accessible: yes
Interoperable: yes

Reusable: yes

Reproducibility given for the Catalog-Creation WF, **Reusability** for the Rapid-Response WF

Thank you for your attention!

If you have further questions, feel free to contact us:
iris.christadler@lmu.de m.marchandon@lmu.de
Both workflows could be adapted to other regions and (maybe) even to other software

Geo-INQUIRE is a joint effort of 51 institutions

Geo-INQUIRE is funded by the European Commission under project number 101058518 within the HORIZON-INFRA-2021-SERV-01 call.