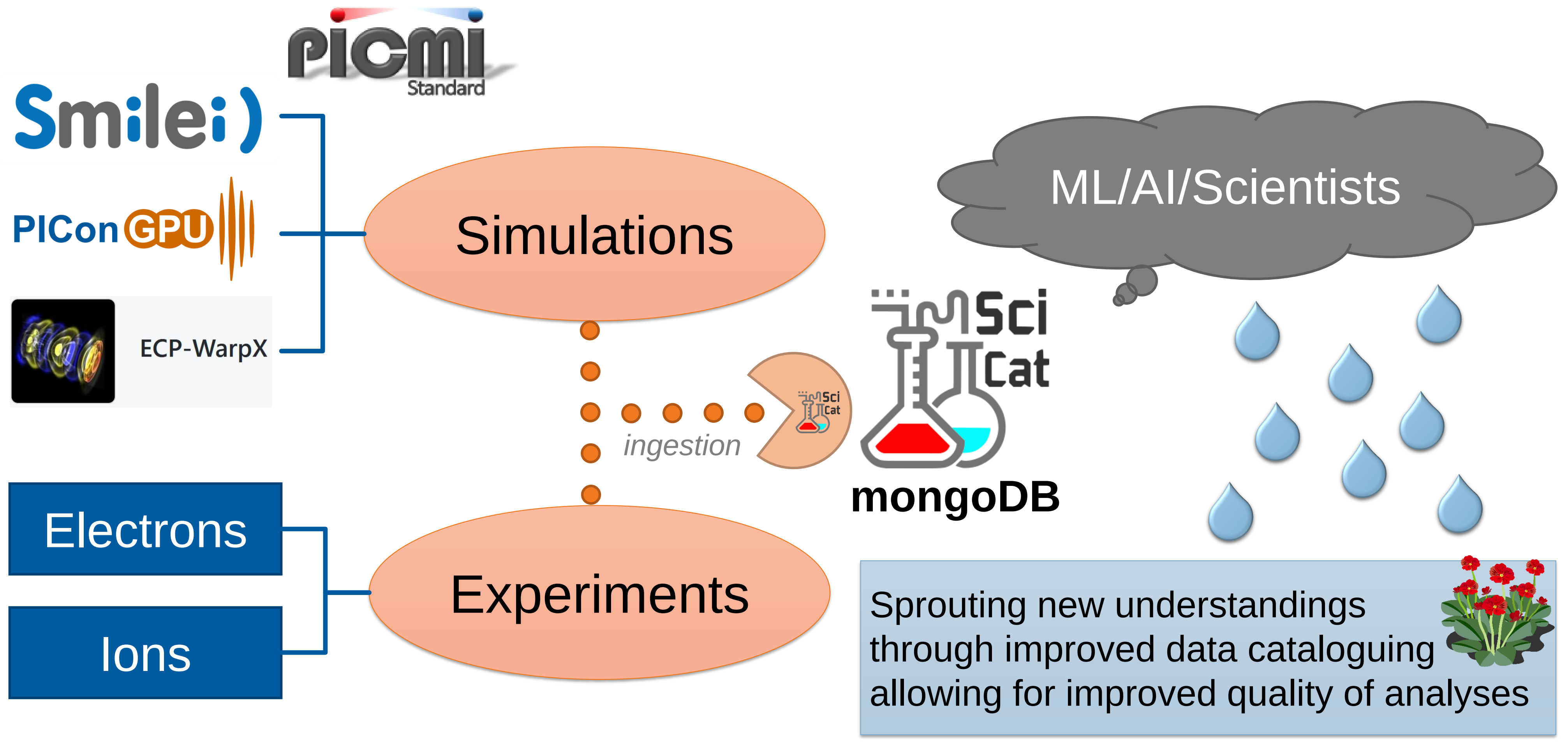
Enabling FAIRer Simulation and Experiment Data Cataloging at Laser Particle Acceleration

Kristin Tippey, Hans-Peter Schlenvoigt, Thomas Kluge, Brian Marre, Oliver Knodel, David Pape

Long-term Goals

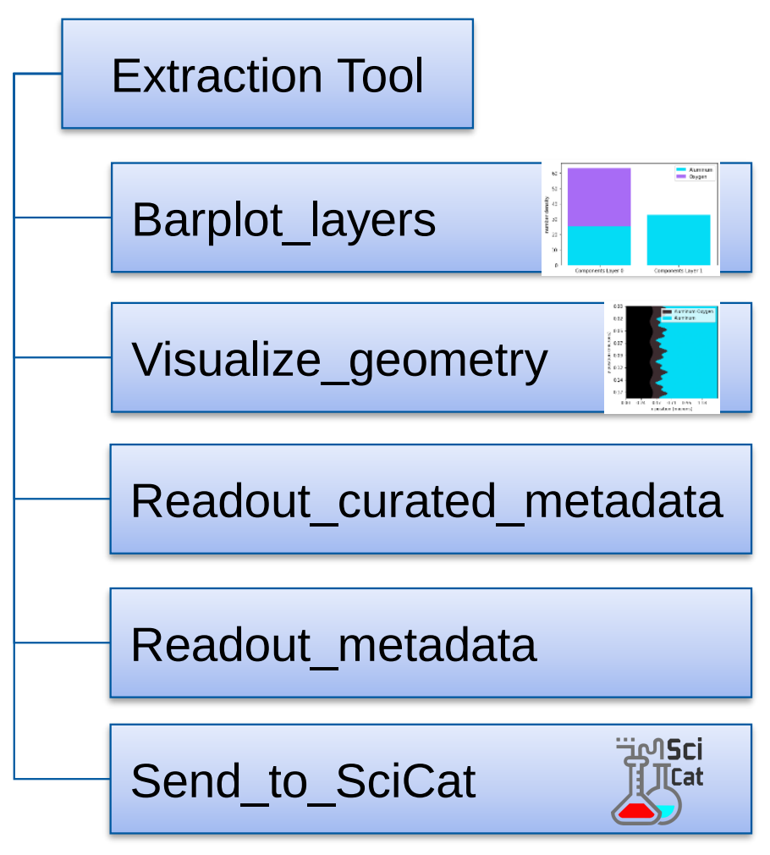
- Developing and deploying a clear and consistent **metadata schema**
- To enable FAIR (Findable Accessible Interoperable Reusable) qualities
- Scraping/collecting this metadata from simulation and experimental datasets, using the existing architecture as much as possible
- Uploading to a dynamic and queryable databases such as SciCat or mongoDB
- Goal: **Enabling more robust human- and machine- learning**

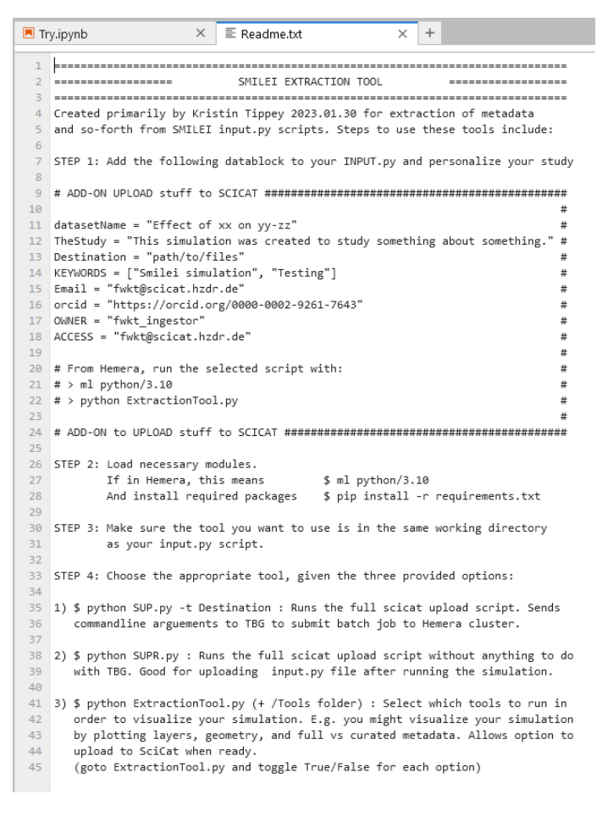


Sprouting new understandings through improved data cataloguing allowing for improved quality of analyses

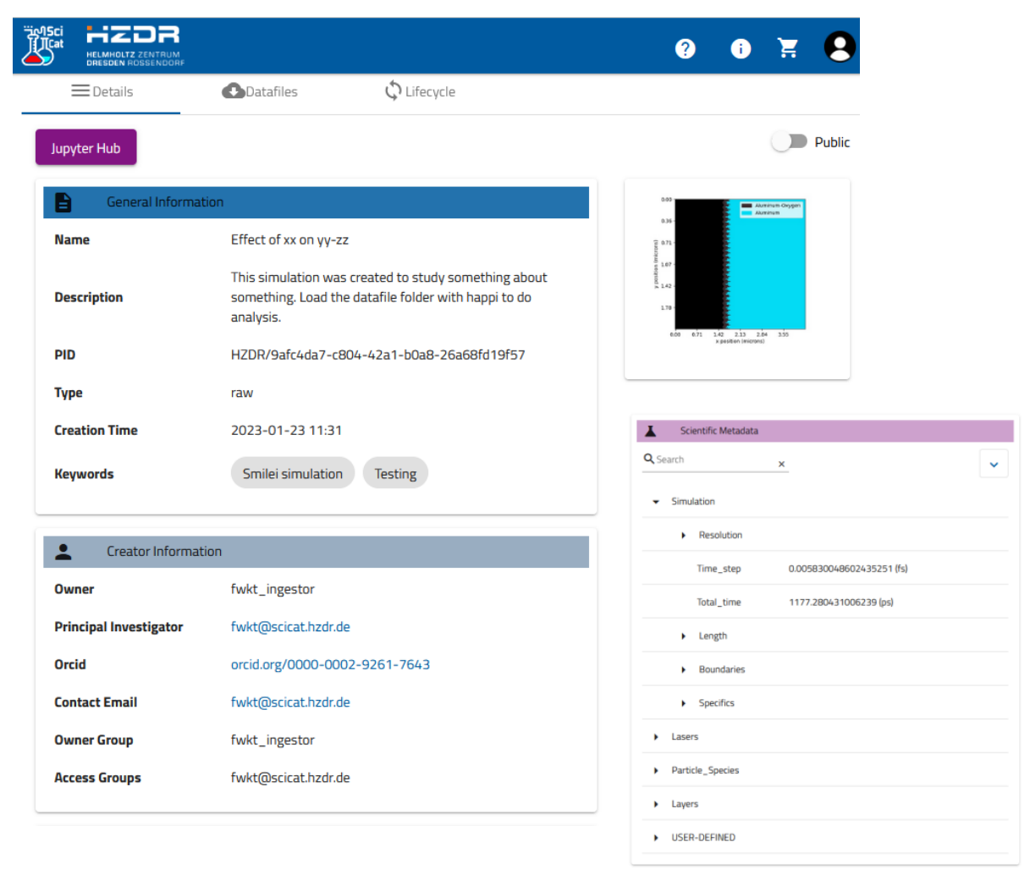
Metadata Extraction Toolkits for Simulations

- Developed a toolkit for extracting and passing metadata from **SMILEI** input files to SciCat
- Extended implementation to **wrap around TBG** script, leaving user with minimal add-on work to upload and review metadata with SciCat
- Scripts can be wrapped around or run separately from simulation. ***Status: Being tested/used***





```
================================================================================
=================            SMILEI EXTRACTION TOOL            =================
================================================================================
Created primarily by Kristin Tippey 2023.01.30 for extraction of metadata
and so-forth from SMILEI input.py scripts. Steps to use these tools include:

STEP 1: Add the following datablock to your INPUT.py and personalize your study

# ADD-ON UPLOAD stuff to SCICAT ###############################################
                                                                              #
datasetName = "Effect of xx on yy-zz"                                         #
TheStudy = "This simulation was created to study something about something." #
Destination = "path/to/files"                                                 #
KEYWORDS = ["Smilei simulation", "Testing"]                                   #
Email = "fwkt@scicat.hzdr.de"                                                 #
orcid = "https://orcid.org/0000-0002-9261-7643"                               #
OWNER = "fwkt_ingestor"                                                       #
ACCESS = "fwkt@scicat.hzdr.de"                                                #
                                                                              #
# From Hemera, run the selected script with:                                  #
# > ml python/3.10                                                            #
# > python ExtractionTool.py                                                  #
                                                                              #
# ADD-ON to UPLOAD stuff to SCICAT ############################################

STEP 2: Load necessary modules.
        If in Hemera, this means         $ ml python/3.10
        And install required packages    $ pip install -r requirements.txt

STEP 3: Make sure the tool you want to use is in the same working directory
        as your input.py script.

STEP 4: Choose the appropriate tool, given the three provided options:

1) $ python SUP.py -t Destination : Runs the full scicat upload script. Sends
   commandline arguements to TBG to submit batch job to Hemera cluster.

2) $ python SUPR.py : Runs the full scicat upload script without anything to do
   with TBG. Good for uploading  input.py file after running the simulation.

3) $ python ExtractionTool.py (+ /Tools folder) : Select which tools to run in
   order to visualize your simulation. E.g. you might visualize your simulation
   by plotting layers, geometry, and full vs curated metadata. Allows option to
   upload to SciCat when ready.
   (goto ExtractionTool.py and toggle True/False for each option)
```



General Information

Name	Effect of xx on yy-zz
Description	This simulation was created to study something about something. Load the datafile folder with happi to do analysis.
PID	HZDR/9afc4da7-c804-42a1-b0a8-26a68fd19f57
Type	raw
Creation Time	2023-01-23 11:31
Keywords	Smilei simulation Testing

Creator Information

Owner	fwkt_ingestor
Principal Investigator	fwkt@scicat.hzdr.de
Orcid	orcid.org/0000-0002-9261-7643
Contact Email	fwkt@scicat.hzdr.de
Owner Group	fwkt_ingestor
Access Groups	fwkt@scicat.hzdr.de

Simulation
Resolution
Time_step 0.005830048602435251 (fs)
Total_time 1177.280431006239 (ps)
Length
Boundaries
Specifics
Lasers
Particle_Species
Layers
USER-DEFINED

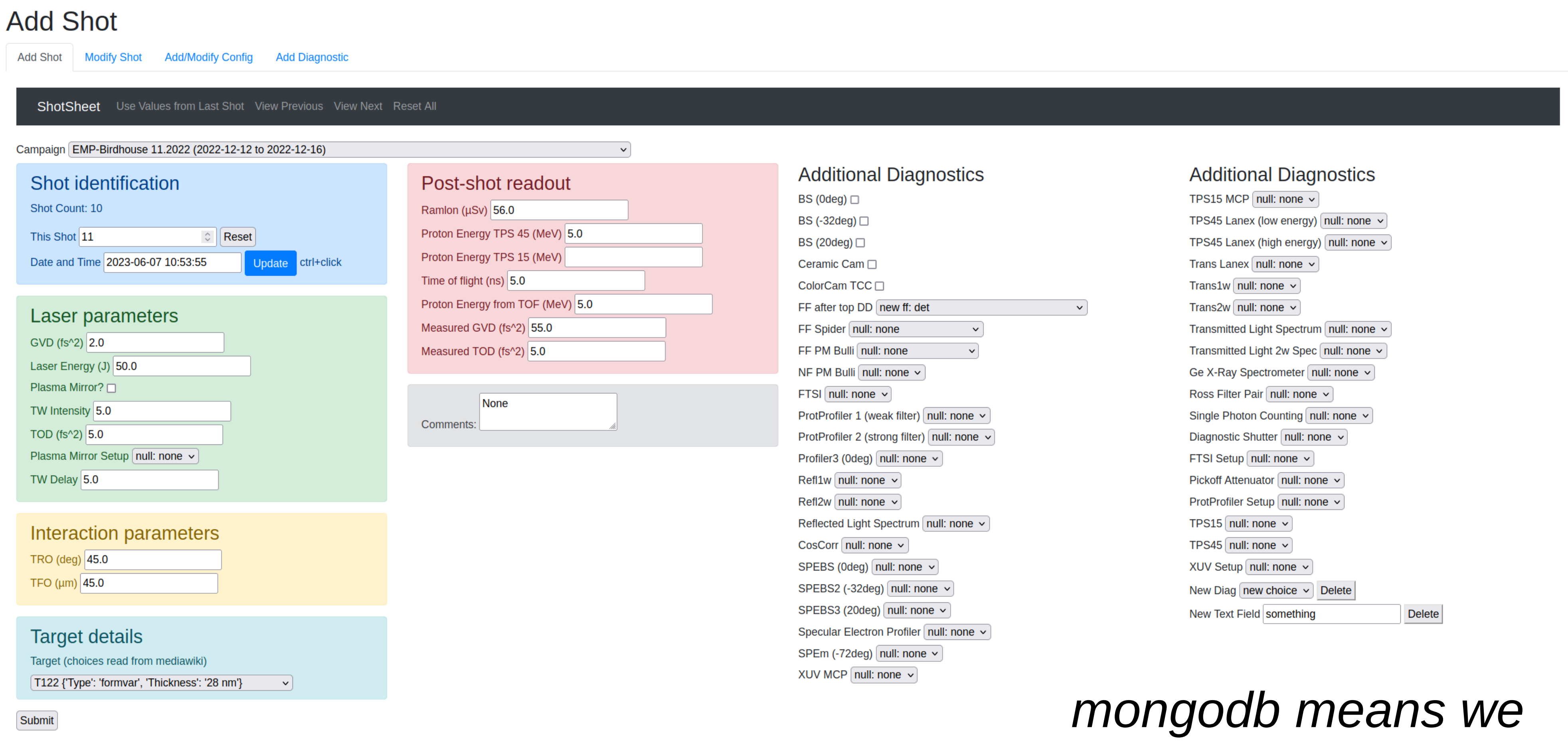
- ***Saved and shared on GitLab, alongside script for searching through SciCat***

https://codebase.helmholtz.cloud/tippey27/working-with-scicat
https://codebase.helmholtz.cloud/tippey27/parse_picongpu

- Adapted scripts to support scraping and uploading from **picmi, PIConGPU,** and **WarpX** codes
- Future work involves working with simulationists to better support broader variety of input conventions and with PIC developers to share SciCat options

Python-flask-WTForms-mongoDB ShotSheet App for Experiment data

- Developing an application that intakes shot data and configurations using python-flask-wtforms, mediawiki and javascript; and saves to mongoDB, for scripted analysis and/or visualization with Grafana
- Allows dynamic adding/modifying of shots, configurations, and diagnostics … mongoDB allows for more dynamicism than SQL
- **Soon: Real trials alongside existing MSAccess ShotSheet**



Campaign: EMP-Birdhouse 11.2022 (2022-12-12 to 2022-12-16)

Shot identification
Shot Count: 10
This Shot: 11
Date and Time: 2023-06-07 10:53:55

Laser parameters
GVD (fs^2): 2.0
Laser Energy (J): 50.0
Plasma Mirror? □
TW Intensity: 5.0
TOD (fs^2): 5.0
Plasma Mirror Setup: null: none
TW Delay: 5.0

Interaction parameters
TRO (deg): 45.0
TFO (μm): 45.0

Target details
Target (choices read from mediawiki): T122 {'Type': 'formvar', 'Thickness': '28 nm'}

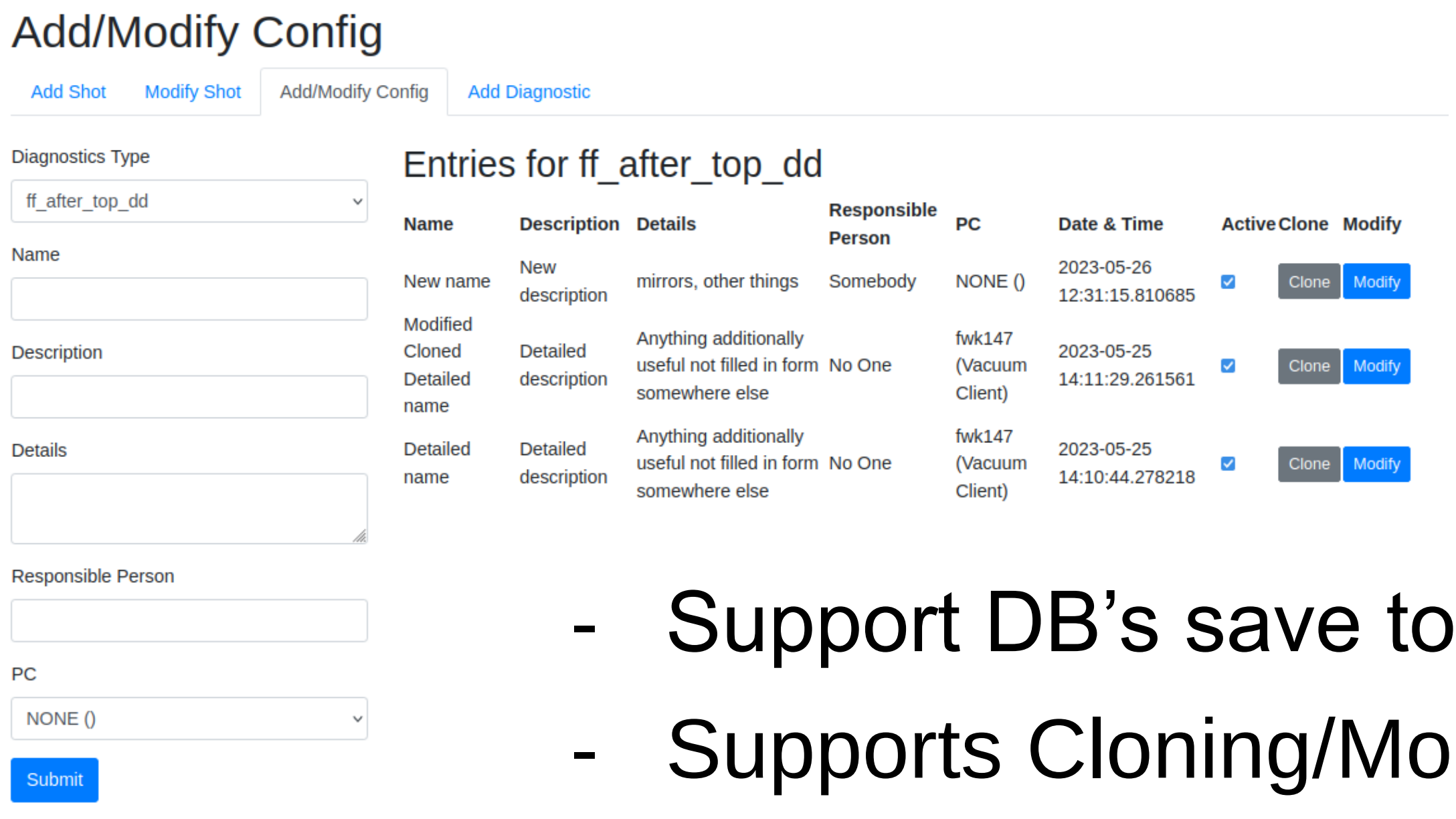
Post-shot readout
Ramlon (μSv): 56.0
Proton Energy TPS 45 (MeV): 5.0
Proton Energy TPS 15 (MeV):
Time of flight (ns): 5.0
Proton Energy from TOF (MeV): 5.0
Measured GVD (fs^2): 55.0
Measured TOD (fs^2): 5.0

Comments: None

Additional Diagnostics
BS (0deg) □
BS (-32deg) □
BS (20deg) □
Ceramic Cam □
ColorCam TCC □
FF after top DD: new ff: det
FF Spider: null: none
FF PM Bulli: null: none
NF PM Bulli: null: none
FTSI: null: none
ProtProfiler 1 (weak filter): null: none
ProtProfiler 2 (strong filter): null: none
Profiler3 (0deg): null: none
Refl1w: null: none
Refl2w: null: none
Reflected Light Spectrum: null: none
CosCorr: null: none
SPEBS (0deg): null: none
SPEBS2 (-32deg): null: none
SPEBS3 (20deg): null: none
Specular Electron Profiler: null: none
SPEm (-72deg): null: none
XUV MCP: null: none

Additional Diagnostics
TPS15 MCP: null: none
TPS45 Lanex (low energy): null: none
TPS45 Lanex (high energy): null: none
Trans Lanex: null: none
Trans1w: null: none
Trans2w: null: none
Transmitted Light Spectrum: null: none
Transmitted Light 2w Spec: null: none
Ge X-Ray Spectrometer: null: none
Ross Filter Pair: null: none
Single Photon Counting: null: none
Diagnostic Shutter: null: none
FTSI Setup: null: none
Pickoff Attenuator: null: none
ProtProfiler Setup: null: none
TPS15: null: none
TPS45: null: none
XUV Setup: null: none
New Diag: new choice [Delete]
New Text Field: something [Delete]

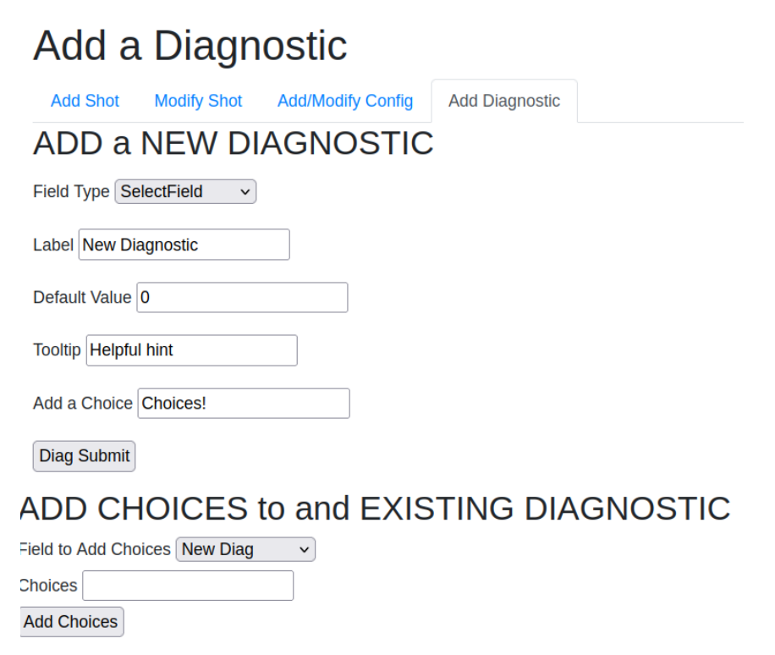
mongodb means we can add fields without migrating databases



Diagnostics Type: ff_after_top_dd
Name
Description
Details
Responsible Person
PC: NONE ()

Entries for ff_after_top_dd

Name	Description	Details	Responsible Person	PC	Date & Time	Active	Clone	Modify
New name	New description	mirrors, other things	Somebody	NONE ()	2023-05-26 12:31:15.810685	☑	Clone	Modify
Modified Cloned Detailed name	Detailed description	Anything additionally useful not filled in form somewhere else	No One	fwk147 (Vacuum Client)	2023-05-25 14:11:29.261561	☑	Clone	Modify
Detailed name	Detailed description	Anything additionally useful not filled in form somewhere else	No One	fwk147 (Vacuum Client)	2023-05-25 14:10:44.278218	☑	Clone	Modify



ADD a NEW DIAGNOSTIC
Field Type: SelectField
Label: New Diagnostic
Default Value: 0
Tooltip: Helpful hint
Add a Choice: Choices!

ADD CHOICES to and EXISTING DIAGNOSTIC
Field to Add Choices: New Diag
Choices:

- Support DB's save to their own mongodb collections
- Supports Cloning/Modifying of config entries
- Supports Active/Inactive qualifications for configs

https://codebase.helmholtz.cloud/tippey27/shotsheet

HZDR · Laser Particle Acceleration Division
Kristin Tippey, PhD · k.tippey@hzdr.de · www.hzdr.de