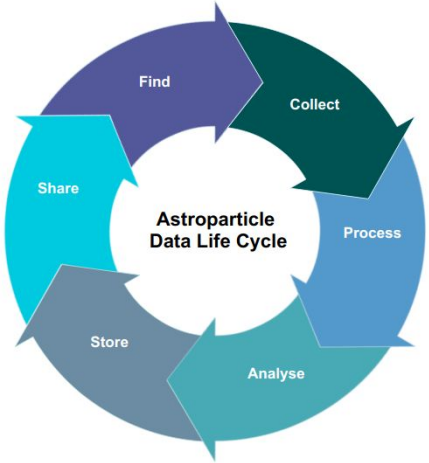
German-Russian Astroparticle Data Life Cycle Initiative to foster Big Data Infrastructure for Multi-Messenger Astronomy

Victoria Tokareva for the GRADLCI Collaboration

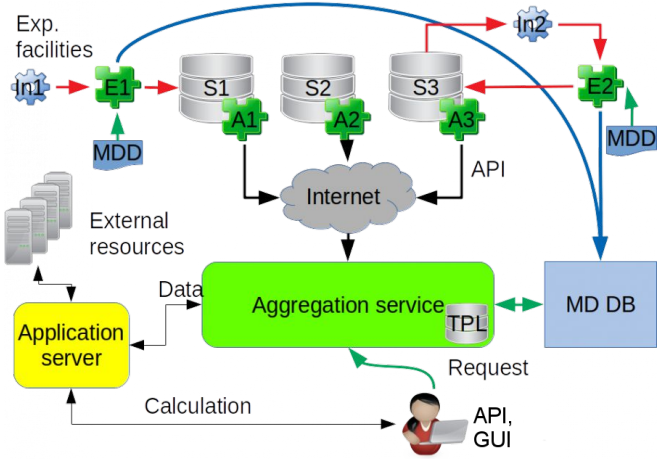
PoS(ICRC2021)938



Motivation

- Need for efficient management of large and big data
- Providing environment for joint analysis and multi-messenger astronomy
- With respect to the experiments' DLC

Prototype analysis and data center



- S_i — remote data storages
- In_i — data sources
- **MDD** — metadata description
- E_i — metadata extractors
- A_i — adapters, provide API for data access
- **TPL** — template library
- **MDDB** — metadata database

KCDC[1] update

- OCEANUS - Nov 2019:
  - \+ LOPES data
  - \+ increase of the processing speed up 10 to 50x
- PENTARUS - May 2020:
  - \+ introduced the first COMBINED DataShop for joint data analysis of the KASCADE and GRANDE detector arrays together with matching simulations
- SKARAGAN - February 2021:
  - \+ data for the 'Maket-Ani' experiment
  - \+ introduce KCDC API
  - \+ add the 100th (!) spectrum

https://kcdc.iap.kit.edu/

See also: PoS(ICRC2021)422

JupyterHub for the application server

- Login via KCDC credentials
- Administration using Docker Swarm

https://jupyter.iap.kit.edu/

Aggregation server

Technical realization

- MySQL DB, TimescaleDB for MDDB
- Flask + JSON-RPC for user interface
- Docker for virtualization
- Kaitai for metadata extraction from binary files
- Python3.8, C++

Datasets

Data of different reconstruction level and simulations for: KASCADE, KASCADE-Grande, LOPES, Tunka-133, TunkaRex, Maket-Ani

Web API

Possible requests

- Data requests
- Request status
- List of requests
- Remove request from the list
- Download file

Request states

- Running
- Scheduled
- Finished
- Failed
- Deleting
- Expired

https://gradlc-dc.ikp.kit.edu/

Machine learning for data analysis

- Analysis of EAS detected by the TAIGA and KASCADE experiments
- Research directions: identification of primary particle type [1], reconstruction of spectrum mass composition [2], evaluation of EAS energy [3]
- Datasets generation - CORSIKA [4]
- Machine learning methods: Decision Trees, Random Forest, Convolutional Neural Networks
- Frameworks: sklearn, TensorFlow, pyTorch

See also:
PoS(ICRC2021)319

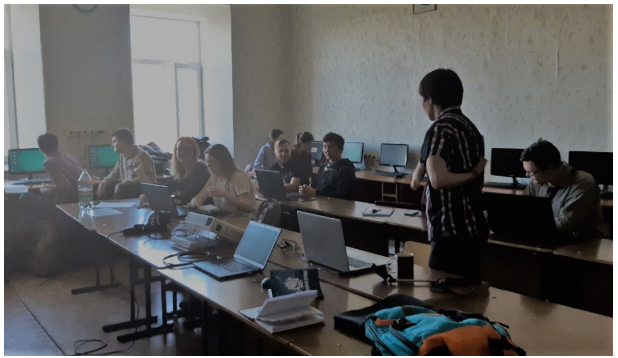
References

[1] *Postnikov E. et al., arXiv:1907.10480, 2019*
[2] *Postnikov E. et al., arXiv:1812.01551, 2018*
[3] *Postnikov E. et al., arXiv:1811.11822, 2018*
[4] *Heck D. et al., Report fzka 6019, 1998*

Outreach and education

- ~10 masterclasses
- course ”Introduction to astroparticle physics” at ISU
- > 300 students
- Multimessenger astronomy
- Online materials on https://astroparticle.online/
- in close collaboration with KCDC

See also:
PoS(ICRC2021)1378, PoS(ICRC2021)1373



Tutorial organized by the astroparticle.online team

This work was supported by Russian Science Foundation Grant №18-41-06003 and the Helmholtz Society Grant №HRSF-0027. Author acknowledges the support by the Doctoral School "Karlsruhe School of Elementary Particle and Astroparticle Physics: Science and Technology (KSETA)"

tokareva.victoria@kit.edu