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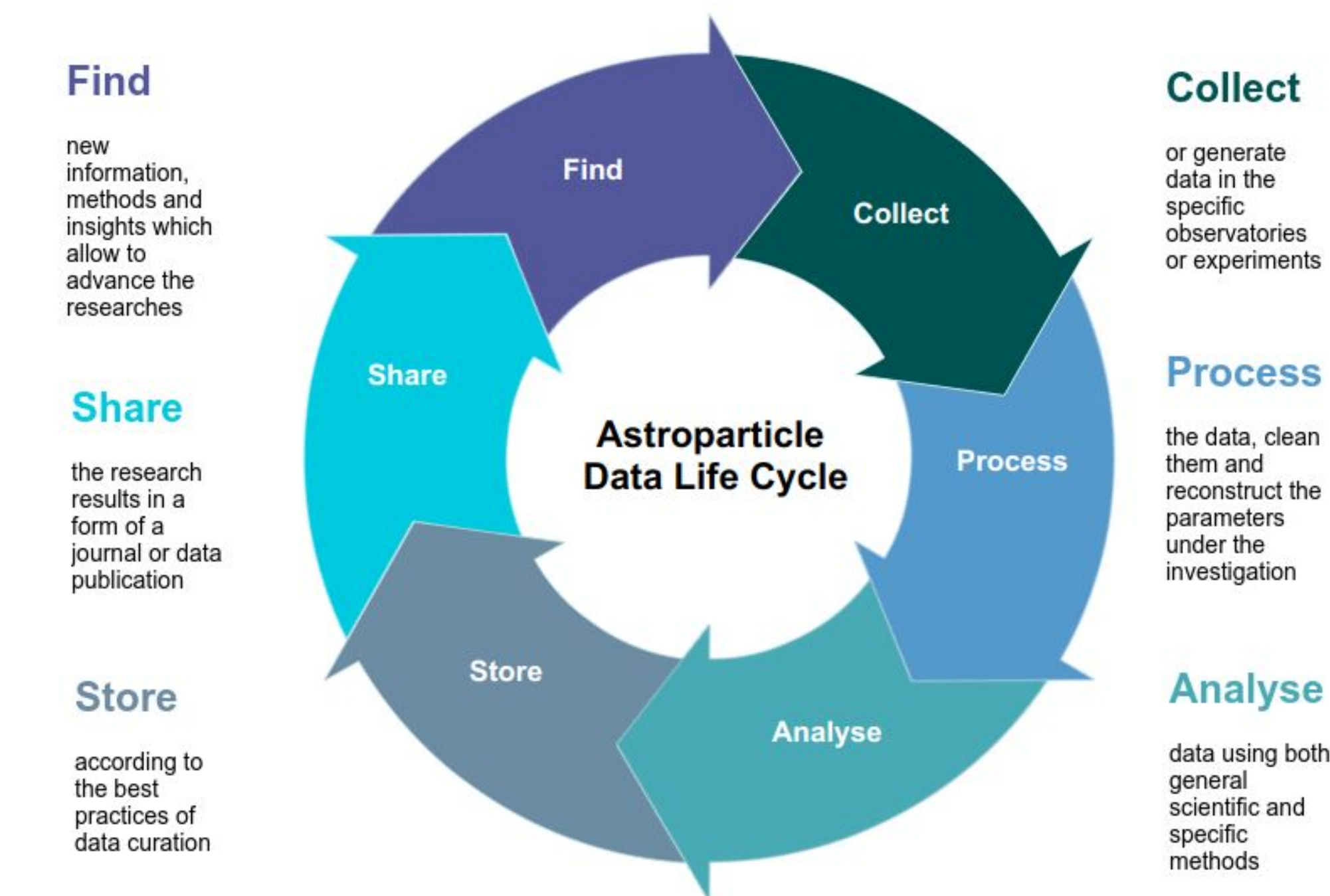 German-Russian Astroparticle Data Life Cycle Initiative to foster Big Data Infrastructure for Multi-Messenger Astronomy

GRADLCI

ONLINE ICRC 2021
THE ASTROPARTICLE PHYSICS CONFERENCE
Berlin | Germany
37th International Cosmic Ray Conference
12–23 July 2021

KIT
Karlsruhe Institute of Technology
Institut für Astroteilchenphysik (IAP)

Victoria Tokareva[+] for the GRADLC Initiative

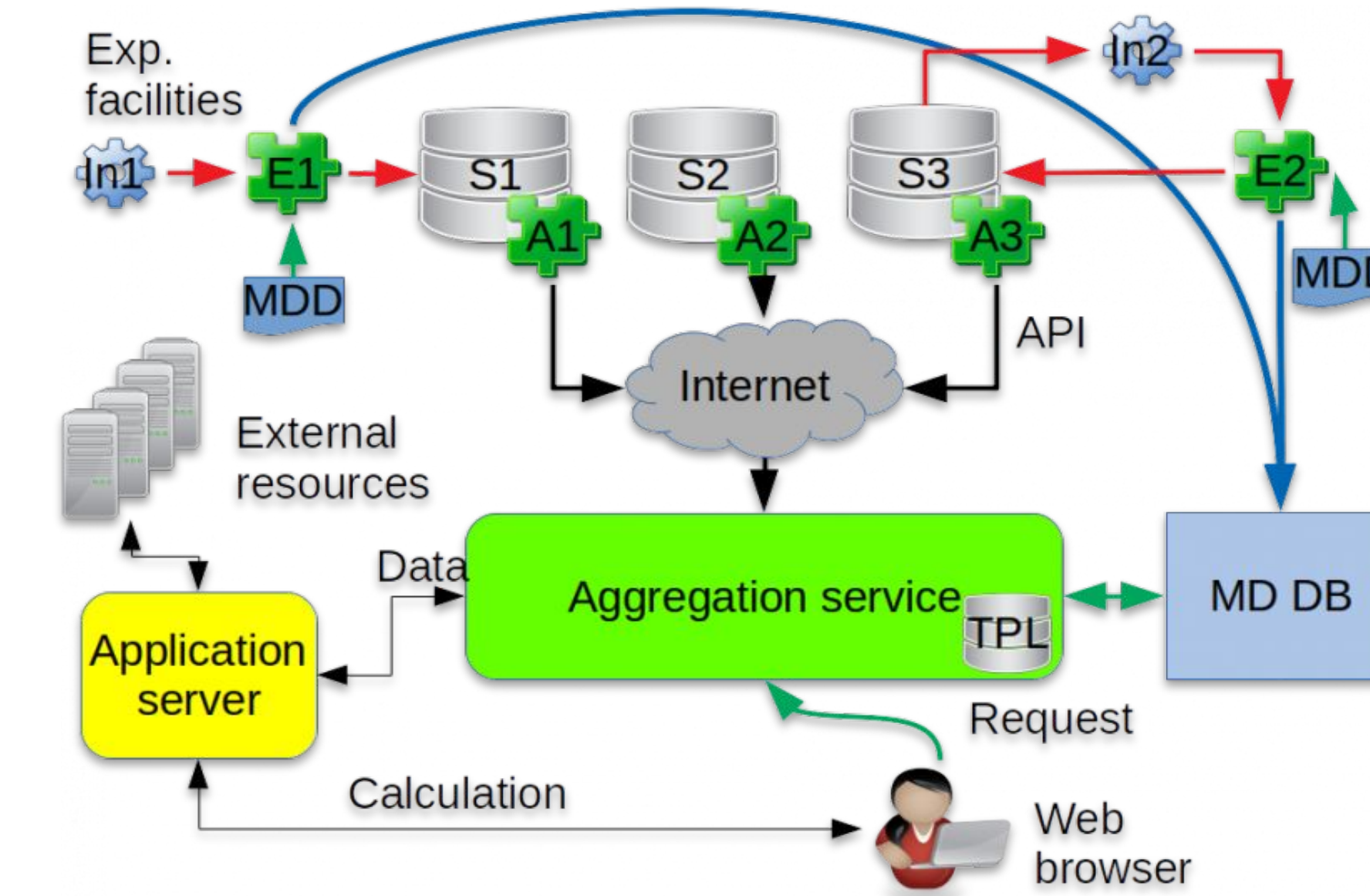[+]tokareva.victoria@kit.edu

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 of analysis and data center for astroparticle physics

- Si - local data storages;
- Ini - data sources of different types;
- MDD - metadata description;
- Ei - metadata extractors;
- Ai - adapters, provide API for data access;
- TPL - template library;
- MD DB - Metadata database;

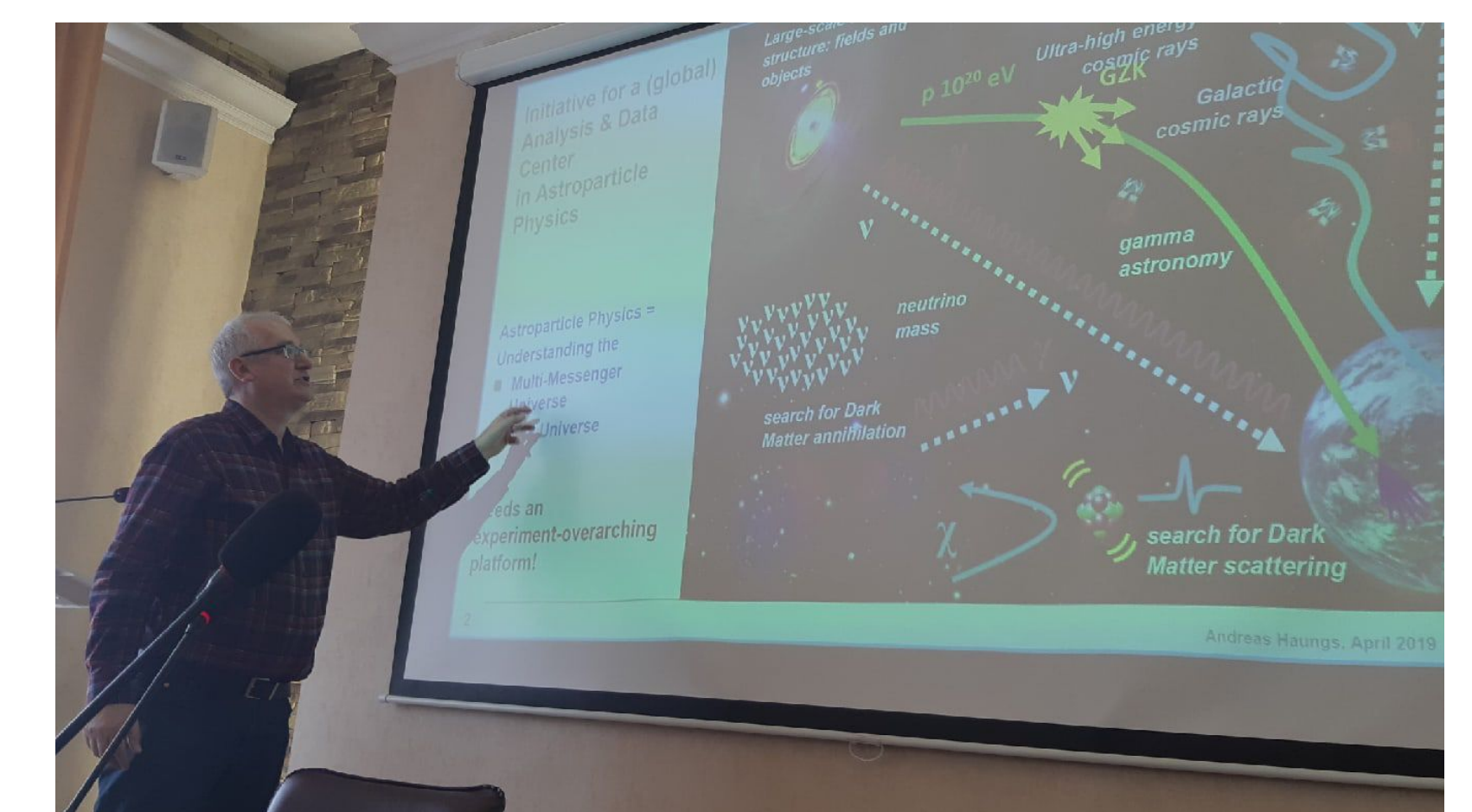## Machine learning for data analysis

- 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
- Machine learning methods: Decision Trees, Random Forest, Convolutional Neural Networks
- Frameworks: sklearn, TensowFlow, pyTorch

See also: PoS (ICRC 21) 319

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
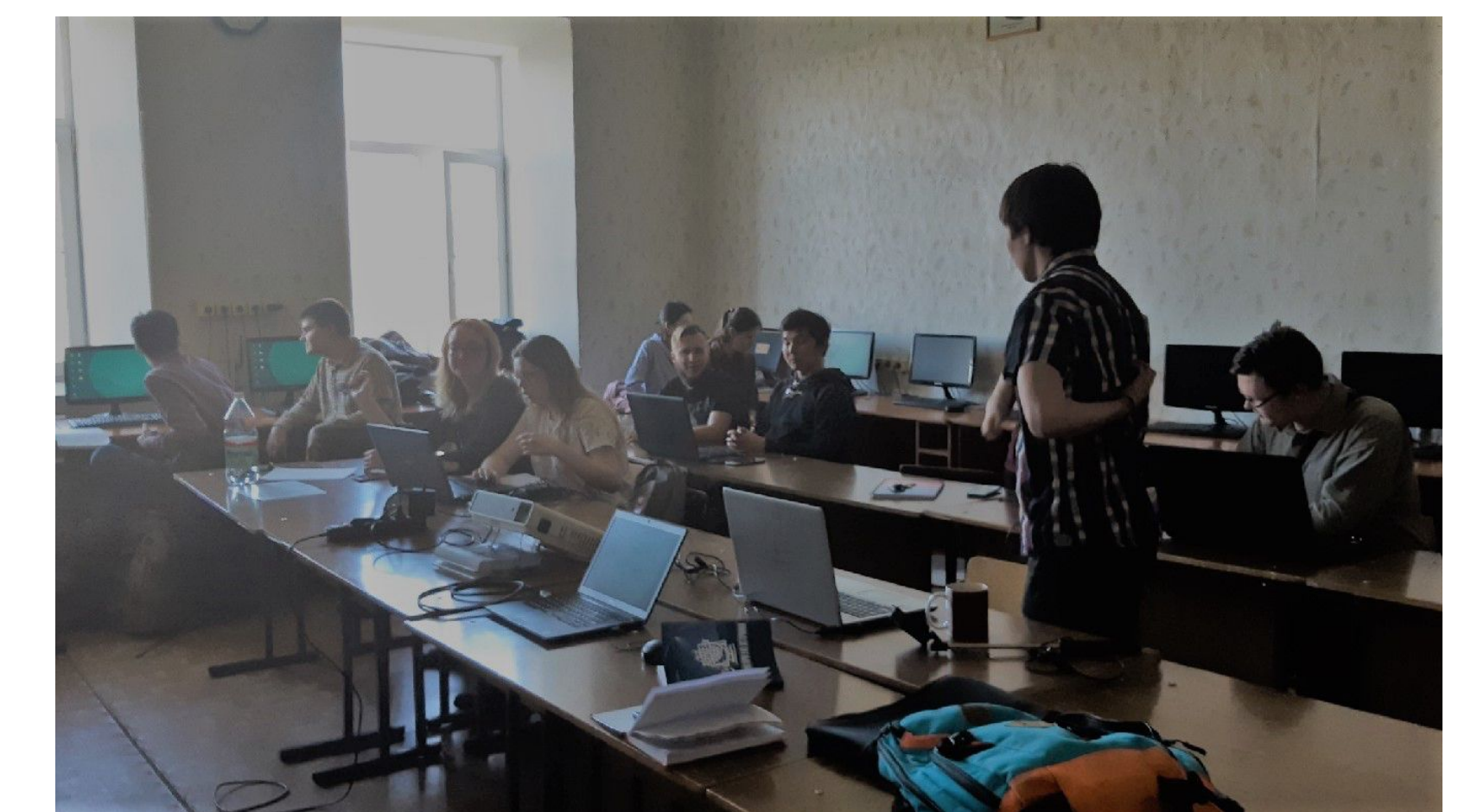
Tutorial on data analysis, ISU, 2018

KASCADE Cosmic-Ray Data Center 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 'Maket-Ani' experiment
  - introduce KCDC API
  - add the 100th (!) spectrum

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

See also:
PoS (ICRC 21) 422

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
- Python 3.8, C++

Datasets

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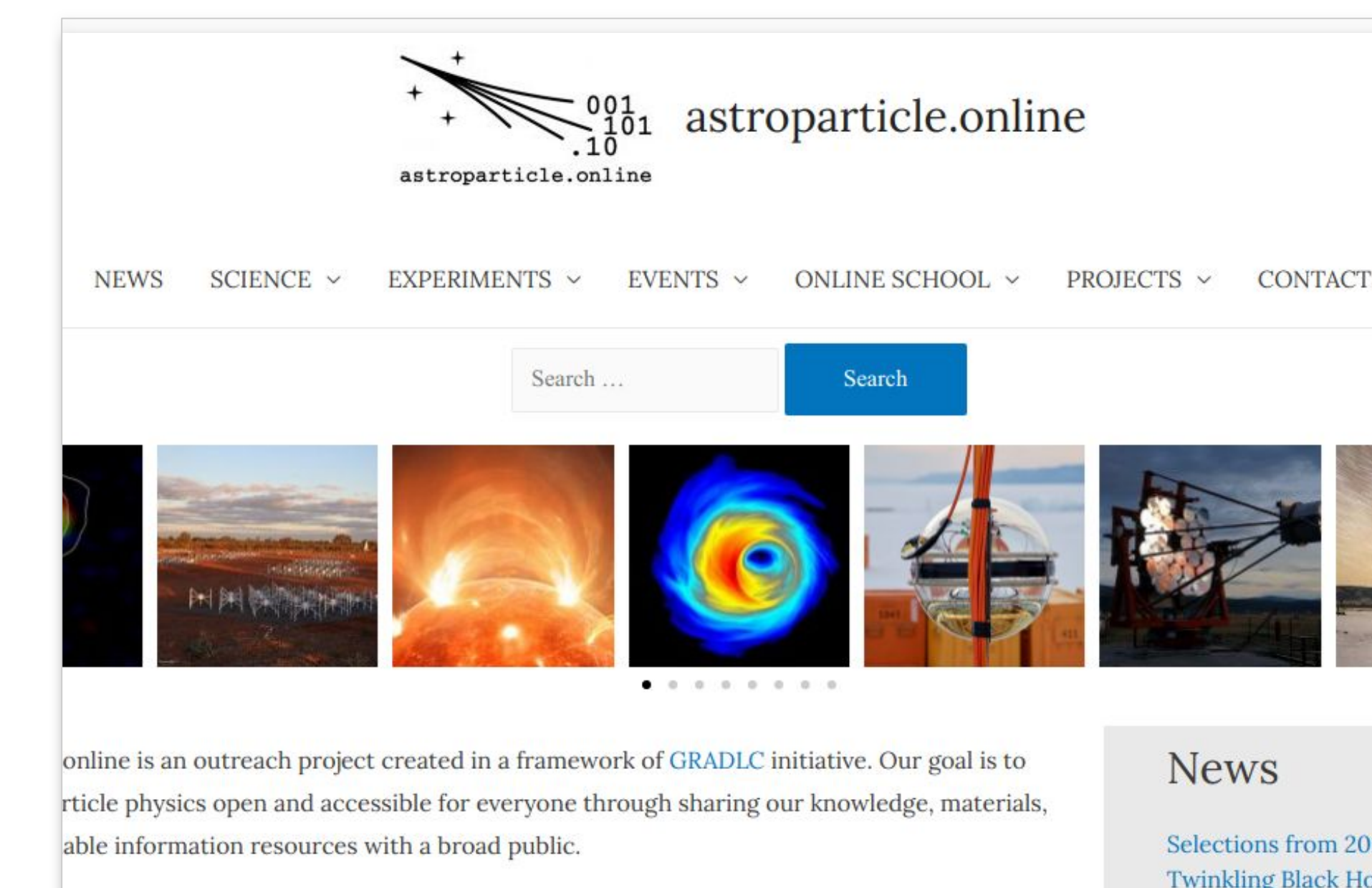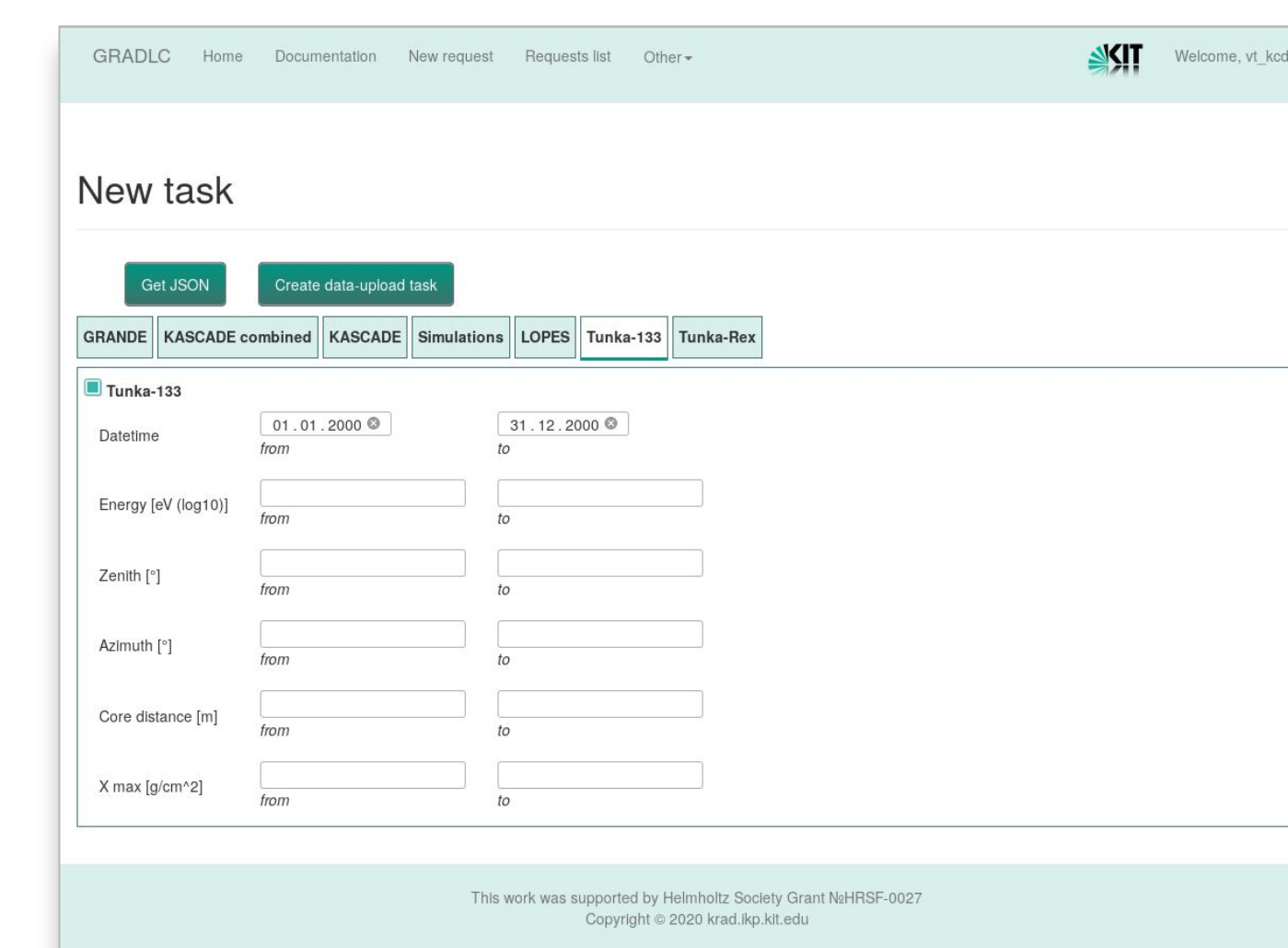

Main page of the astroparticle.online website

See also:
PoS (ICRC 21) 1378, PoS (ICRC 21) 1373

Dr. A. Haungs gives a lecture on multi-messenger astronomy for broad audience, ISU, 2019

JupyterHub for application server

- Login with KCDC or GRALDCI credentials
- Administration using Docker Swarm
- https://jupyter.iap.kit.edu/

jupyterhub

Web API

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

Request states
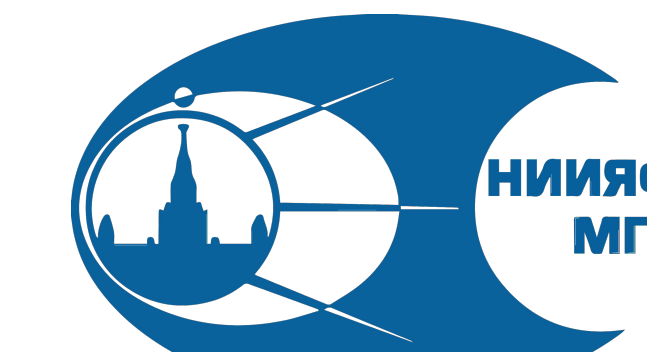- Running
- Scheduled
- Finished
- Failed
- Deleting
- Expired

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*

Acknowledgements

KASCADE - Grande
KArlsruhe Shower Core and Array DEtector - Grande

TAIGA