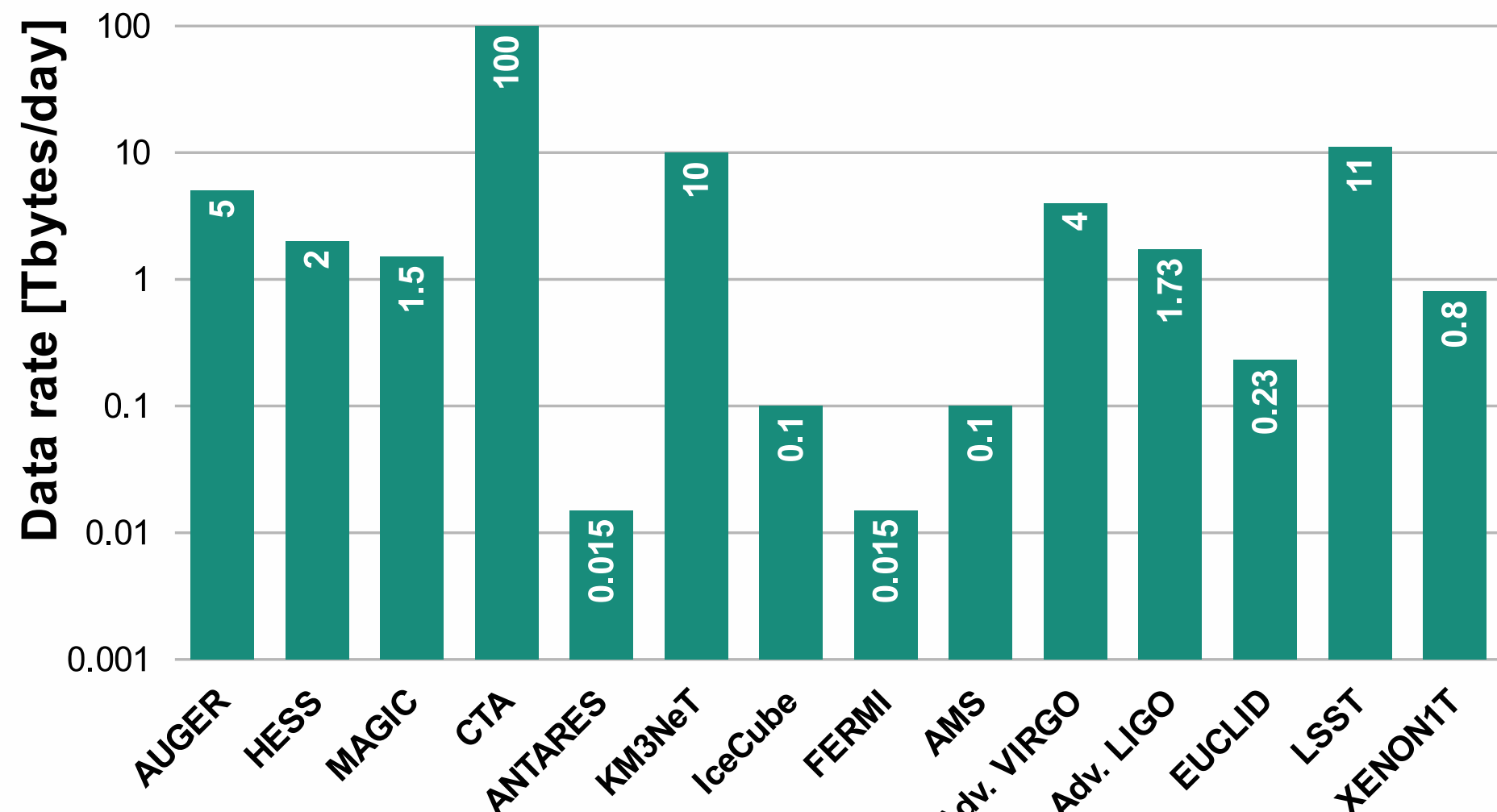# KIT
Karlsruhe Institute of Technology
Institut für Kernphysik (IKP)

mailto: victoria.tokareva@kit.edu

Setting up technical requirements for an astroparticle data life cycle

Victoria Tokareva for the GRADLCI Consortium
ACAT19, 11−15.03.19, Saas Fee, Switzerland

Big Data in astroparticle physics (APP)

Data rate [Tbytes/day] — AUGER 5, HESS 1.5, MAGIC 1.5, CTA 100, ANTARES 0.015, KM3NeT 10, IceCube 0.1, FERMI 0.015, AMS 0.1, Adv. VIRGO 4, Adv. LIGO 1.73, EUCLID 0.23, LSST 11, XENON1T 0.8

Data rates for modern astroparticle experiments [Tbytes/day][†]

- Wide range of experiments;
- More than hundred years of experiments to measure cosmic ray particles;
- Looking at the same sky with different detectors;
- Common data rate for astroparticle physics experiments all together is a few PBytes/year, which is comparable to the current LHC output[†];
- Big Data suitable for deep learning applications.

[†] Berghöfer T., Agrafoti I. et al., *Towards a model for computing in European astroparticle physics*, Astroparticle Physics European Coordination Committee, 2016

Requirements for the data storage

- Several experiments (TAIGA, KASCADE, etc.);
- Remote access to query results as local file systems;
- On-demand data transfer;
- Automatic real-time updates;
- No changes to the existing site infrastructure, only add-ons;
- Smart data mining.

KASCADE-Grande

- 1989−2013, most of the data available on KCDC web portal;
- Aimed at studying high-energy (galactic) cosmic rays by observing extensive air showers (EAS);
- Consisted of: 252 KASCADE scintillator detector stations; 37 Grande scintillator stations; hadronic calorimeter; digital radio antenna array LOPES.

TAIGA facility

Tunka-133
- 165 optical modules
- air-Cherenkov light

Tunka-Rex
- 63 radio antennas
- EAS radio emission

Tunka-Grande
- 380 scintillator detectors

Tunka-HiSCORE
- 47 low-threshold optical modules
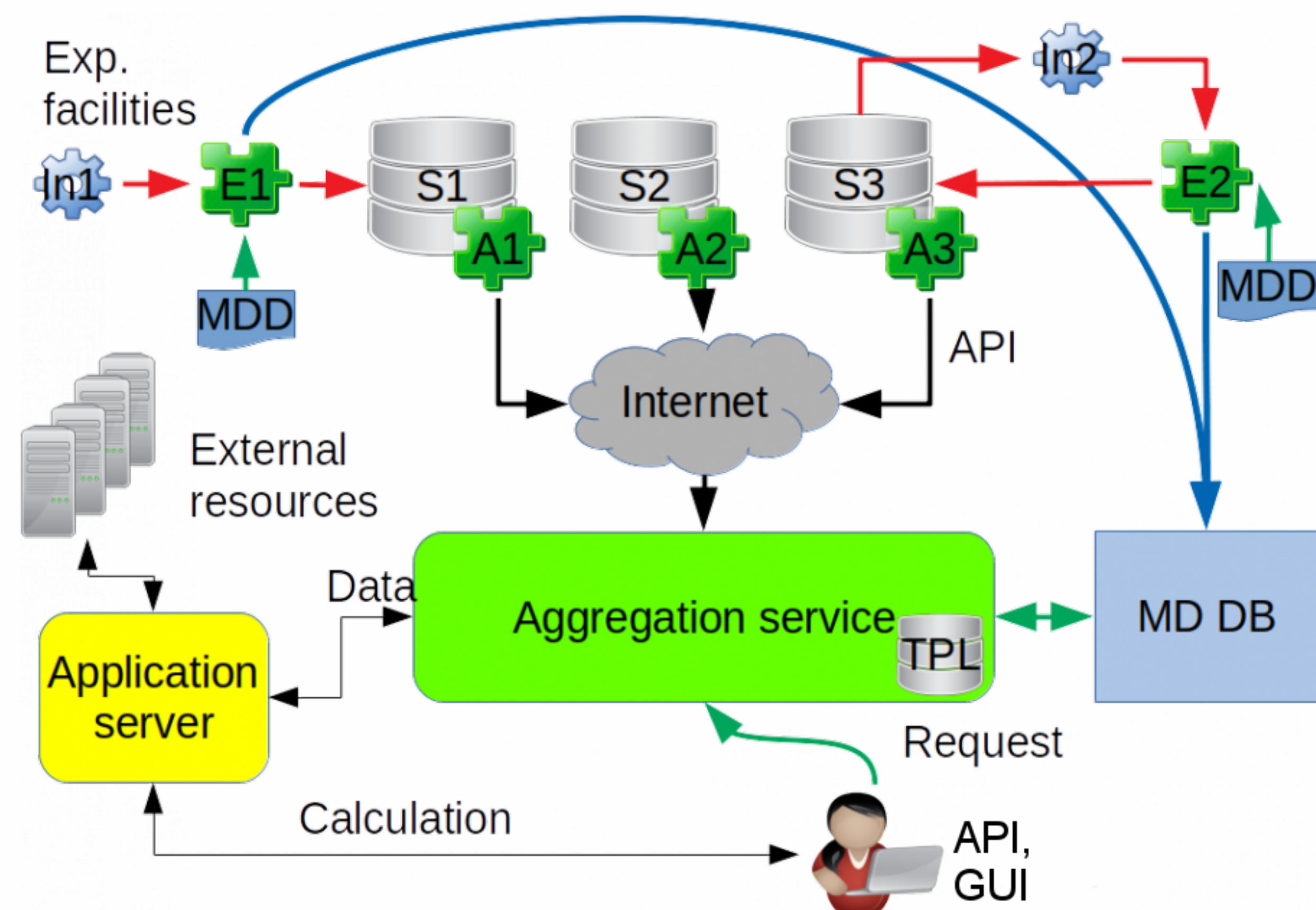
Tunka-IACT
- 2 imaging air cherenkov telescopes
- is being extended

KASCADE + TAIGA data rates

- current TAIGA rates:
  - ~ 50 TB of raw data;
  - ~ 8 TB/year of reconstructed data:
    - HiSCORE: 6.5 TB/year
    - IACT: 1 TB/year
    - others: 0.5 TB/year
- planned TAIGA rate:
  - ~ 20 TB/year of reconstructed data:
    - HiSCORE: ~ 18 TB/year
    - IACT: 1.5 TB/year
    - others: 0.5 TB/year
- KASCADE:
  - 450 000 000 events
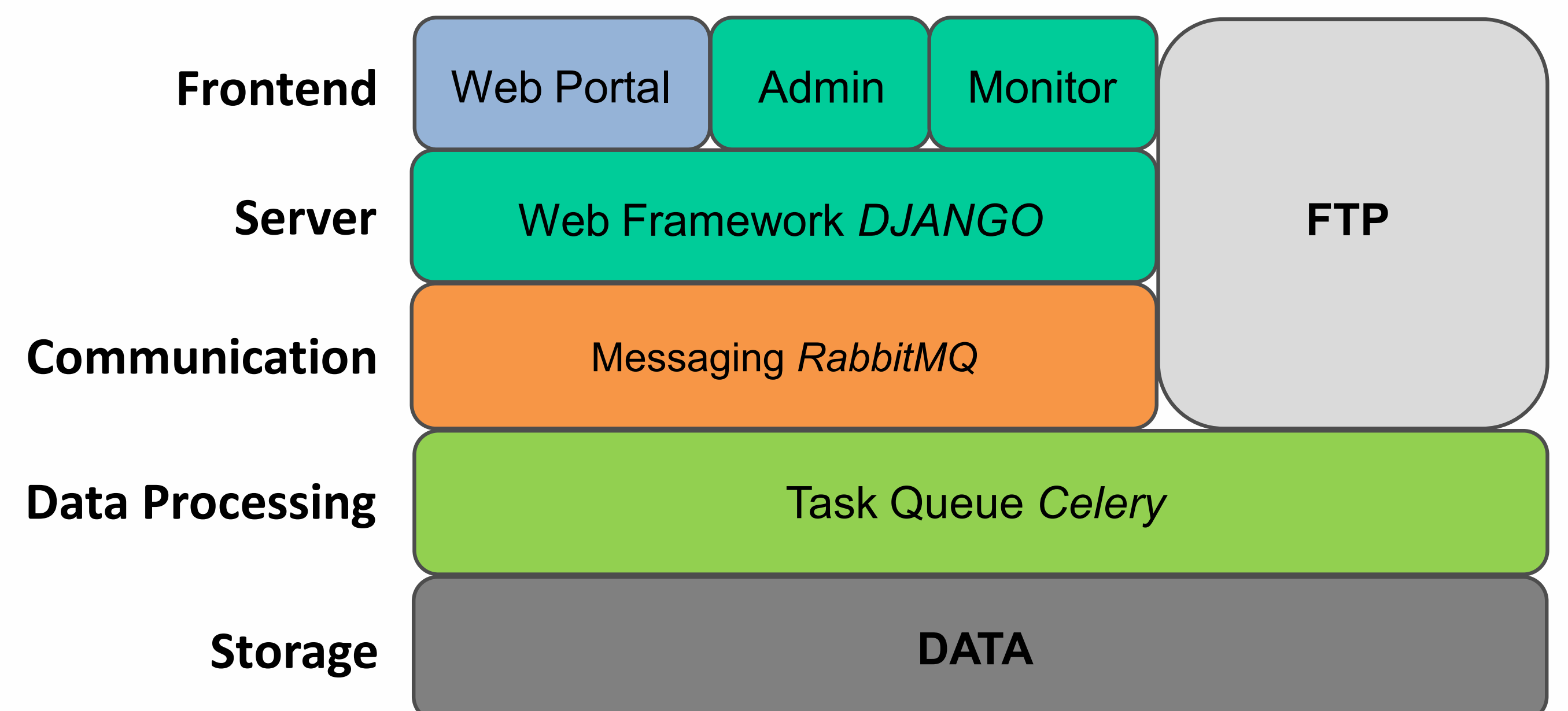 - ~ 4 TB of reconstructed data

Distributed architecture

Exp. facilities — In1, In2, E1, E2, S1, S2, S3, A1, A2, A3, MDD, API, Internet, External resources, Application server, Aggregation service, TPL, MD DB, Data, Request, Calculation, API, GUI

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.

Possible solutions: CVMFS, PostgreSQL, TPL, HTCondor.

KASCADE Cosmic Ray Data Centre (KCDC)

Frontend	Web Portal	Admin	Monitor
Server	Web Framework *DJANGO*		FTP
Communication	Messaging *RabbitMQ*		
Data Processing	Task Queue *Celery*		
Storage	DATA		

- Providing free, unlimited, reliable open access to KASCADE cosmic ray data at https://kcdc.ikp.kit.edu;
- Almost all KASCADE data is available;
- Selection of fully calibrated quantities and detector signals;
- Information platform: physics and experiment backgrounds, tutorials, meta information for data analysis;
- Archive of KASCADE software and data;
- Uses modern and open source web technologies.
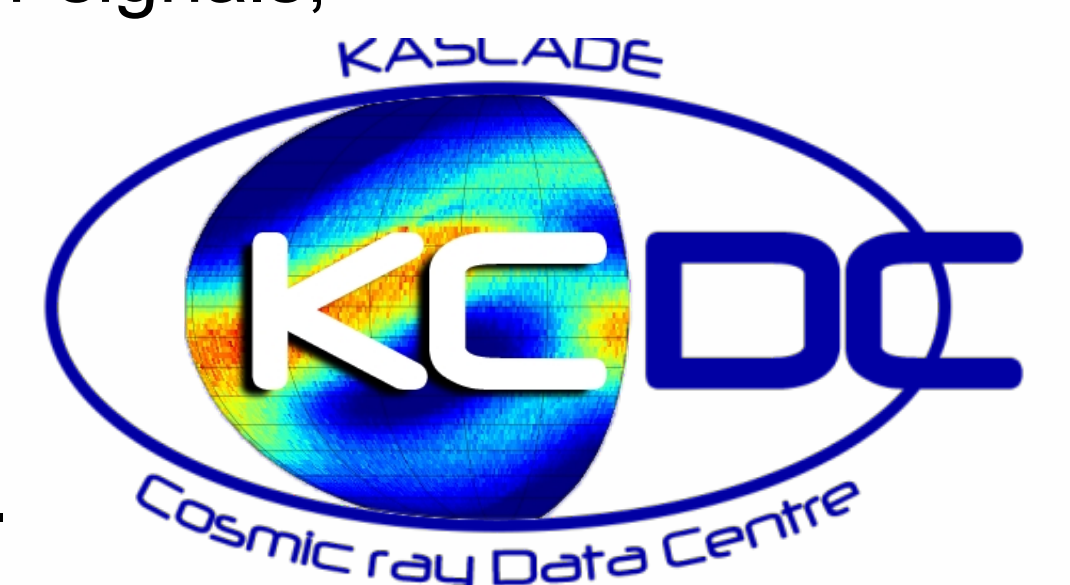
German-Russian Astroparticle Data Life Cycle Initiative

KASCADE - Grande
KArlsruhe Shower Core and Array DEtector - Grande

KIT — Karlsruher Institut für Technologie

TAIGA

SCC — Steinbuch Centre for Computing

Matrosov Institute for System Dynamics and Control Theory

НИИЯФ МГУ

astroparticle.online

001
101
.10

Conclusion

- The Astroparticle Data Life Cycle ecosystem is in development by the GRADLC Consortium, based on the existing data portal KCDC;
- The first stage of development includes building aggregation data server and performing the proof-of-principle joint analysis of data from KASCADE and TAIGA experiments;
- Possible solutions were offered based on the data rates and specificities of astroparticle physics;
- Most of the KASCADE-Grande data are available on the KCDC web-portal https://kcdc.ikp.kit.edu.

See also: I. Bychkov et. al. Russian–German Astroparticle Data Life Cycle Initiative. Data 2018, 3(4), 56.