



APPDS

Astroparticle Physics Data Storage (APPDS) is a project for design and development of distributed data storage for cosmic ray experiments such as [TAIGA](#) and [KASCADE](#). This project has financial support from [RSF](#) and [Helmholtz society](#) (grant No 18-41-06003).

Annotation

The project will address the problem of management of contemporary scientific data. Nowadays the exponential growth of the amount of experimental data can be observed. While there was 1-10 Tb of data per year in astrophysics 10-15 years ago, new experimental facilities generate data sets ranging in size from 100's to 1000's of terabytes per year. It can be illustrated by a growth of the amount of data acquired by satellites. While the Integral satellite [<http://sci.esa.int/integral>] downloaded to the ground 1.2 Gb of data per day in 2002, now the Gaia spacecraft (2013) transfers about 5 Gb of data per day. The other example is the ground-based experiment LSST [<https://www.lsst.org>], providing over 3 gigapixels per image with an exposure every 15 seconds. It is expected to take ~10 petabytes of information per year.

These trends give rise to a number of emerging issues of big data management. It's obvious that various activities must be performed continually across all stages of the data life cycle to help support effective data management (see an example in <https://pubs.usgs.gov/of/2013/1265/pdf/of2013-1265.pdf>): the collection and storage of data, its processing and analysis, refining the physical model, making preparations for publication, and data reprocessing taking refinement into account. An important topic for modern science in general and astroparticle physics in particular is open science, the model of free access to data (e.g. Paul A. David, *Industrial and Corporate Change*, Volume 13, Number 4, pp. 571-589; <http://ec.europa.eu/research/openscience/index.cfm>): data are accessible not solely to collaboration members but to all levels of an inquiring society, amateur or professional. This approach is especially important in the age of Big Data, when a complete analysis of the experimental data cannot be performed within one collaboration.

The present project will strive to develop an open science system to be able to collect, store, and analyze astrophysical data having the TAIGA [<http://taiga-experiment.info>] and KASCADE [<https://web.ikp.kit.edu/KASCADE>] experiments as the examples. The novelty of the proposed approach can be seen in developing integrated solutions including:

- development and adaptation of distributed data storage algorithms and techniques with a common meta-catalog to provide a common information space of the distributed repository;
- development and adaptation of data transmission algorithms as well as simultaneous data transmission from several data repositories thus significantly reducing load time;
- development of machine-learning techniques for identifying mass groups of particles and their properties in a fully remote access mode;
- installation of the KCDC-based prototype system of Big Data analysis and exporting the experimental data from KASCADE and TAIGA for testing technology of data life cycle management.

We will also create an educational system on the HubZero platform [<https://hubzero.org>] dedicated to astroparticle physics.

This is an innovative approach that will be used in astroparticle physics research for the first time. Plans are underway to expand the number of experiments by exporting data from other scientific collaborations, it will rapidly advance the research of fundamental properties of matter and the universe. It's noteworthy that the suggested approach can be used not only in the specified field of science but also adapted to other scientific disciplines.

Collaboration

- Russian institutes:
 - SINP MSU, Moscow
 - IGU, Irkutsk
 - IDSTU, Irkutsk
- Germany institutes:
 - KIT, Karlsruhe

Principal members

- Germany principal members
 - [Andreas Haungs](#), KIT - group leader. [Home page](#)
 - [Dmitry Kostunin](#), KIT
- Russian principal members
 - [Alexander Kryukov](#), SINP MSU - group leader
 - [Igor Bychkov](#), IDSTU SB RAS
 - [Elena Korosteleva](#), SINP MSU
 - [Julia Kazarina](#), ISU

Mailing list: [APPDS](#)

Project structure

WP's

- [WP1](#) (KCDC extension): software extension of KCDC to include Tunka-related data
- [WP2](#) (Big Data Science Software): Run first jobs for KCDC or Tunka on a Tier/Grid environment (i.e. concept for Docker, Container, etc...)
- [WP3](#) (Multi-messenger Data Analysis): find/select 2 PhDs and define appropriate physics topics.
- [WP4](#) (go for public): define/create concept and design for KRAD webpages

Schedule and Milestones

Considering the above mentioned specific tasks, we defined the schedule and a couple of milestones to be reached within this project:

2018

1. Development of standard astrophysical data format and preparation of the TAIGA data (wp1)
2. Preparing the environment to move KCDC to the computing facilities in KIT and MSU (wp2)
3. Workout of a coherent concept for a dedicated Data Life Cycle Lab (wp2)
4. Defining physics tasks for common data analysis (wp3)
5. Creating webpage and announcing KRAD to the world (wp4)

2019

1. Inclusion of TAIGA data within KCDC (wp1)
2. Final installation of extended KCDC at large-scale computing facilities in KIT and MSU (wp2)
3. Installing the data life cycle lab (wp2)
4. Performing multi-messenger analysis using the extended KCDC (wp3)
5. Publishing of the Data Life Cycle Lab (wp4)

2020

1. Generalization of the Data Life Cycle Lab (wp2)
2. Workout of a concept for global astroparticle physics data centre (wp2)
3. Finalizing and publication of specific data analyses using the new environment (wp3)

From:

<https://theory.npi.msu.su/dokuwiki/> - **THEORY**

Permanent link:

<https://theory.npi.msu.su/dokuwiki/doku.php/appds/main>

Last update: **08/11/2018 01:20**

