



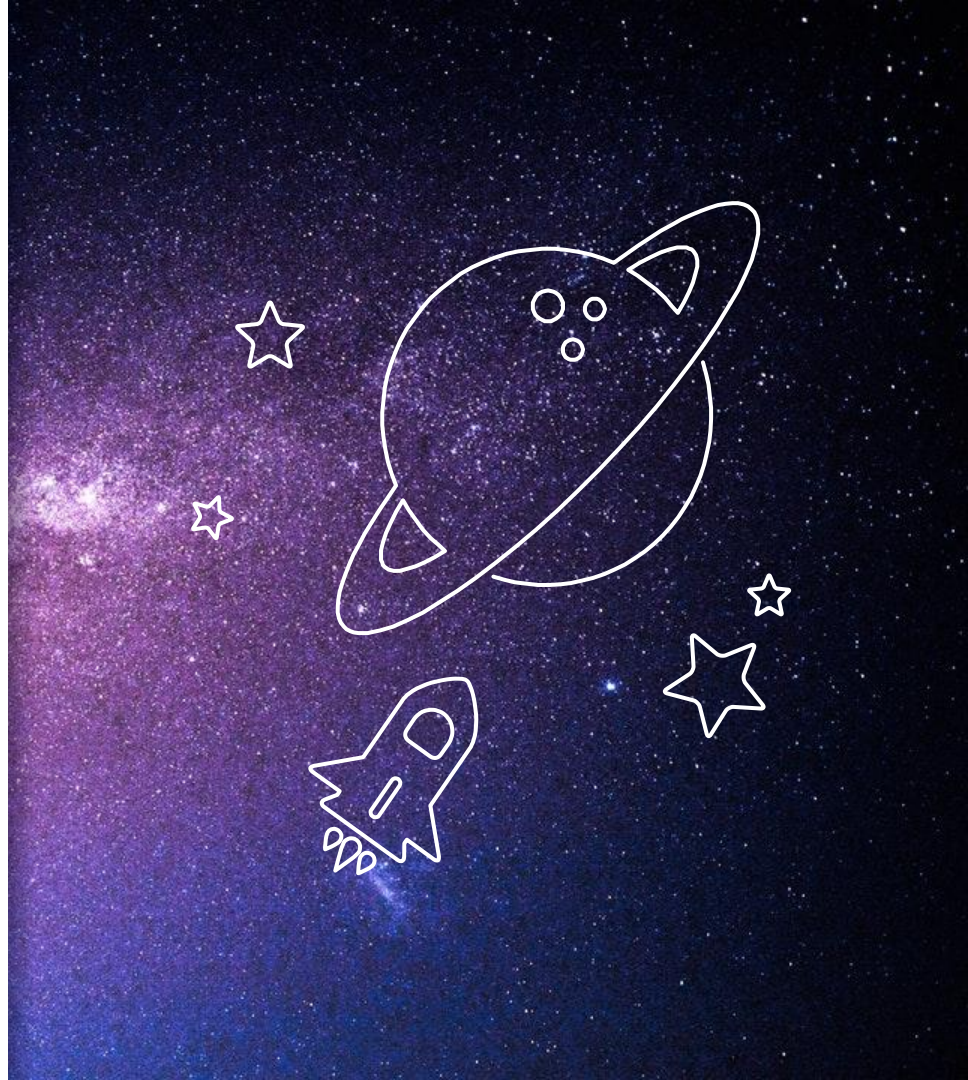
KCDC overview for German–Russian Astroparticle Data Life Cycle Initiative

Victoria Tokareva (IKP KIT)

GRAD/APPDS f2f meeting, Karlsruhe, 2018

Contents:

- ★ Introduction
- ★ Tutorial
- ★ Technologies
- ★ Structure
- ★ Data Workflow
- ★ Future Plans/Discussion



The KASCADE Cosmic-ray Data Centre (KCDC)



- ★ Web portal, there the data of KASCADE-Grande are made available for the interested public;
- ★ <https://kcdc.ikp.kit.edu>;
- ★ Data for more than 20 years of accumulation;
- ★ More than 433 million air showers;
- ★ Simulations for three different high energy interaction models;
- ★ In run since 2013.





KCDC tutorial

Exploring KCDC Data Shop

KCDC Homepage

KCDC Motivation

KCDC Regulations

► Information

► Announcements

FAQs

► User Account

▼ Data Shop

New Request

Review Requests

Preselections

► Simulations

Spectra

Publications

► Report a Bug

Education/Lehre

Welcome to KCDC

The aim of the project **KCDC** (**KASCADE Cosmic Ray Data Centre**) is the installation and establishment of a public data centre for high-energy astroparticle physics based on the data of the KASCADE experiment. KASCADE was a very successful large detector array which recorded data during more than 20 years on site of the KIT-Campus North, Karlsruhe, Germany (formerly Forschungszentrum, Karlsruhe) at 49,1°N, 8,4°E; 110m a.s.l. KASCADE collected within its lifetime more than 1.7 billion events of which some 433.000.000 survived all quality cuts and are made available here for public usage.



KASCADE
Karlsruhe Shower Core
and Array Detector

been published !!! For details check the 'Developer News' page +++

Institute for Nuclear Physics
(IKP)
KIT Campus North

Address:
Institute for Nuclear Physics
Karlsruhe Institute of Technology
Hermann-v.Helmholtz-Platz 1
D-76344 Eggenstein-
Leopoldshafen

Postal Address:
Institute for Nuclear Physics
Karlsruhe Institute of Technology
Postbox 3640
D-76021 Karlsruhe

Phone: +49/721/608-23546
Fax: +49/721/608-23548

E-Mail:
[ikp-kcdc\[at\]lists.kit.edu](mailto:ikp-kcdc[at]lists.kit.edu)

Downloads
[KCDC Manual \(en\)](#)
[KCDC Simulations Manual \(en\)](#)

KCDC OPEN -BETA - VERSION NABOO 2.2 BASED ON: KAOS (1.0.0)

To get access to the KCDC DataShop, you need to:

- ★ Register;
- ★ Login with your user account at

<https://kcdc.ikp.kit.edu>



KASCADE Cosmic Ray Data Centre (KCDC) / Open β

KCDC Homepage

KCDC Motivation

KCDC Regulations

► Information

► Announcements

FAQs

► User Account

► Data Shop

New Request

Review Requests

Preselections

► Simulations

Spectra

Publications

► Report a Bug

Education/Lehre

KCDC Data Shop

Components Available	Components Selected	Quantities and Cuts
<div>KASCADE</div> <div>Calorimeter</div> <div>GRANDE</div>	<div>General Info</div>	<input type="checkbox"/> Toggle all Please select a component on the left by adding it to the selected components. If you want to set the quantities, please click on its name afterwards.

Verify & Submit Request

Hadron Calorimeter Info

Event data of the KASCADE Central Hadron Detector like reconstructed energy sum of all hadrons and the number of hadrons reconstructed.

Select the detector component and the quantities

KCDC Homepage

KCDC Motivation

KCDC Regulations

► Information

► Announcements

FAQs

► User Account

► Data Shop

New Request

Review Requests

Preselections

► Simulations

Spectra

Publications

► Report a Bug

Education/Lehre

KCDC Data Shop

Components Available	Components Selected	Quantities and Cuts	
Calorimeter	General Info		
GRANDE	KASCADE		
		<input type="checkbox"/> Toggle all	
		Air Temperature range: -20 to 50 °C Add Cut	
		Air Pressure range: 960 to 1040 hPa Add Cut	
		<input checked="" type="checkbox"/> DateTime range: 1998-05-08 to 2013-01-15 Add Cut <div>2000-01-01 to 2001-01-01</div>	
		<input type="checkbox"/> Global Time range: 8.946e+8 to 1.358e+9 sec Add Cut	
		<input type="checkbox"/> Mt range: 0 to 9.99e+8 ns	
		<input type="checkbox"/> Run Number range: 877 to 7417 Add Cut	
		<input type="checkbox"/> Event Number range: 1 to 3e+6	
		<input type="checkbox"/> e/y E-Deposit range: 0 to 3e+4 MeV	
		<input type="checkbox"/> μ E-Deposit range: 0 to 1000 MeV	
		<input type="checkbox"/> Arrival Times range: -1550 to 2550 ns	
		<input type="checkbox"/> Grande Deposit range: 0 to 1e+5 MeV	
		<input type="checkbox"/> Grande Arrival range: 1000 to 1e+4 ns	

Verify & Submit Request

Welcome to the Datashop

On the left hand site you may select available detector components. Hovering the mouse over such a component, will give you some information on it. Once selected, you may click on the components name to view and select quantities associated with that detector. You may also deselect components using the left arrow. In right most column, you may select quantities for shipping and add cuts, that will be used to select only events passing these. On verification: Yellow means it has been corrected, red means you have to adjust your input. Is it a valid number? Is the lower bound larger than the upper bound?

[\[details -> KCDC Manual\]](#)

Applying data cuts

KCDC Homepage

KCDC Motivation

KCDC Regulations

► Information

► Announcements

FAQs

► User Account

► Data Shop

New Request

Review Requests

Preselections

► Simulations

Spectra

Publications

► Report a Bug

Education/Lehre

KCDC Data Shop

Components Available	Components Selected	Quantities and Cuts	
Calorimeter	General Info		
GRANDE	KASCADE		
		<input type="checkbox"/> Toggle all	
		Energy range: 13 to 19 eV [log10] <input type="button" value="Add Cut"/>	
		<input checked="" type="checkbox"/> X Core Position range: -91 to 91 m	<input type="text" value="14"/> to <input type="text" value="18"/> eV [log10] <input type="button" value="Add Cut"/>
		<input type="checkbox"/> Y Core Position range: -91 to 91 m	<input type="button" value="Add Cut"/>
		<input checked="" type="checkbox"/> Zenith Angle range: 0 to 60 °	<input type="text" value="18"/> to <input type="text" value="24"/> ° <input type="button" value="Add Cut"/>
		<input type="checkbox"/> Azimuth Angle range: 0 to 360 °	<input type="button" value="Add Cut"/>
		<input checked="" type="checkbox"/> Electron Number range: 2 to 8.7 [log10]	<input type="text" value="4"/> to <input type="text" value="8.7"/> [log10] <input type="button" value="Add Cut"/>
		<input checked="" type="checkbox"/> Muon Number range: 2 to 7.7 [log10]	<input type="button" value="Add Cut"/>
		<input type="checkbox"/> Shower Age range: 0.1 to 1.48	<input type="text" value="2.5"/> to <input type="text" value="7.7"/> [log10] <input type="button" value="Add Cut"/>

Verify & Submit Request

Welcome to the Datashop

On the left hand site you may select available detector components. Hovering the mouse over such a component, will give you some information on it. Once selected, you may click on the components name to view and select quantities associated with that detector. You may also deselect components using the left arrow. In right most column, you may select quantities for shipping and add cuts, that will be used to select only events passing these. On verification: Yellow means it has been corrected, red means you have to adjust your input. Is it a valid number? Is the lower bound larger than the upper bound?

[details -> KCDC Manual]

Applying data cuts

KCDC Homepage
KCDC Motivation
KCDC Regulations

Information

Announcements

FAQs

User Account

Data Shop

New Request

Review Requests

Preselections

Simulations

Spectra

Publications

Report a Bug

Education/Lehre

KCDC Data Shop

Check your selections and submit request

KASCADE

Energy	range:	14	to	18	eV [log10]	user cut
Zenith Angle	range:	18	to	27	°	user cut
Electron Number	range:	4	to	8.7	[log10]	user cut
Muon Number	range:	2.5	to	7.7	[log10]	user cut

General Info

DateTime	range:	2000-01-01	to	2001-01-01	user cut
Run Number	range:	877	to	7417	full range
Event Number	range:	1	to	3e+6	full range

Output Format: ☒ :ROOT ☐ :HDF5 ☐ :ASCII

Back

Submit

Welcome to the Datashop

On the left hand site you may select available detector components. Hovering the mouse over such a component, will give you some information on it. Once selected, you may click on the components name to view and select quantities associated with that detector. You may also deselect components using the left arrow. In right most column, you may select quantities for shipping and add cuts, that will be used to select only events passing these. On verification: Yellow means it has been corrected, red means you have to adjust your input. Is it a valid number? Is the lower bound larger than the upper bound?

[details -> KCDC Manual]

Overview on the selected quantities and cuts



KASCADE Cosmic Ray Data Centre (KCDC) / Open β



User Review Page

Your last requests were:

Submitted: **2018-10-29 13:01:14 UTC**, Data Format: **ROOT**, Status: **PENDING**

Details

Resubmit

Cancel

Submitted: **2018-08-17 10:31:58 UTC**, Data Format: **ROOT**, Status: **DLEXPIRED**

Details

Resubmit

Delete

User Review Page

This page holds all jobs submitted by the user.

The **'Details'** button gives a list of detector components, quantities and cuts applied as well as the status information of the job.

To resubmit the job with the same or with different cuts press **'Resubmit'**. To cancel a running job indicated by the status **'STARTED'** press **'Cancel'**. To delete a job which has been successfully processed press **'Delete'**.

To download a processed data set via ftp press **'Download'**. If the status is set to **'DLEXPIRED'** the **'Download'** button will be vanished, no further access to your data is possible.

[details -> KCDC Manual]

Your requests overview



KASCADE Cosmic Ray Data Centre (KCDC) / Open β

KCDC Homepage

KCDC Motivation

KCDC Regulations

Information

Announcements

FAQs

User Account

Data Shop

New Request

Review Requests

Preselections

Simulations

Spectra

Publications

Report a Bug

Education/Lehre

User Review Page

Your last requests were:

Submitted: **2018-10-29 13:01:14 UTC**, Data Format: **ROOT**, Status: **STARTED**

Hide

Resubmit

Cancel

You have selected the following parameters and cuts for download:

General Info

Event Number	range:	1.0 to	3000000.0	full range
Run Number	range:	877 to	7417	full range
DateTime	range:	2000-01-01 to	2001-01-01	user cut

KASCADE

Zenith Angle	range:	18 to	27 °	user cut
Electron Number	range:	4 to	8.7 [log10]	user cut
Energy	range:	14 to	18 eV [log10]	user cut
Muon Number	range:	2.5 to	7.7 [log10]	user cut

Submitted: **2018-08-17 10:31:58 UTC**, Data Format: **ROOT**, Status: **DLEXPIRED**

Details

Resubmit

Delete

User Review Page

This page holds all jobs submitted by the user. The 'Details' button gives a list of detector components, quantities and cuts applied as well as the status information of the job.

To resubmit the job with the same or with different cuts press 'Resubmit'. To cancel a running job indicated by the status 'STARTED' press 'Cancel'. To delete a job which has been successfully processed press 'Delete'. To download a processed data set via ftp press 'Download'. If the status is set to 'DLEXPIRED' the 'Download' button will be vanished, no further access to your data is possible.

[details -> KCDC Manual]

Your requests overview

Data Preselections for download

ASCII 'Data Preselections' for direct download

[ReducedData-KASCADE_runs_0877-4683_ASCII](#)

all KASCADE quantities in the run range 877-4683, no cuts applied,
no data arrays, data format **ascii**

Full data sample from the releases VULCAN and MERIDIAN size: 7.1 Gb [details](#)

[ReducedData-KASCADE_runs_4685-7417_ASCII](#)

all KASCADE quantities in the run range 4685-7417, no cuts applied,
no data arrays, data format **ascii**

size: 12.9 Gb [details](#)

[ReducedData-GRANDE_runs_4775-7398_ASCII](#)

all GRANDE quantities in the run range 4775-7398, no cuts applied,
no data arrays, data format **ascii**

size: 5.0 Gb [details](#)

[ReducedData-CALOR_runs_877-5496_ASCII](#)

all CALORIMETER and KASCADE quantities in the run range 877-5496,
no cuts applied,
no data arrays, data format **ascii**

size: 9.8 Gb [details](#)

[HighEnergyData_runs_0877-7417_ASCII](#)

Events with a reconstructed primary energy above $10^{15.7}$ eV
no data arrays, data format **ascii**

size: 179 MB [details](#)

[SmallDataSample_noDataArrays_runs_0877-7417_ASCII](#)

Data sample with every 400th event of the whole data set
included are all detector components, no data arrays, data format **ascii**

size: 64 MB [details](#)

Data Preselections

If you are interested in 'Preselections' this menu offers the option to download the data sample directly without selecting quantities and applying cuts in the DataShop.

'details' provides a more detailed information page of the respective data set. The 'Small Data Samples' offer the opportunity to check your own requirements on a small data sample.

To download click on the 'Set Name'.

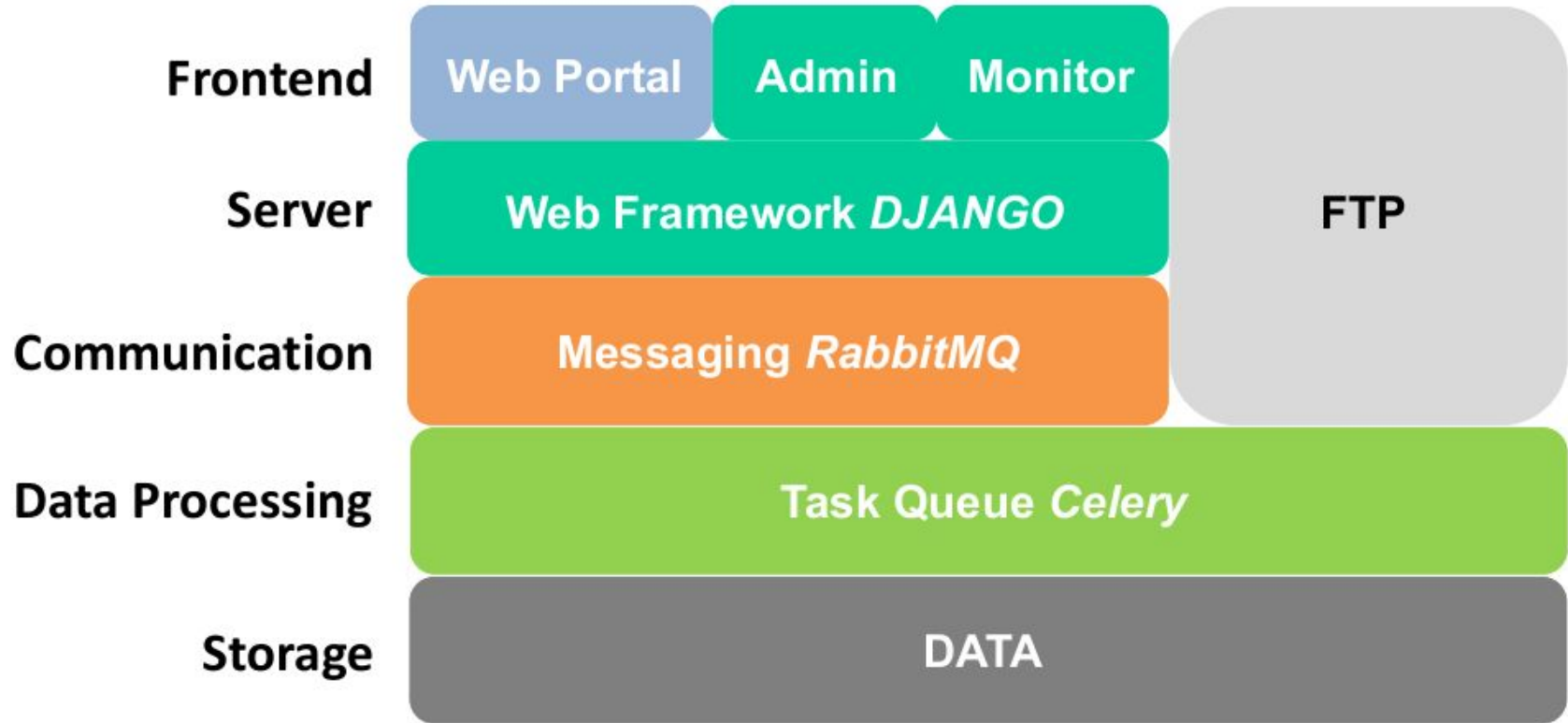
[details -> [KCDC-Manual](#)]

Data preselections



Technologies and structure

KCDC IT Structre



KAOS – Karlsruhe Astroparticle physics Open data Software



Has been written in the context of the KCDC, implemented using a plugin based design with a focus on easy extensibility and modifiability in order to work also outside the context of KCDC.

Consists of the following plugins:

- | | | |
|------------------|-----------------|-----------------|
| ★ kaos_bugreport | ★ kaos_email | ★ kaos_newsfaqs |
| ★ kaos_celery | ★ kaos_spectra | ★ kaos_papers |
| ★ kaos_datashop | ★ kaos_mainmenu | ★ kaos_user |

DJANGO



Django is a free and open-source web framework, written in Python, which follows the model-view-template (MVT) architectural pattern.

Django's primary goal is to ease the creation of complex, database-driven websites. Django emphasizes reusability and "pluggability" of components, less code, low coupling, rapid development, and the principle of don't repeat yourself. Python is used throughout, even for settings files and data models. Django also provides an optional administrative create, read, update and delete interface that is generated dynamically through introspection and configured via admin models.

* <https://www.djangoproject.com/>

RabbitMQ



RabbitMQ is an open source message broker software (sometimes called message-oriented middleware) supporting the Advanced Message Queuing Protocol (AMQP), Streaming Text Oriented Messaging Protocol (STOMP), Message Queuing Telemetry Transport (MQTT), and other protocols.

The RabbitMQ server program is written in the Erlang programming language and is built on the Open Telecom Platform framework for clustering and failover. Client libraries to interface with the broker are available for all major programming languages.

[*<http://www.rabbitmq.com>](http://www.rabbitmq.com)

Celery



Celery is an asynchronous task queue/job queue based on distributed message passing. It is focused on real-time operation, but supports scheduling as well.

The execution units, called tasks, are executed concurrently on a single or more worker servers using multiprocessing, *eventlet*, or *gevent*. Tasks can execute asynchronously (in the background) or synchronously (wait until ready).

Celery is written in Python, but the protocol can be implemented in any language. It can also operate with other languages using *webhooks*. There is also a Ruby client, a PHP client, a Go client, and a Node.js client.

*<http://www.celeryproject.org/>

MongoDB



MongoDB is a free and open-source cross-platform document-oriented database program. Classified as a NoSQL database program, MongoDB uses JSON-like documents with schemata. MongoDB is developed by MongoDB Inc., and is published under a combination of the Server Side Public License and the Apache License.

MongoDB supports field, range query, and regular expression searches. Fields in a MongoDB document can be indexed with primary and secondary indices. MongoDB provides high availability with replica sets. MongoDB scales horizontally using sharding.

*<https://www.mongodb.com>

SQL vs. NoSQL

SQL

- ★ ACID (Atomicity, Consistency, Isolation, Durability);
- ★ Automatic handling of interlocks, collisions, and data consistency;
- ★ The structure is known from the outset and is relatively stable;
- ★ Data types are determined in advance;
- ★ Preferable for centralized storage: usually a better performance for a single node.

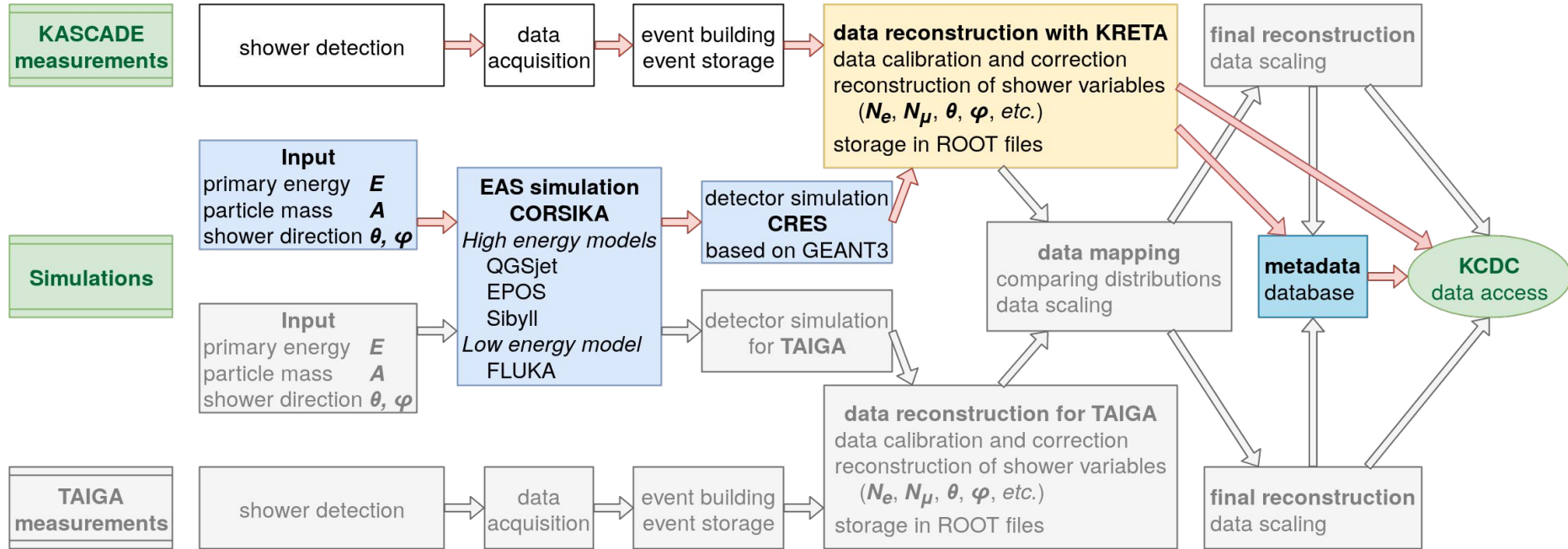
NoSQL

- ★ *BASE (basic availability, soft state, eventual consistency)*
- ★ Could require manual handling of interlocks, collisions, and data consistency;
- ★ The structure is unspecified, vague, or could change during development;
- ★ New data types can be added on the fly;
- ★ Preferable for cloud storage: easier scaling due to the built-in replication and sharding.

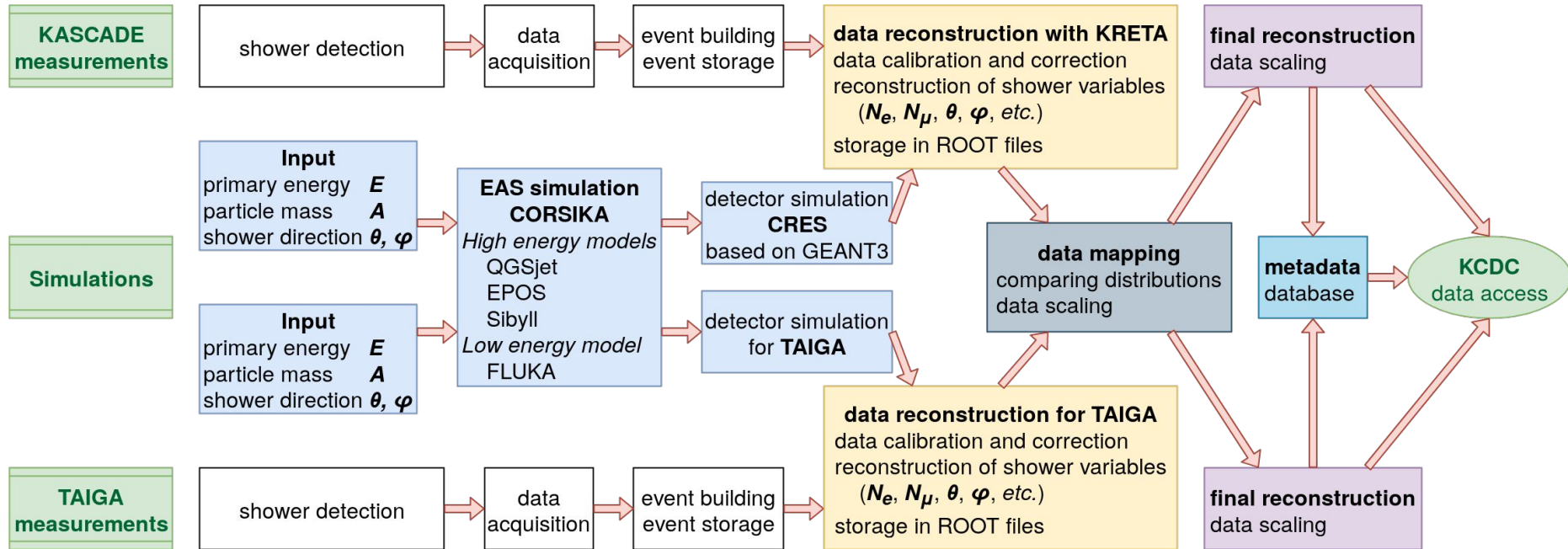


KRAD/APPDS data workflow

Currently



Joint data access scheme



Future plans and discussion

- ★ Online analysis:
 - Interface: user-made scripts or set of pre-defined actions?
 - Security: isolating user code, avoiding DoS attacks.
- ★ Distributed storage and analysis:
 - How to integrate with current site architecture?
 - What WMS to choose?
- ★ Responsibility

Thanks!

Any questions?

- victoria.tokareva@kit.edu
- <https://kcdc.ikp.kit.edu>
-

