Information Systems for the SPD Experiment

Fedor Prokoshin [1], Anisenkov. A [2], Budtueva Z.[1], Dvizov B.[3], Gussalov S.[3], Magkaev R.[3]

[1]Joint Institute for Nuclear Research

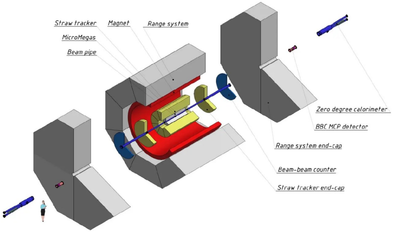
[2]Budker Institute of Nuclear Physics (BINP)

[3]North Ossetian State University (NOSU)

July 8, 2025

SPD as a Data Source

- **First stage**: polarized/unpolarized HI (up to Ca), polarized p-p, d-d

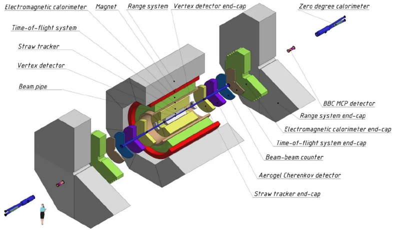


- protons and deutrons
  - $\sqrt{s} = 9.4$ GeV
  - luminosity up to $10^{31} cm^{-2} s^{-1}$
  - $\sqrt{s} = 4.5$ GeV/nucleon
  - luminosity up to $10^{30} cm^{-2} s^{-1}$
- Interaction rate: up to 0.4 MhZ
- Number of channels: 177 K
- SPD Designed to operate at high luminosity but low track multiplicity
- High event rate, no simple signatures $\Rightarrow$ Triggerless DAQ

- Event rate: up to 300 KhZ
- Online filter x20 $\Rightarrow$ 15 kHz
- Events per year: up to 100 bn
- Raw data flow: up to 1 GB/s
- Duty cycle 0.3 $\Rightarrow$ 10 PB/year
- Online filter x20 $\Rightarrow$ 0.5 PB/year

SPD as a Data Source

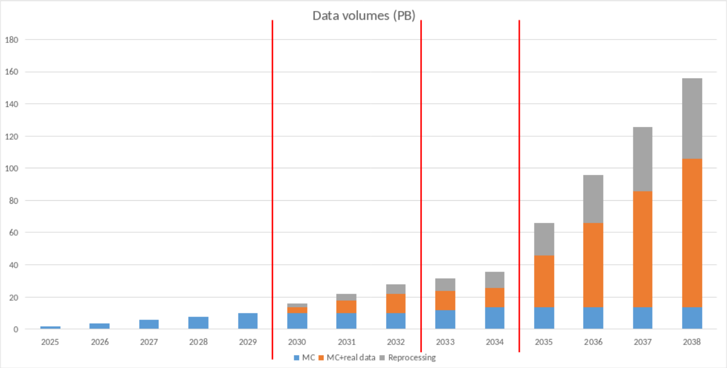
- **Second stage stage**: polarized p-p, d-d



- protons and deutrons
  - $\sqrt{s} = 27$ GeV
  - luminosity up to $10^{32} cm^{-2} s^{-1}$
  - $\sqrt{s} = 13.5$ GeV/nucleon
  - luminosity up to $10^{31} cm^{-2} s^{-1}$
- Interaction rate: up to 4 MHz
- Num. channels: 283 or 608 K
- SPD Designed to operate at high luminosity but low track multiplicity
- High event rate, no simple signatures $\Rightarrow$ Triggerless DAQ

- Event rate: up to 3 MHz
- Online filter x20 $\Rightarrow$ 150 kHz
- Events per year: up to 1 tn
- Raw data flow: up to 20 GB/s
- Duty cycle 0.3 $\Rightarrow$ 200 PB/year
- Online filter x20 $\Rightarrow$ 10 PB/year

Data Volumes



- Preparation for the experiment. MC Simulation: 2 PB per year. Total per stage: 10 PB.
- Stage I: Low lumi. MC and real data: 4 PB/y. Reprocessing: 2 PB/year. Total: 18 PB.
- Upgrade of the setup. MC simulation: 2 PB/y. Reprocessing: 2 PB/year. Total: 8 PB.
- Stage II: Max lumi. MC and real data: 20 PB/y. Reprocessing: 10 PB/y. Total: 120 PB.

Total for all stages: 156 PB.

SPD Information Systems

Data

- Distributed Computing and Data Management
- EventIndex
- Production and versions registry

Collaboration

- Personnel and Publication Databases

Detector

- Hardware and Mapping Database
- Geometry registry
- Magnetic field maps registry
- Monitoring Information Systems
- Logging and Bookkeeping

Distributed computing and data management

- Number of events, amounts of data, comparable to the LHC experiments.
- Distributed computing systems will be used for processing of data.
- Already existing methods and proven ready-made solutions will be adapted.

- **Computing Resource Information Catalog (CRIC)**: information about computing and storage resources, access protocols, entry points, etc...
- **PanDA** WFMS: manages data processing at the level of jobs (files), tasks (datasets) and chains of tasks
- **Rucio** DMS: Data management: dataset and file catalog, transfer and deletion tools, data integrity and data lifetime strategies
- **FTS** DTS: enables massive data transfers

- **Identity and Access Management** Service **(IAM)**:
  stores user profiles, their roles and rights to perform certain actions

EventIndex

- SPD needs a tool to effectively access trillions of physics events.
- Event Index: Complete catalog of SPD events, real and simulated data

Event Index use cases

- Event picking: selecting by ID events in some format and processing version
- Searching, counting and selection if events based on some parameters
- Production completeness and consistency checks

- A prototype of storage and client for the EventIndex has been tested
- PostgreSQL DBMS is used for storage and processing of data
- Various methods were investigated to speed up writing data to the DB.
- The estimated data flow from 10s k up to 150 k events per second
- Using of asynchronous bulk loading showed sufficient performance

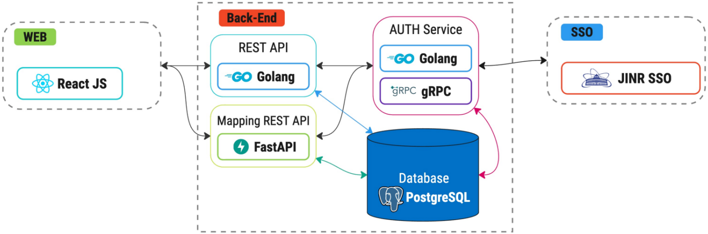
- EventIndex development will go on as data model will be defined

Hardware Database

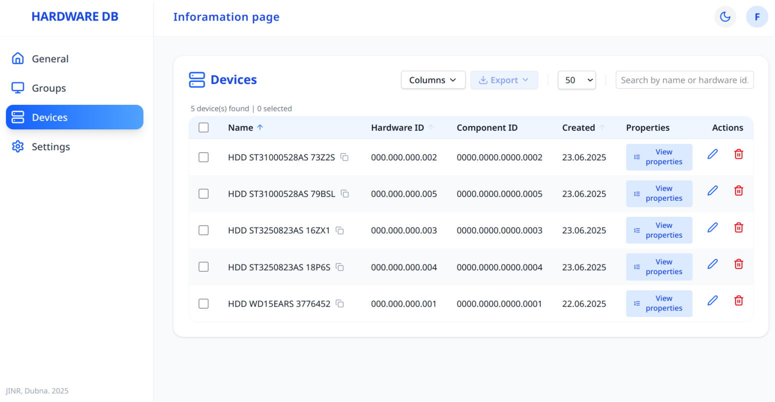
- A catalog of hardware components that SPD detector consist of.
- It should contain the information about the detectors and the electronic parts, as well as the location history of all items
- It include equipment models, provider, parameters and other (semi)permanent characteristics
- This should help in maintenance of the detector systems and especially helpful in knowledge transfer between team members.
- Filling of the hardware database should take place gradually, and updates will be rare
- The requirements for the speed of recording information in the database are low, as well as for reading of data
- The system should be tailored for the needs of the hardware development and maintenance

Hardware Database: architecture

- A prototype system is being developed
    - PostgreSQL for a data storage
    - Back-end providing endpoints for access through the REST API
    - Web front-end providing interface for fill, search and display data
    - JINR SSO for authorization, access can be requested from web front
- The architecture of the project is built on the basis of microservices
- A stack of modern platforms is used to ensure good performance, flexibility and ease of development and maintenance

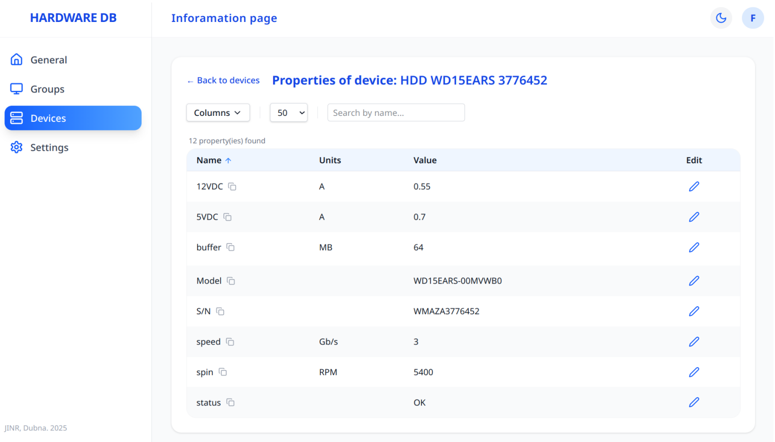


Hardware Database: interface



- Development of the HW database is put on hold, no urgent demand
- Developers shifted to other projects with higher priorities

Hardware Database: interface



- Development of the HW database is put on hold, no urgent demand
- Developers shifted to other projects with higher priorities

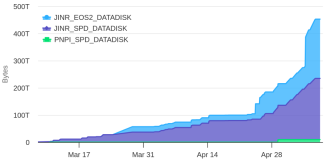
Registries

- Development of the SPD software and ongoing large scale production campaign requires creation of a number of databases.
  - Production registry
  - Software version registry
  - Geometry registry
  - Magnetic field map registry

Common features

- These are quite small projects, should be available fast
- No high loads expected
- Low update frequency
- Payloads are normally stored separately, in files or sqlite
- Have to be developed independently, but need for data exchange may arise
- Similar platforms should be used to simplify development and maintenance
- IAM should be used for authentication and access management

Production



- Production of Monte-Carlo data started from March 2025 on a regular basis
- New production appear twice a week
- 2-steps: simulation + reconstruction
- More than 400M events produced, 200TB of data
- Production requests had different parameters:
  - Numbers of events (from 5M to 40M), software versions, production seeds, energy, polarization, detector configuration, etc...
  - Production scripts names and paths, output datasets names, etc..
  - Need a database to keep records of the production
- More production campaigns will be held
- More production steps will be added, parameter list may be changed
- Real data for some detectors (BBC) may appear this year!

Production registry

- A prototype of the production registry has been created
- Expected to be filled by 2-3 responsible persons
- Will be used by all members related to the analysis and production
- Employs a framework **fwkweb**, similar to one used in CRIC

Export ▾ Columns 17/23 Filter Reload

Processing Request list

100

Request	campaign	status	# procs	swproject	version	stage	C	E	P	Events	EF	Tag	S	S	Geometry	description
17 PROD2025-001	SPD MC 2025	DONE	0	spdroot-dev	4.1.7.0	S1	pp	10	UU	5000000	4000	minbias-P8-spdroot417-dev test	1	1250	Micromegas, TS, ECal, RS, BBC, ZDC (sketch)	PRE-PRODUCTION obsolete
16 PROD2025-002	SPD MC 2025	DONE	0	spdroot-dev	4.1.7.0	S1	pp	10	UU	20000000	4000	minbias-P8-spdroot417-dev	1	5000	Micromegas, TS, ECal, RS, BBC, ZDC (sketch)	/eos/nica/spd/users/elenazem/productions/PROD2025-002_recofiles.txt PROD2025-002_paramfiles.txt
15 PROD2025-003	SPD MC 2025	DONE	0	spdroot-dev	4.1.7.0	S1	pp	10	UU	20000000	4000	minbias-P8-spdroot417-dev	5001	10000	Micromegas, TS, ECal, RS, BBC, ZDC (sketch)	/eos/nica/spd/users/elenazem/productions/PROD2025-003_recofiles.txt PROD2025-003_paramfiles.txt
14 PROD2025-004	SPD MC 2025	DONE	0	spdroot-dev	4.1.7.0	S1	pp	10	UU	40000000	4000	minbias-P8-spdroot417-dev	1	10000	Micromegas, TS, ECal, RS, BBC, ZDC (sketch)	/eos/nica/spd/users/elenazem/productions/PROD2025-004_recofiles.txt PROD2025-004_paramfiles.txt

- Production record contains name, description, software version, number of requested events, initial seed and processing types.
- On the second stage of development automatic generation of the production tasks specification may be implemented

Production registry

- A prototype of the production registry has been created
- Expected to be filled by 2-3 responsible persons
- Will be used by all members related to the analysis and production
- Employs a framework **fwkweb**, similar to one used in CRIC

SPD Data processing   Requests Config API   prof@jinr.ru

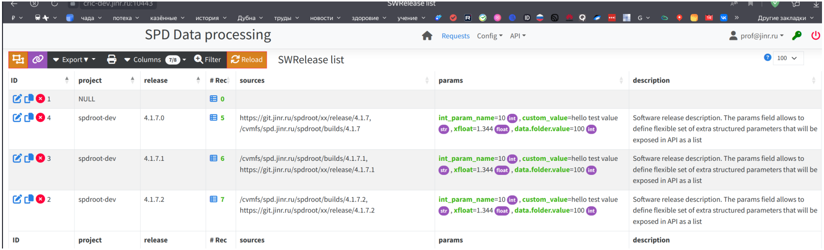
Export Columns 8/10 Filter Reload   Processing Request list   100

ID	Request	type	script	dsn_pattern	input_datasets	output_datasets	description
3	3 PROD2025-016	simu	/cvmfs/spd.jinr.ru/production/MC/minbias-P8-spdroot4172-dev.27GeV.V02/simu.C	MC2025_S2:minbias-P8-spdroot4172-dev.27GeV-UU.PROD2025-016.SIM*		MC2025_S1:MC2025_S1.minbias-P8-spdroot4172-dev.27GeV-UU.PROD2025-016.SIM.1.S, MC2025_S1:MC2025_S1.minbias-P8-spdroot4172-dev.27GeV-UU.PROD2025-016.SIM.1.P, MC2025_S1:MC2025_S1.minbias-P8-spdroot4172-dev.27GeV-UU.PROD2025-016.SIM.1.log	P8, FARICH PID for pi/K/p fixed, additional to PROD2025-012
4	4 PROD2025-016	reco	/cvmfs/spd.jinr.ru/production/MC/minbias-P8-spdroot4172-dev.27GeV.V02/reco.C	MC2025_S2:minbias-P8-spdroot4172-dev.27GeV-UU.PROD2025-016.RECO*		MC2025_S1:MC2025_S1.minbias-P8-spdroot4172-dev.27GeV-UU.PROD2025-016.RECO.1.log, MC2025_S1:MC2025_S1.minbias-P8-spdroot4172-dev.27GeV-UU.PROD2025-016.RECO.1.R	eos/nica/spd/users/elenazem/productions/PROD2025-016_recofiles.txt PROD2025-016_paramfiles.txt

- For every processing type (simu, reco...) location of the macros are specified, as well as output datset templates and names.
- On the second stage of development automatic generation of the production tasks specification may be implemented

Software Versions registry

- Registry will keep information on versions of software used



ID	project	release	# Rec	sources	params	description
1	NULL		0			
4	spdroot-dev	4.1.7.0	5	https://git.jinr.ru/spdroot/xx/release/4.1.7, /cvmfs/spd.jinr.ru/spdroot/builds/4.1.7	int_param_name=10 int , custom_value=hello test value str , xfloat=1.344 float , data.folder.value=100 int	Software release description. The params field allows to define flexible set of extra structured parameters that will be exposed in API as a list
3	spdroot-dev	4.1.7.1	6	/cvmfs/spd.jinr.ru/spdroot/builds/4.1.7.1, https://git.jinr.ru/spdroot/xx/release/4.1.7.1	int_param_name=10 int , custom_value=hello test value str , xfloat=1.344 float , data.folder.value=100 int	Software release description. The params field allows to define flexible set of extra structured parameters that will be exposed in API as a list
2	spdroot-dev	4.1.7.2	7	/cvmfs/spd.jinr.ru/spdroot/builds/4.1.7.2, https://git.jinr.ru/spdroot/xx/release/4.1.7.2	int_param_name=10 int , custom_value=hello test value str , xfloat=1.344 float , data.folder.value=100 int	Software release description. The params field allows to define flexible set of extra structured parameters that will be exposed in API as a list
ID	project	release	# Rec	sources	params	description

- Contains software tags, packages used, version descriptions links to the repositories, configuration an condition data compatibilities
- Will be filled by the person responsible for the deployment of the software frameworks (SPDRoot, SAMPO)
- Will be available through GUI and endpoints for use from other ISs

Geometries and Magnetic field maps registry

Geometries registry

- The GeoModel class library, which is presently in use by both the ATLAS and FASER experiments, will be used for the SPD detector description
- Geometry will be held held as a single file, should be used by a dedicated SAMPO framework tool
- DB record: tag for a geometry, version, metadata, link to a file on the cvmfs
- Metadata content should be defined lately

Magnetic field maps registry

- B-field maps, each having few GB in size will be held in a sqlite DB.
- Updates should be quite rear, few times a year
- DB record: tag for a map, its version, metadata, pointer to the map
- A special tool for SAMPO is being created to access B-field map, it will define metadata for the field maps and how it should be accessed.

Further development

- Development of the B-field and geometry registries
- Application of the Hardware database to the actual needs of the detector and computing systems
- Development of the mapping database
- Development of the Condition and calibration databases
- Introduction of modern DevOps practices (CI/CD, etc...)