



Data Management & Computing

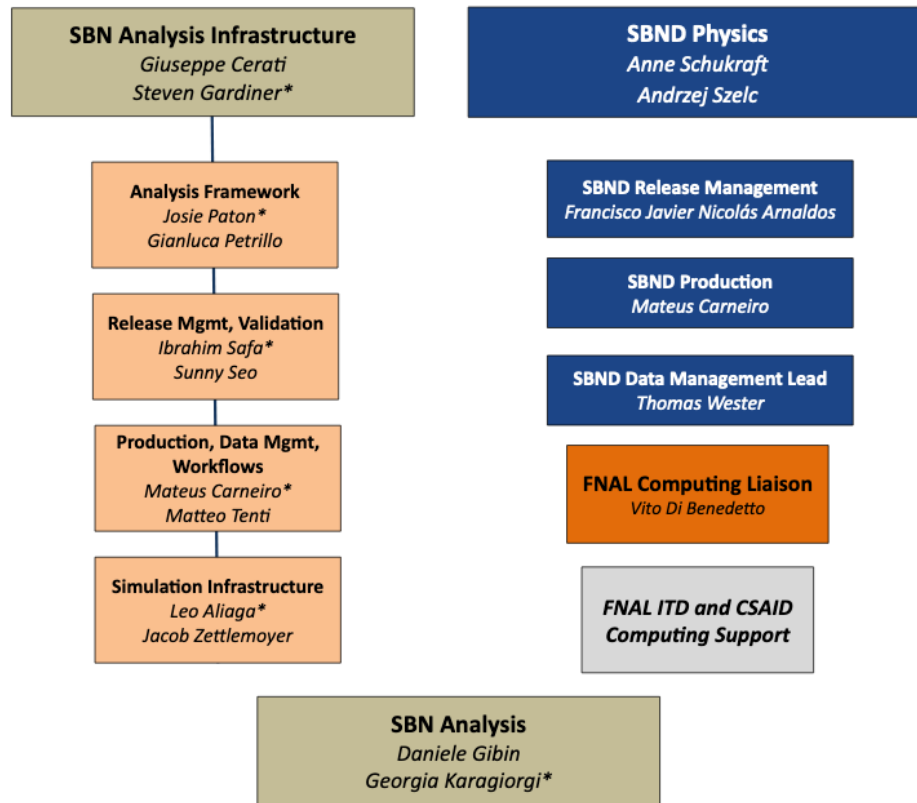
Steven Gardiner (FNAL CSAID)
SBND Operations Readiness Review
February 21-22, 2024



Organization



- SBN Analysis Infrastructure (AI) group oversees computing for both ICARUS and SBND
- Technical work to enable physics objectives from SBN Analysis group and SBND/ICARUS Physics groups
- Tools, strategies, and activities coordinated across SBN program





Charges to Analysis Infrastructure working groups

- **Analysis Frameworks**

- Support Common Analysis Format (CAF) data files across experiments, analyses, and workflows
- Oversee development of the CAF infrastructure used by analyzers, including documentation and interfaces with external tools

- **Release Management & Validation**

- Prepare and document releases of SBN-level software tools (sbncode, etc.)
- Coordinate software releases with managers of upstream (LArSoft) and downstream (sbndcode, icaruscode) parts of the stack.
- Develop, maintain, and support the use of automated tools for validating SBN software releases. Integrate them into release management process.



Charges to Analysis Infrastructure working groups

- **Production, Data Management, & Workflows**

In coordination with experiment-specific production teams:

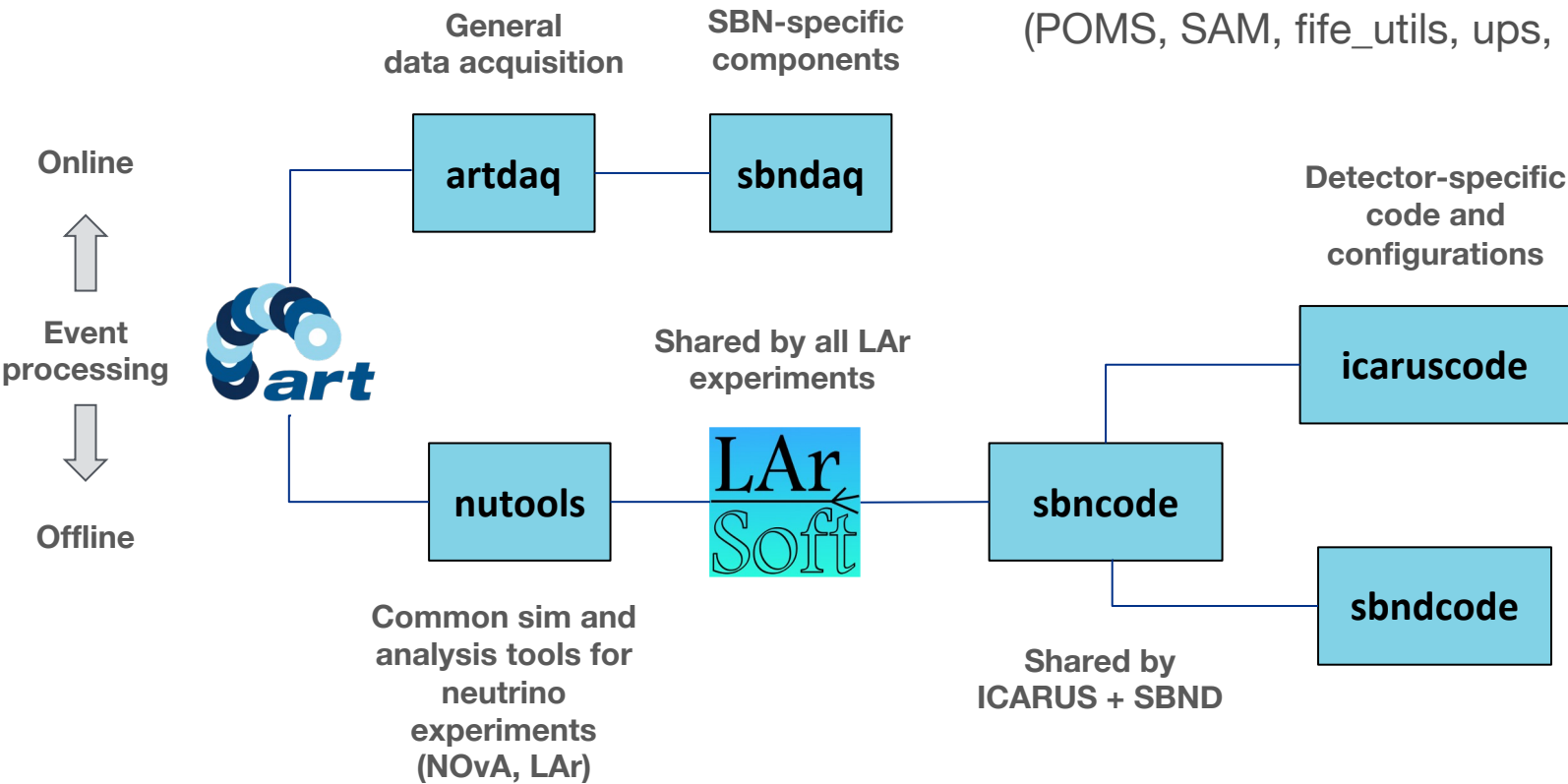
- Execute data processing and fulfillment of MC sample requests, including documentation
- Benchmark production workflows and pursue performance improvements
- Manage centralized data storage areas (disk/tape) and metadata definitions

- **Simulation Infrastructure**

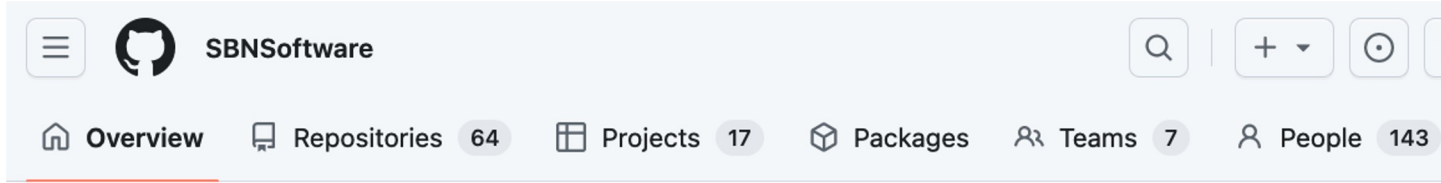
- Coordinate technical improvements to all aspects of SBN simulation codes (beam production, interaction physics, detector response)
- Ensure consistency between ICARUS and SBND simulation implementations, including knobs for correlation of uncertainties
- Infrastructure for simulation of cosmics, including CORSIKA and data overlays

Major software frameworks

Many more FNAL-provided tools also essential but not shown here (POMS, SAM, fife_utils, ups, ...)



Software release management

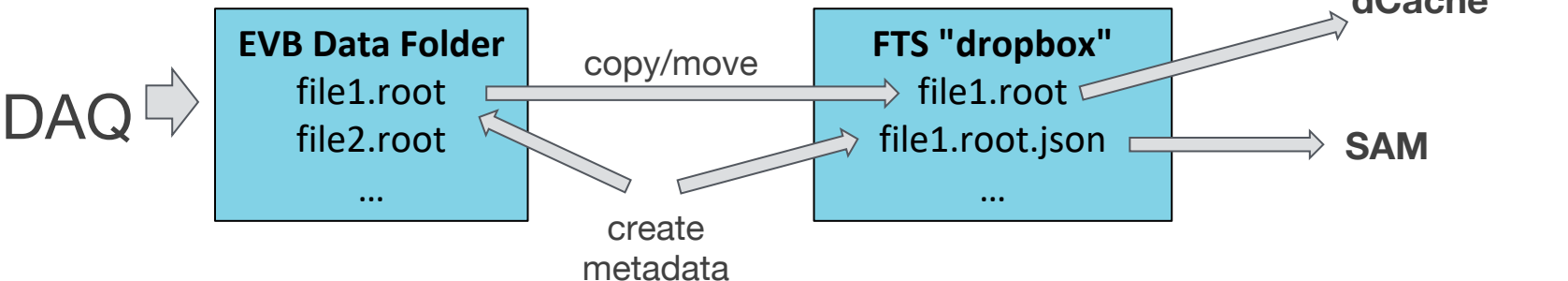


SBN Software

8 followers <https://sbnssoftware.github.io/>

- All SBN- and SBND-specific code is open source and hosted at <https://github.com/SBNSoftware>
- Regular sbndcode releases documented on software wiki page: [https://sbnssoftware.github.io/sbndcode_wiki/Releases/List of SBND code releases.html](https://sbnssoftware.github.io/sbndcode_wiki/Releases/List_of_SBND_code_releases.html)
 - Mature code base: **353 tagged versions** so far

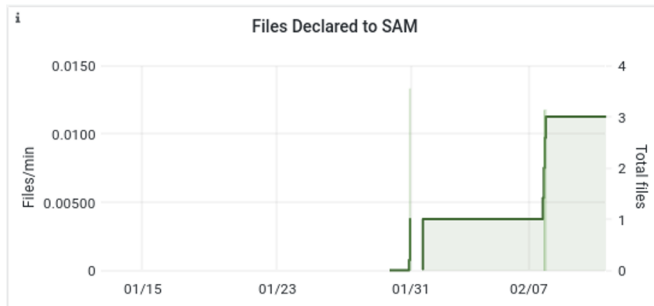
Journey of a raw data file: FTS handoff



- Four event builder servers create finalized **raw data files** in artroot format
- **File Transfer Service** provided by CSAID makes them available for centralized processing
 - Documentation: <https://cdcvs.fnal.gov/redmine/projects/filetransferservice/wiki>
- Metadata declared to Sequential Access via Metadata (SAM) database
 - Used for file management at all stages of processing

FTS status

- Successful tests of system and monitoring during commissioning
- Work ongoing to integrate metadata from run conditions database, finalize definitions
- Thomas Wester appointed operations Data Management Lead to oversee further work



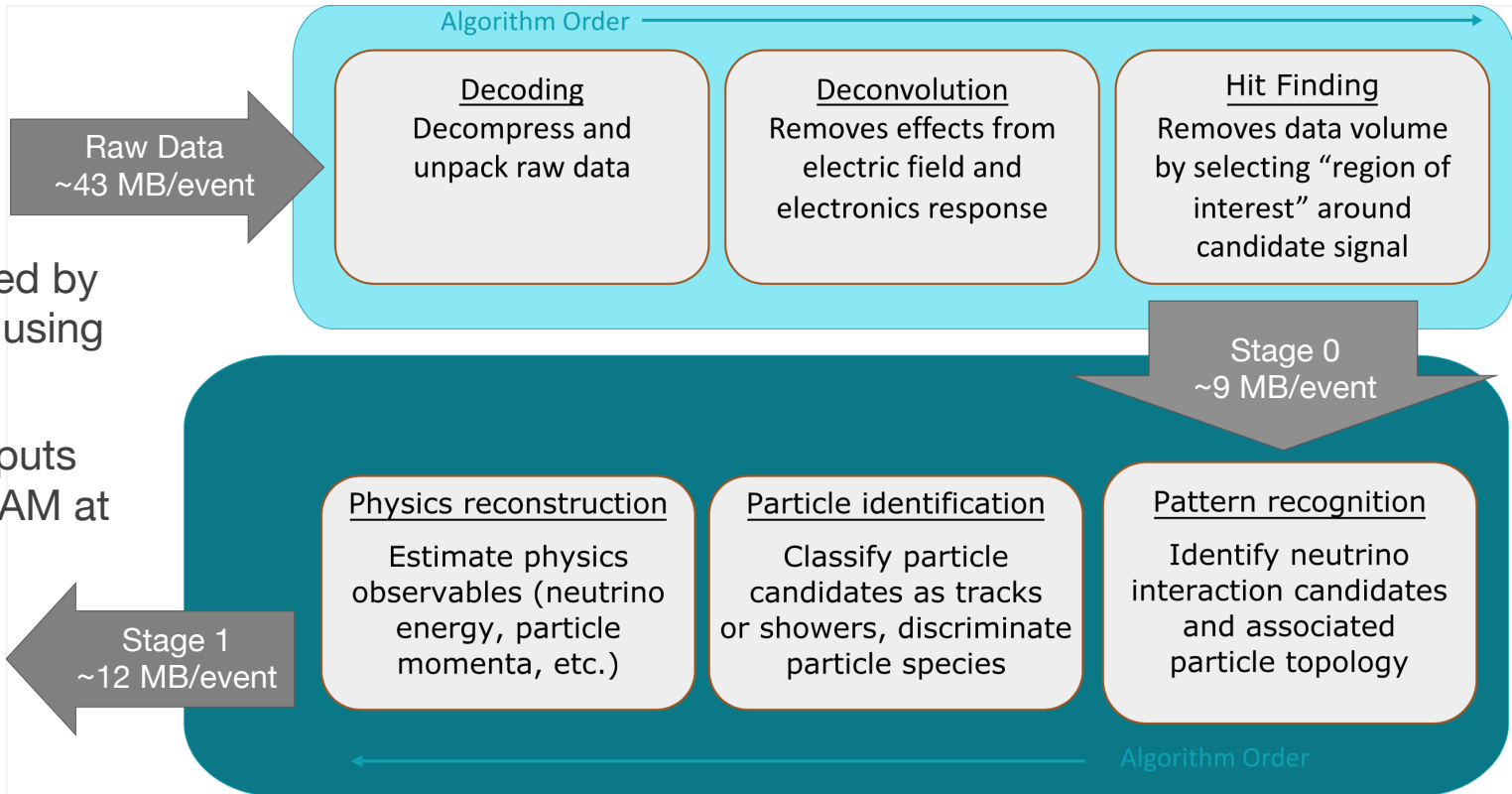
```

{
  "file_name": "data_evb01_run10880_1_20240208T182119.root",
  "file_id": 10417351,
  "create_date": "2024-02-08T23:45:27+00:00",
  "user": "sbndpro",
  "file_size": 236390859,
  "checksum": [
    "enstore:2542554270",
    "adler32:74bf4c9f"
  ],
  "content_status": "good",
  "file_type": "data",
  "file_format": "artroot",
  "data_tier": "raw",
  "data_stream": "commissioning",
  "configuration.name": "standard",
  "sbn.experiment": "sbnd",
  "sbn_dm.beam_type": "none",
  "sbn_dm.detector": "sbn_nd",
  "sbn_dm.event_count": 20,
  "sbn_dm.file_day": 8,
  "sbn_dm.file_month": 2,
  "sbn_dm.file_year": 2024,
  "sbnd.random": 0.0935383130878161,
  "sbnd.random_run": 0.913600375359277,
  "sbnd_project.name": "sbnd_daq_v1_09_00",
  "sbnd_project.stage": "daq",
  "sbnd_project.version": "v1_09_00",
  "runs": [
    [
      10880,
      "commissioning"
    ]
  ]
}

```

FTS monitoring web page: <https://fifemon.fnal.gov/monitor/d/000000032/fts?&var-experiment=sbnd>

Journey of a raw data file: centralized processing



Grid jobs managed by production team using POMS

Metadata for outputs registered with SAM at each stage

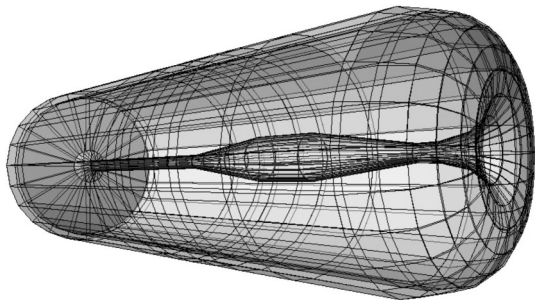


Journey of a raw data file: analysis formats

- Stage 1 artroot files are also converted to a lightweight **Common Analysis Format** (CAF) distributed for use in physics studies
 - CAF size is ~0.3 MB / event
- Specialized ntuple formats are also created for certain applications
 - Commissioning, calibrations, etc.
- Data processing tools have been successfully exercised during commissioning and MC production
 - MC adds **simulation stage** before the normal data processing stages

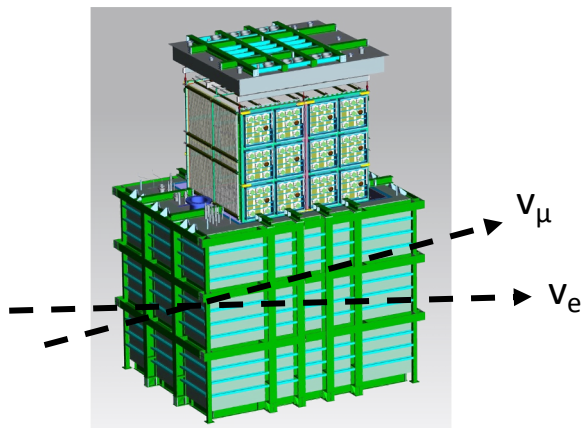
SBND analysis requires comprehensive simulations

Beam production

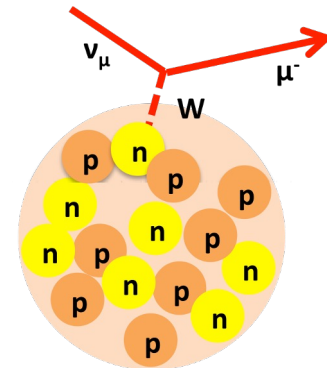


BNB horn geometry from [Phys. Rev. D 79, 072002 \(2009\)](#)

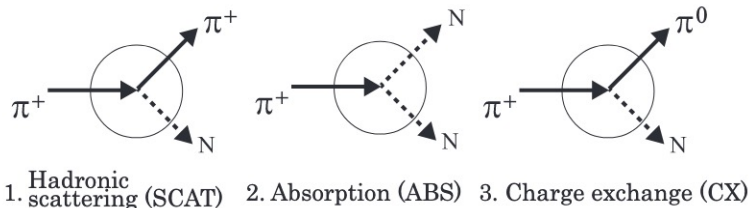
Flux propagation



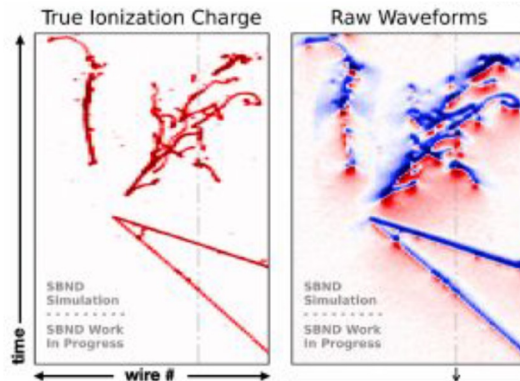
Neutrino interactions



Particle transport



Detector response



Ongoing Fermilab support for these tools is critical to our science goals





Production procedures (1)

- Campaigns planned by AI organization in coordination with physics leadership from both experiments
 - Ensures consistency for joint ICARUS + SBND studies
 - Production release for first 2024 campaign ("MC2024A") nearing completion
- Online form used to aggregate MC sample requests from working groups
 - Proposed workflow, contact person, desired event count, etc.
 - Only workflows tested locally are run by production, form asks for expected resource usage
 - Form populates a spreadsheet that is used by the production group to organize and document each sample configuration



Production procedures

- Mini-productions defined in advance, used for final validations
- Fulfilled requests documented for analyzers on SBN Software wiki:
https://sbnsoftware.github.io/sbn/sbnprod_wiki/sample
 - SAM definitions for CAF files and upstream stages
 - Latest entries are from second MC campaign from last year (“MC2023B”)
- GitHub [issue tracker](#) used to manage software feature requests



FCRSG Review Process

- **Fermilab Computing Resource Scrutiny Group**
 - Oversees resource allocations for the lab, see [slides](#) from ALD Jim Amundson
 - AI coordinators present joint ICARUS + SBND resource requests annually
 - 2023 presentation [here](#)
- **Last year's presentation:** 5-year predicted CPU/disk/tape usage, some technical questions (data lifetimes, etc.)
- Updates to resource requests for 2024 in progress, next review meeting date not yet finalized
 - Estimates in EOP and in these slides are based on **2023 FCRSG numbers**

Resource summary and file families

- Estimates developed based on current data/MC processing workflows and performance
- All raw data to be preserved long-term on tape storage
- **File families** used to group similar files together for efficient retrieval from tape
 - Several expected based on separate data streams: *commissioning, test, cosmics, minbias, beam, and calibration*
 - Subdivided based on processing stage (*raw, stage0, stage1, caf*)

Table 6: SBND computing resource estimates for 2024

Grid time (CPU hours)	21 M
Persistent disk	2.9 PB
Dedicated write (shared with ICARUS)	2 PB
Tape	4.3 PB
/exp/sbnd/app storage	8 TB
/exp/sbnd/data storage	35 TB

Computing support from Fermilab: ITD



Ongoing contributions from ITD and CSAID are essential for SBND operations and analysis

Table 4: Listing of services provided by Fermilab's Information Technology Division (ITD) in support of SBND operations

Authentication and Directory Services	Standard KCA and DNS services provided.
Central Web Hosting	Support for the SBN central web server, including the SBN online portal, and SBN DocDB and FNAL Indico.
Database Hosting	Database hosting and database infrastructure used by SBND.
Desktop Services	Windows and Mac desktop support for the computers covered by the Managed Services contract.
Fermilab (Data Center) Facilities	Support for laboratory space for DAQ test stands and collaboration common computing nodes.
Network Services	Standard support for detector facilities. Essential SBND network devices are supported for 24x7 service.
Networked Storage Hosting	Support for home areas and NAS attached data disks
Service Desk	Issue and notification reporting, handling and tracking.

Computing support from Fermilab: CSAID



Table 5: Listing of services provided by Fermilab's Computational Science and Artificial Intelligence Directorate (CSAID) in support of SBND operations

Grid and Cloud Computing	Batch processing on Grid accessible systems at Fermilab as well as offsite through the Open Science Grid and HEPCloud. Jobsub, GlideinWMS, CVMFS, “POMS” production software, and other software for enabling processing and analysis.
Scientific Collaboration Tools	SBND code repositories hosted through cdcvs.fnal.gov , redmine, GitHub, and the electronic log-book application.
Scientific Computing Systems	Support for control room and SBND computing systems and workstation administrative support. Support for interactive, batch processing, simulation and analysis computing systems at Fermilab.
Data Acquisition Servers	Maintenance and repairs provided by the SLAM team
Scientific Data Management	SAM, IFDH, FTS, RUCIO, and other data handling software and systems that are essential to online data transfer.
Scientific Data Storage and Access	Enstore-based tape storage services. Tape handling and curation. dCache-based data disk services and systems.
Scientific Databases	Applications and database infrastructure for identified SBND online and offline databases, including hardware mapping database, calibration database, run history database, and the IFBEAM database.
Scientific Software	Support for LArSoft, artdaq, art, ROOT, and other software tools.
Simulation Software	Support for Geant4, GENIE, and other simulation codes.



Support needed for upcoming transitions

- Key tools related to scientific computing infrastructure continue to evolve
- SBND computing leadership is planning for these transitions, but continuing FNAL involvement to help us through them will be necessary
- Some representative examples:
 - Scientific Linux 7 → Alma Linux 9
 - ups → Spack
 - SAM → MetaCat + DataDispatcher
(see slides from recent mini-workshop: <https://indico.fnal.gov/event/62858/>)
- Recent migration of our interactive node storage from NAS to CephFS was highly successful
 - We hope to work effectively with lab experts on these future changes as well

Summary



- SBND has a great computing team in place together with a mature software stack
 - Built on a strong foundation of prior work by FNAL experiments
 - Close cooperation with ICARUS in all aspects of our computing program
- Continuing support from ITD and CSAID will be essential for our success
 - Resources and technical expertise in many relevant areas
 - Separate annual review process (FCRSG) keeps our computing strategy in sync with the laboratory