



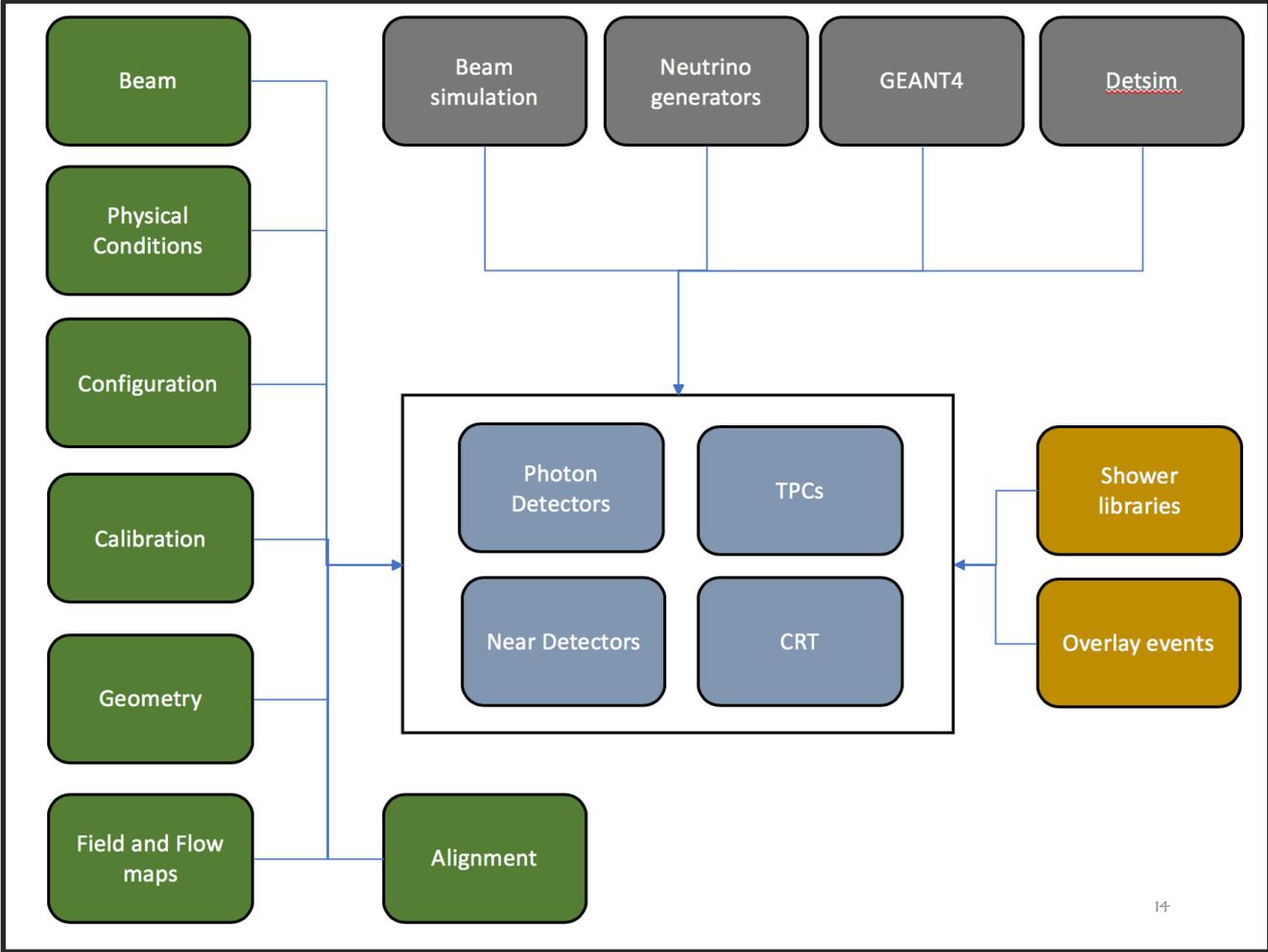
Offline Processing

Needs and Interfaces

Wesley Ketchum (FNAL)

DUNE Computing Workshop I – Data Model

Very helpful slide from Heidi



A silly and doomed attempt

- I want to try to list the data needed to do analyses that should be coming from our data model
 - In some ways, trying to think of this as a checklist to know that we have concrete plans for each
 - Paying less attention to:
 - Exactly where the data comes from
 - Optimization of getting the data
- ... but of course those matter and shouldn't be forgotten
- See google doc here for some of my first stabs
 - https://docs.google.com/document/d/1_74XeYwo3DcNSTlxG7K5KS4nHVlljzodBduVfbFXjCw/edit?usp=sharing
 - Of course I will have forgotten something obvious and essential
 - Thank you for your patience, your feedback welcome

Need the basic detector data

- Two main classes of events:
 - Extended in time
 - E.g. SNB. We will read out whole detector over long period of time (100 s).
 - Luckily processing and analysis can handle smaller parts (per APA, per drift period or less) at a time
 - Non-extended in time
 - E.g. beam, cosmics, atmospheric, proton decay, etc. Read out (up to the) whole detector, for ~one drift period.
 - Analysis may need to activity over an extended range in space
- Can be thought of in two phases:
 - First stage signal processing:
 - Need raw waveforms, raw light detection information, etc. extracted from the DAQ data payloads
 - Produce deconvolved/semi-calibrated signals/hits
 - Second-stage processing:
 - Use above, and perform 3D reconstruction and particle ID

Handling raw data

- Information necessary to unpacking detector data
 - Offline needs to know how/be able to convert raw DAQ data to offline formats
 - E.g. readout window lengths
- Information necessary to verifying detector data integrity
 - Need to know the data obtained is valid and (re)perform integrity checks
 - E.g. checksums, status bits, etc.

Trigger information

- Trigger bits / data selection algorithm and decision information
 - Need to know conditions/decisions under which data was selected for collection
- Trigger time / GPS time
 - Need to know exactly when the data corresponds to
- In simulation, simulated information used for making trigger decision
 - Need to be able to repeat trigger algorithms on simulation, and study efficiency/uncertainty on efficiency/etc.
 - E.g. raw waveforms *around simulated activity*

Run configuration information

- Detector data readout configuration/settings information for a given run and given timestamp
 - Need to know how to deconvolve/calibrate/reconstruct detector data
 - E.g. gain settings, shaping, readout thresholds, etc.
- “Run type” information and relevant per-run configuration/settings information
 - To mark data for specific studies/processing
 - e.g. physics run, calibration run, etc.

Calibration systems information

- Auxiliary/calibration detector information, whether on per-run, per-time-range, or per-trigger-readout basis
 - For association of calibration information to other ‘standard’ detector data
 - e.g. test pulse, laser, radioactive sources, etc.

Detector uptime information

- Enabled/disabled detector components during a given run and given timestamp
 - Need to know the expected fraction of detector live/operating/exposure
- Active detector components and dead/bad channels during a given timestamp
 - Need to know the actual fraction of detector live/operating/exposure at data collection
 - this may include any missing data from DAQ, any readout/DAQ deadtime, and dead/bad channels from quality monitoring or offline analysis
- (Should this be a basis for determining run/subrun?)

Detector conditions and calibration information

- High voltage settings/measurements
 - For determination of drift time, basic recombination, etc.
- Wire bias settings/measurements
 - For signal processing
- Argon purity measurements
 - For determination of lifetime
 - Online (purity monitors/Ar39 measurements?) and offline measurements needed
- Other detector conditions data for verifying data quality
 - For verifying data quality likely an (eventually) well-defined subset of total detector conditions data
 - e.g. HV stability metrics, wire bias stability, temperature and pressure, trigger primitive rates?, etc.
- EField and argon flow maps
 - For reconstruction of position and charge corrections

Beam information

- Per-spill beam information
 - For determining protons on target, beam timing, beam quality

Other

- Geometry and alignment information
 - For basic reco (likely doesn't change often)

And ... that's all I have right now

- Comments welcome
- Items missing especially welcome
- I can guess where many of the elements on this list will come from, but would be good to agree on data source, when and how often that information will be accessed and used, and many of the other details of course