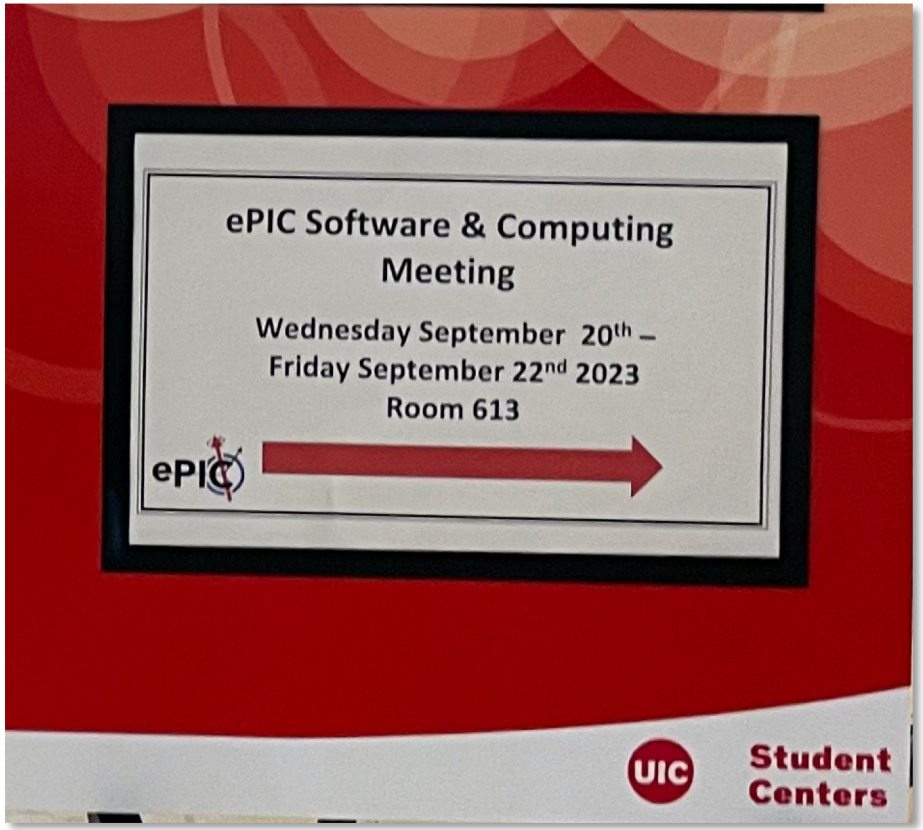
ePIC Software & Computing

Report from Meeting at University of Illinois at Chicago

Meeting Summary

- **September 20–22**
- **Indico**: https://indico.bnl.gov/event/20159/
- **43 Participants, in person and remote.**



Thanks to Olga Evdokimov for hosting us!

- **Purpose**: Collectively drive forward priority targets and provide an avenue for new collaboration members to actively engage.

  In-person component invaluable.

- **Priority targets**:
  - **Reconstruction Workflow**: What is the algorithmic workflow for a holistic reconstruction of physics events?
  - **Streaming Computing Model**: Open questions prior to the review?
  - **User Learning**: How can we expedite the process of creating instructional material for onboarding new users and developers?
  - **Validation**: How can a collaboration member access the accuracy and quality of our simulations?

Highly Productive Discussions

- Summarized detailed discussions in 12 pages:
  - **Direct link**
- Identified **priority action items**.

Thank you Torre!

Live Notes

Meeting	ePIC Software & Computing at University of Illinois at Chicago
Date	September 20-22, 2023
Indico	https://indico.bnl.gov/event/20159/
Timetable	Timetable on One Page Timetable Indico Style

Meeting Objectives

The purpose of the meeting was to collectively drive forward priority targets and provide an avenue for new collaboration members to actively engage:

- **Reconstruction Workflow**: What is the algorithmic workflow for a holistic reconstruction of physics events?
- **Streaming Computing Model**: What open questions require attention prior to the ePIC Software & Computing review on October 19 and 20? What is the progress status of the review preparations? The charge of this review covers an assessment of the ePIC Computing Model in preparation for the EIC Resource Review Board (RRB) meeting in December 2023.
- **User Learning**: How can we expedite the process of creating instructional material for onboarding new users and developers?
- **Validation**: How can a collaboration member access the accuracy and quality of our simulations?

Next F2F Meeting

Working on a plan to have the next meeting at CERN. ~~Feb~~? Apr? May? Feb is very shortly after collab meeting. A CERN meeting acknowledges the international nature of our collaboration, and gives us the opportunity to meet and discuss with the developers in the CERN EP-SFT group who contribute to many of the software packages in our stack. Also a good opportunity to follow up on MC generator work with the Theory group and HSF collaborators there.

Important that we announce it as soon as possible, preferably 6mo in advance, to accommodate visa timescales.

Reconstruction Workflow

- podio schema evolution is supported now, it was not a year ago. Perhaps not well tested yet.
- python tooling that can work directly with the data model is autogenerated (as well as C++). Some people use it, seems to work.
- EDM4hep is gaining traction and modernizing its technologies.
- reco currently expects to see events, mapping onto evgen events. Not set up to deal with the many events in a frame.
- Jana2 does have a built-in workflow to support frames in and events out
- effectively you are going to have software that finds events in the timeframes, much like a trigger if you walk through the beam crossings reconstructing as you go, combining the info you have across detectors with their appropriate time slice scales, and identify events. Different points of view. Stepping through 1 by 1 could present a significanant bookkeeping need.
- 1ms is 500 beam crossings, 10MB, workable amount of data, in-mem is easy
- each subdetector does its reconstruction on its natural time slice scale
- after detector-appropriate reco at this data stream level, based on time slice scale appropriate to the detector, we look to higher level reco combining across detectors ultimately identifying and emerging with physics events as the outcome of the reco
- conclusion: "different approaches" are in violent agreement on the same approach.
- need the full, wide MAPS integration window to do the tracking
- real physics collision event rate, ~1 every 2 microsecs.
- individual ePIC events are very small. There is advantage to processing many at once, not just small chunks, especially for using accelerators. Already having this structure of 100's of events together is an advantage.
- frame building infrastructure post-Geant4 post-digi is there and ready. Just specify frame size.
- background makes one 1ms timeframe expensive to simulate. Hence importance of embedding/reuse of backgd.
- adapting reco to working with frames is a substantial effort. Have to fit it into the pre-TDR effort, we need TDR to reflect working with frames-aware reco. e.g. does frame-based tracking work. Re-qualify the tracking with this new time dimension and in presence of full timeframe worth of background.
- seed events from the fast detectors, fit to the slow detectors, emerge with events
- need to find the right strategy and timeline for getting in frame based reco. We do need it pre-TDR.
- start with a sub-project in the reco to work from the frames, a proof of principle. Don't need to and can't do everything at once.
- at some point will be doing hw tests using our sw. Need to be ready for that. Plan to have needed pieces like time-dependent channel maps. And the calib DB to support the lookup.

- apparatus for combined reco across subdetectors, defining events and then refining the (re)reco/fitting of objects within the event, isn't in yet, needed e.g. for particle flow, refining tracking in light of PID, etc.
- digi should be independent of reco, its own module, outside EICrecon. We'll need a capability for algorithm sharing.
- GPUs are not a priority right now.
- ML inference in reco is important. onnxruntime has been working well. [also in ATLAS.]
- should address metadata, conditions db in the reqmts/decision process. Our next priority after the review. AI/ML also by the end of the year.
- in latest podio which we are using, the file contains the full schema info describing the data
- need to set milestones intermediary to the TDR. Reviewers like milestones!
- people have the opportunity to take ownership and responsibility for issues. Own a part of the software ecosystem! Already in a much better state than a year ago.
- go through existing priority list, update and prioritize them in light of these discussions

Essential Action Items

- **Framework Tests and Development**:
  - Confirm and test the stability of the newly supported Podio schema evolution.
  - Look into the integration and effectiveness of Jana2's built-in workflow for supporting frames in and events out.
- **Reconstruction (Reco) Process**:
  - Adapt the reco process to work with frames, making it frames-aware. This involves setting up the system to manage events within the timeframes and the integration of detectors with their respective time slice scales.
  - Initiate a project for proof of principle on working from frames.
- **Data and Infrastructure Management**:
  - Ensure the output files contain complete schema information describing the data.
  - Utilize the full, wide MAPS integration window for tracking purposes.
  - Separate digitization from reco, and evaluate the need for algorithm sharing capabilities.
  - Implement and utilize the frame-building infrastructure post-Geant4 post-digi.
- **Planning and Prioritization**:
  - Develop a strategy and timeline for implementing frame-based reco, ensuring it's incorporated pre-TDR.
  - Review and update the current priority list based on the discussions, ensuring necessary tasks are properly prioritized.
  - Address the need for metadata and the conditions database in the requirements/decision process.
  - Prioritize AI/ML development by the end of the year.
  - Postpone focus on GPUs after the TDR.
- **Collaboration and Future Development**:

Priority Targets

Reconstruction Workflow: What is the algorithmic workflow for a holistic reconstruction of physics events?

Streaming Computing Model: Open questions prior to the review?

User Learning: How can we expedite the process of creating instructional material for onboarding new users and developers?

Validation: How can a collaboration member access the accuracy and quality of our simulations?

Reconstruction Workflow

- **Framework Tests and Development**:
  - Confirm and test the stability of the newly supported Podio schema evolution.
  - **Look into the integration and effectiveness of Jana2's built-in workflow for supporting frames in and events out.**
- **Reconstruction (Reco) Process**:
  - **Adapt the reco process to work with frames, making it frames-aware. This involves setting up the system to manage events within the timeframes and the integration of detectors with their respective time slice scales.**
  - Initiate a project for proof of principle on working from frames.
- **Data and Infrastructure Management**:
  - Ensure the output files contain complete schema information describing the data.
  - Utilize the full, wide MAPS integration window for tracking purposes.
  - Separate digitization from reco, and evaluate the need for algorithm sharing capabilities.
  - Implement and utilize the frame-building infrastructure post-Geant4 post-digi.
- **Planning and Prioritization**:
  - **Develop a strategy and timeline for implementing frame-based reco, ensuring it's incorporated pre-TDR.**
  - Review and update the current priority list based on the discussions, ensuring necessary tasks are properly prioritized.
  - **Address the need for metadata and the conditions database in the requirements/decision process.**
  - **Prioritize AI/ML development by the end of the year.**
  - Postpone focus on GPUs after the TDR.
- **Collaboration and Future Development**:
  - Promote the opportunity for team members to take ownership and responsibility for different parts of the software ecosystem.

Priority Targets

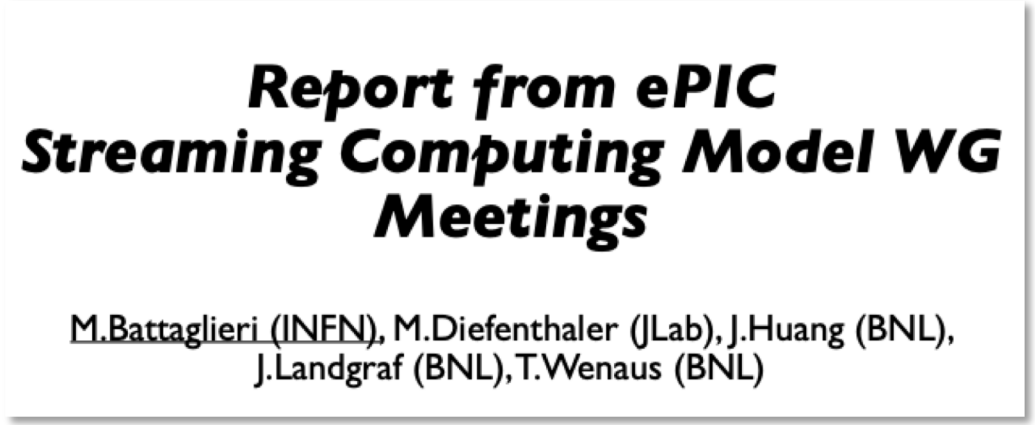
Reconstruction Workflow: What is the algorithmic workflow for a holistic reconstruction of physics events?

Streaming Computing Model: Open questions prior to the review?

User Learning: How can we expedite the process of creating instructional material for onboarding new users and developers?

Validation: How can a collaboration member access the accuracy and quality of our simulations?

Streaming Computing Model



Detailed Update from Marco

Review of ePIC Software & Computing (October 19-20)

Oct. 10 Final ePIC Streaming Computing WG meeting prior to the review.
Oct. 11 Dry run from 8:00 – 11:00 a.m. (EDT)
Oct. 11 Draft for report on the ePIC Streaming Computing Model being shared.
Oct. 13 Deadline for immediate feedback on the draft report.
Nov. 10 Deadline for detailed feedback on the draft report.

Report will be developed into a publication on the ePIC Computing Model.

Note: Updated live during the meeting from Oct. 31 to Nov. 10 to provide the collaboration with a month for feedback.

Review Charge

A committee of experts, the **EIC Computing and Software Advisory Committee** (**ECSAC**), has been formed to **advise Brookhaven National Laboratory (BNL) and Jefferson Lab (JLab)** on the progress and status of computing and software for the ePIC collaboration. Reviews are expected to take place on a regular cadence, with a charge reflective of the EIC schedule, the stage of the ePIC experiment, and impending deadlines.

The charge of this review covers an **assessment of the ePIC computing model** in **preparation for the EIC Resource Review Board (RRB) meeting** in December 2023. The scope of this **review** also **includes the organization of the EIC Computing and Software Joint Institute** (**ECSJI**), which has been formed by BNL and JLab.

- At this stage, approximately ten years prior to data collection, is there a comprehensive and cost-effective long-term plan for the software and computing of the experiment?
- Are the plans for integrating international partners' contributions adequate at this stage of the project?
- Are the plans for software and computing integrated with the HEP/NP community developments, especially given data taking in ten years?
- Are the resources for software and computing sufficient to deliver the detector conceptional and technical design reports?
- Are the ECSJI plans to integrate into the software and computing plans of the experiment sufficient?

Review Agenda

October 19

Executive Session

9:00 AM — **ECSAC Welcome and Executive Session**
ECSAC-Charge-Fall-... | ECSJI - BNL JLab A...

Review

10:00 AM — **Introducing the ePIC Experiment: Exploring Use Cases and Workflows**
Speaker: Rosi Reed (Lehigh University)

10:30 AM — **Q & A**

10:45 AM — **ePIC Software Stack: Status and Plans**
Speaker: Sylvester Joosten (Argonne National Laboratory)

11:30 AM — **Q & A**

11:55 AM — **Break**

12:00 PM — **The ePIC Computing Model**
Speaker: Dr Markus Diefenthaler (Jefferson Lab)

12:45 PM — **Q & A**

Executive Session

1:15 PM — **ESCAC Working Lunch / Dinner**

3:00 PM — **ESCAC Shares ePIC Homework Questions**

October 20

Review

9:00 AM — **ePIC Homework Q&A**

10:00 AM — **Organizational Structure of ECSJI**

10:30 AM — **Q&A**

10:45 AM — **The EIC Computing Organization**

11:05 AM — **Q&A**

11:20 AM — **Break**

11:30 AM — **International Data Facilities Perspective on the Computing Model**
Speaker: Andrea Bressan (University of Trieste and INFN)

11:45 AM — **Q&A**

Executive Session

12:00 PM — **ESCAC Working Lunch / Dinner**

Review

2:00 PM — **Close Out**

The ECSJI and EICO presentations will be given by Amber Boehnlein (JLab) and Eric Lancon (BNL).

Priority Targets

Reconstruction Workflow: What is the algorithmic workflow for a holistic reconstruction of physics events?

Streaming Computing Model: Open questions prior to the review?

User Learning: How can we expedite the process of creating instructional material for onboarding new users and developers?

Validation: How can a collaboration member access the accuracy and quality of our simulations?

User Learning

- **Landing Page Redesign**:
  - Decide the primary audience of the landing page (new users vs frequent returners; an expert on subject A can be a new user on subject B).
  - Reconsider the text-heavy design in favor of a more visual layout, like clickable boxes.
  - Add contact information relevant to individual points.
- **Documentation and FAQs**:
  - Aim to consolidate documentation from various places into one centralized location.
  - Provide simple examples and categorize them (event generation, simulation, analysis).
  - How to strengthen documentation:
    - FAQs are bite-sized documentation tasks that help us advance:
      - Having too many FAQs and needing to restructure them into documentation would actually be a good problem to have.
      - Get conveners to identify common questions and designate individuals to provide answers.
      - Re-evaluate and potentially shrink the FAQ section as items are transferred to other documentation sections.
    - Acknowledgment of documentation as an official service:
      - Consider assigning this as a service task for beginners.

Priority Targets

Reconstruction Workflow: What is the algorithmic workflow for a holistic reconstruction of physics events?

Streaming Computing Model: Open questions prior to the review?

User Learning: How can we expedite the process of creating instructional material for onboarding new users and developers?

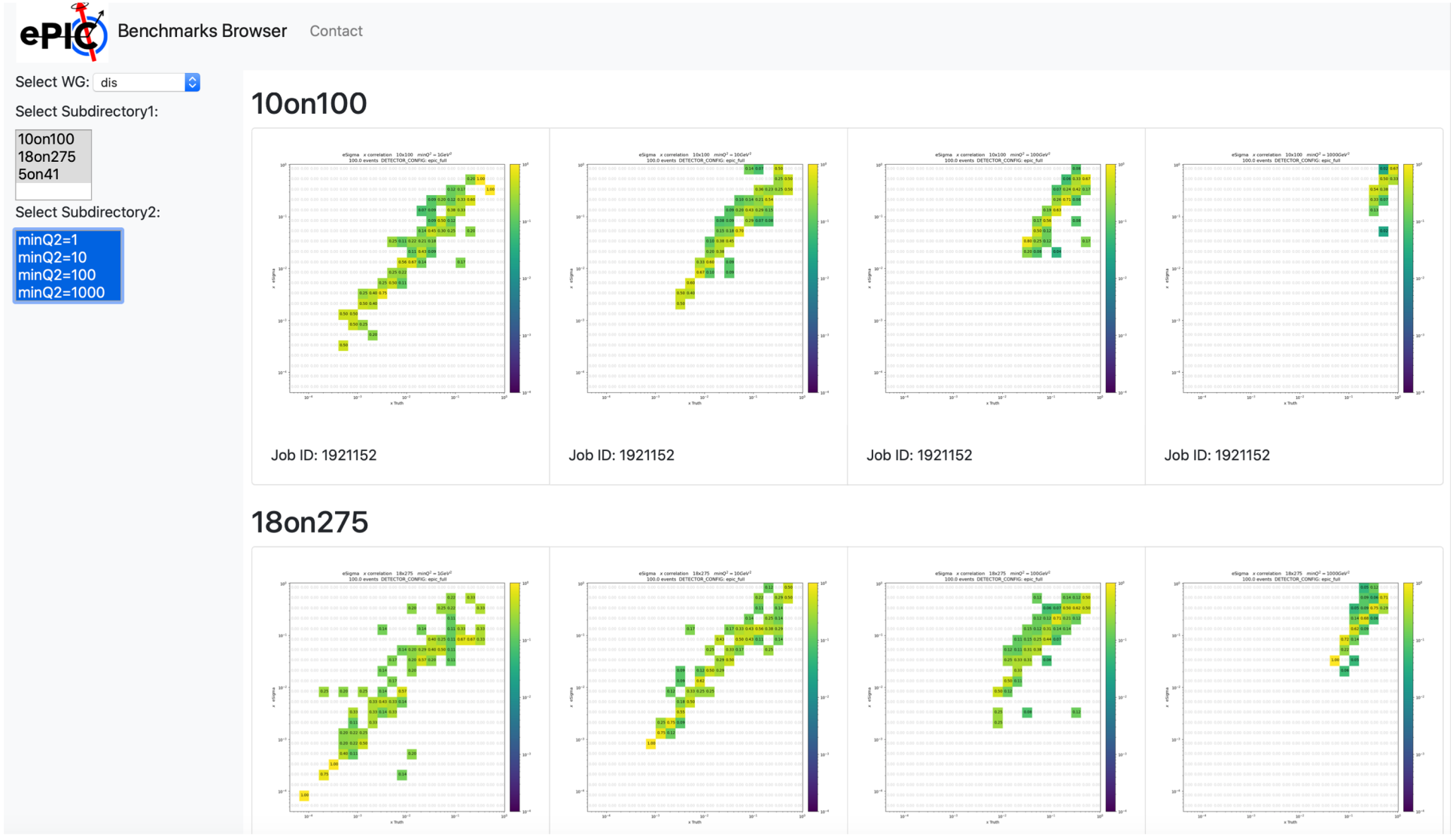
Validation: How can a collaboration member access the accuracy and quality of our simulations?

Validation: Priority 1

Priority 1 – Ongoing Campaign :

- Clarify the Definition of Benchmarks: Ensure a shared understanding of what benchmarks mean in the context of ePIC software development and ePIC simulation campaigns.
- Focus on Detector Benchmarks (in addition to the physics benchmarks we have already received)
- Develop Template for Validation Contributions:
  - Create a standardized example script with associated metadata (inputs, workflow, configuration, outputs) that can serve as a template for validation contributions.
- Presentation/Processing Tool: Continue developing a presentation/processing tool that can be used in the short term.

Presentation/Processing Tool



Validation: Priority 2 and 3

- **Priority 2: Improvements**
  - Benchmark Scripts:
    - Ensure that the validation script is robust against misuse, such as running on unsupported inputs.
    - Incorporate and require descriptions, including plot captions, within the validation script to enhance understanding.
    - Centralize Script Development: Establish a centralized approach to script development.
  - Workflows:
    - Begin with a straightforward workflow managed by a single script, simplifying the process without introducing complex workflow management tools.
    - Consider the integration of more complex workflows into the CI system (EICweb).
  - Benchmarks:
    - Develop tests and presentations that can detect major changes immediately, particularly those related to geometry revisions.
- **Priority 3: Further Ideas for Improvement**
  - Transform benchmarks into performance metrics to emphasize the grading aspect and track performance over time.
  - Include Reference Plots and Statistics.

Planning

- Follow-up on **UIC meeting priorities**.
- **Post-review topics**:
  - Software decision on metadata handling.
  - AI/ML integration in reconstruction and production workflows.

- Joint meetings with PWGs will start at the end of October and will occur monthly.

- Planning to **omit the January campaign** in consideration of the holidays.
  - December campaign will be important input for the collaboration meeting.

- Proposed **parallel session on software and simulation readiness for TDR** during the collaboration meeting.
- **Next in-person meeting**: April/May 2025 at CERN.

News

- **Communicating our Software Progress**:
  - Detailed changelogs structured by subsystems, e.g., https://github.com/eic/EICrecon/releases/tag/v1.6.0

- **New Detector Geometry Matrix** Detector-20230929162408 **Many thanks to EIC Project!**
  - Released on Sept. 29:
    - Overall update to make detector envelopes consistent with the Sept 2023 CAD model.
  - CSV version available**. Highly appreciated! Consistency check ongoing.**

EIC GEOMETRY FRI, 29 SEP 2023 16:24:08

EIC DETECTOR GEOMETRY

INTERACTION POINT 6

Region	Component	Sub-Component	WBS	Length (cm)	Inner Radius (cm)	Outer Radius (cm)	Offset from Center (cm)	Physical Start (cm)	Physical End (cm)	Volume (m^3)	Weight (kg)	Technology	Notes
HADRON DIRECTION END CAP	HD Flux Return (Collar)			170	269	324	414.6	329.6	499.6	17.42	136,685	Iron	Offset: measured from center. Weight estimated as 100% iron.
	Hadron Calorimeter		6.10.06	140	17.5	267	359.6	359.6	499.6	31.22	199,896	FeSc, WSc last segment	Tower size: 5cm x 5cm x 140cm including 10cm readout Offset: measured from face nearest to interaction point Weight: estimated as 79% iron and 21% plastic
	HD Flux Return (Oculus)			22.2	195	267	340.7	329.6	351.8	2.32	18,205	Iron	Offset: measured from center. Weight estimated as 100% iron.
	Electromagnetic Calorimeter		6.10.05	30	14.0	195	329.6	329.6	359.6	3.57	23,048	Pb/Sc	Tower size: 2.5 cm x 2.5 cm x 30 cm including readout 10cm Offset: measured from face nearest to interaction point Weight: estimated as 85% lead glass and 15% steel
	Service Gap			13.6			316	316	329.6				Offset: measured from location nearest to interaction point
	Dual RICH		6.10.04	120	14.0	180	320	200	320	10.47	1,946	Aerogel/Gas	Offset: measured from face farthest from the interaction point Volume: calculated as sum of the sub-sections Weight: based on parametric estimate from CLAS LTCC
		Detector Section		*94*	*14.0*	*180*	*226*	*226*	*320*	*9.51*			*Offset: measured from face nearest to interaction point*