SOFTWARE

EIC Software: Lessons Learned ^{3}

EICUG Software Working Group.

Working Together

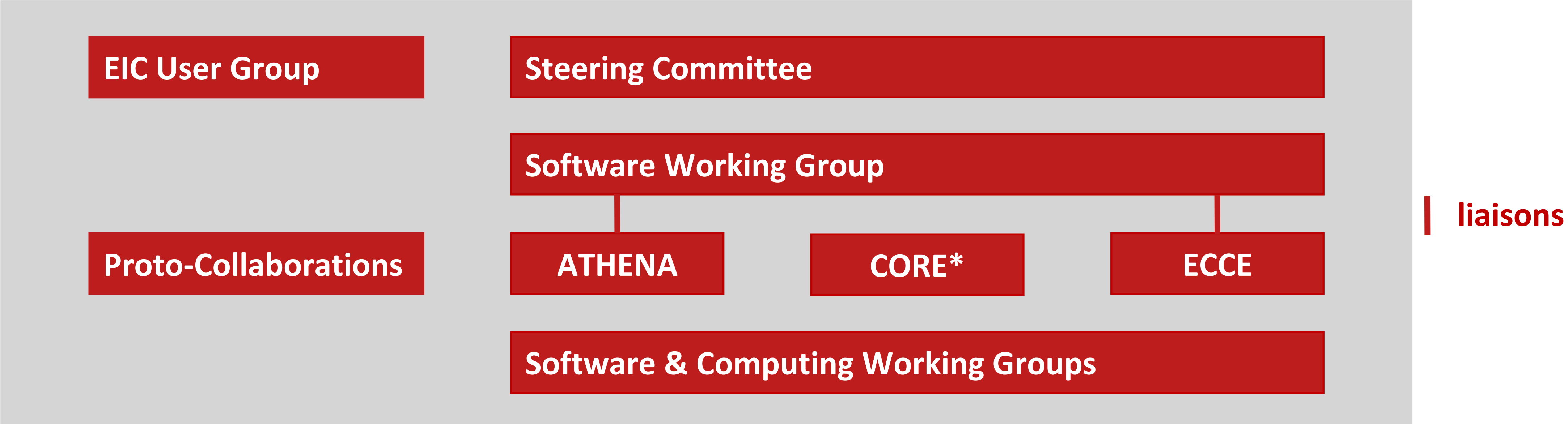
After a presentation on “Breakthroughs in Detector Technology”, Ian Shipsey (Oxford) was asked about the role of software.

Anecdote

"***Software is the soul of the detector,***” Ian Shipsey replied in a poetic way and emphasized the **importance of great software for great science**. He added that we need to **work together**, on a global scale and with other fields, to achieve this goal.

Common Software Effort

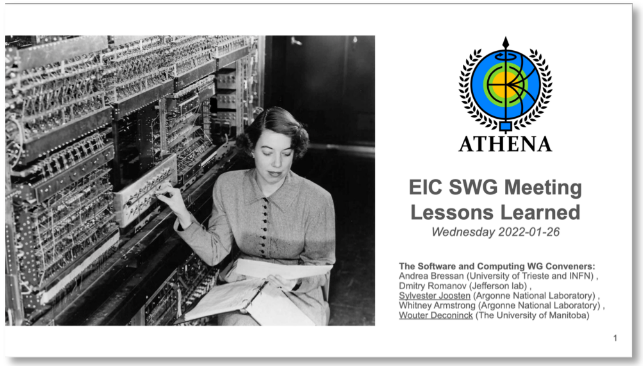
Thank you very much for working together with the EICUG SWG!



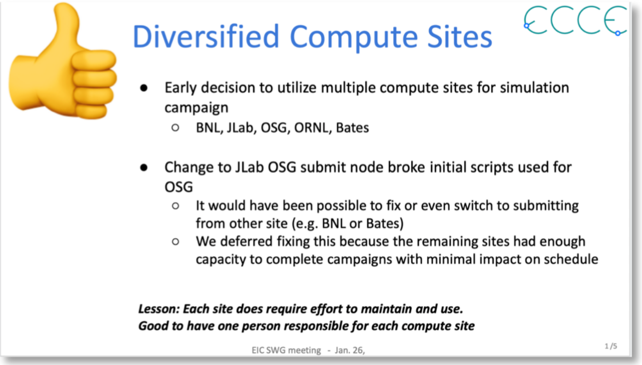
* CORE adapts existing software for their needs and has a far smaller software effort than other proto-collaborations.

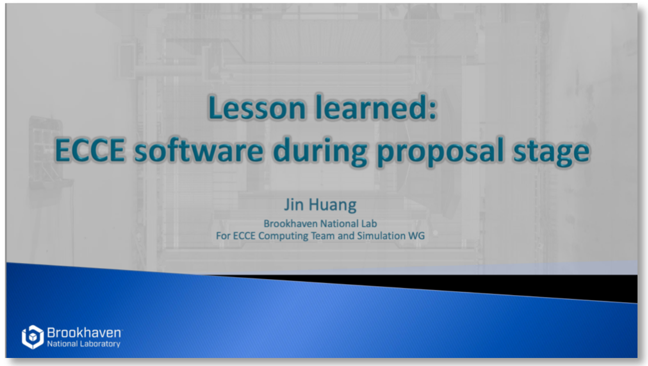
EIC Software: Lessons Learned (https://indico.bnl.gov/event/14319/)

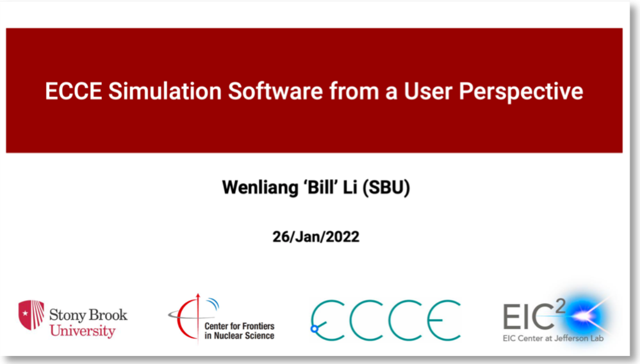
Lessons Learned from ATHENA (Sylvester Joosten, Wouter Deconinck)



Lessons Learned from ECCE (David Lawrence, Jin Huang, and Bill Li)







Software is in a very early life stage.

Work with the EIC community

- Both ATHENA and ECCE have been very successful in large-scale, detailed full detector simulations:
  - ATHENA have successfully developed a modular software stack based on common NHEP software.
  - ECCE have successfully leveraged familiar software.
- As this talk will describe, both the proto-collaboration software communities now agree – together with the EICUG SWG – on proceeding with work on one software stack.

Common Software

- Define requirements for EIC Software and **common software projects**:
  - Software needs of the EIC addressed in Software EoI and the resulting work plan for the EICUG SWG.
  - Evolve with the EIC community and the EIC project.

Right now, after the DPAP closeout and during the formation of the EIC collaboration(s), is the ideal time for doing so.

- Work together on **common software projects** based on these requirements.
  - Avoid duplication of the effort, e.g., workflows for distributed computing.
  - Team up on challenges, e.g., running on heterogeneous computing resources.
- Continue to build a **EIC Software community** with close connections and collaborations to the experts in NHEP:
  - ATHENA made a very deliberate choice to avoid “not-invented-here” syndrome and sits now at the table with HEP developers.

Expression of Interest for Software

1

Expression of Interest (EOI) for Software

Please indicate the name of the contact person for this submission:

Conveners of the Software Working Group:
- A. Bressan, M. Diefenthaler, and T. Wenaus
- eicug-software-conveners@eicug.org

Please indicate all institutions collectively involved in this submission of interest:

29 institutions

ANL	Argonne National Laboratory
BNL	Brookhaven National Laboratory
CEA/Irfu	IRFU at CEA /Saclay institute
EIC-India	Akal University, Central University of Karnataka, DAV College Chandigarh, Goa University, Indian Institute of Technology Bombay, Indian Institute of Technology Delhi, Indian Institute of Technology Indore, Indian Institute of Technology Patna, Indian Institute of Technology Madras, Malaviya National Institute of Technology Jaipur, Panjab University, Ramkrishna Mission Residential College Kolkata
IMP-CAS	Institute of Modern Physics - Chinese Academy of Sciences
INFN	Istituto Nazionale di Fisica Nucleare
JLab	Thomas Jefferson National Accelerator Facility
LANL	Los Alamos National Laboratory
LBNL and UC Berkeley	Lawrence Berkeley National Laboratory and University of California, Berkeley
NCBJ	National Centre for Nuclear Research
OhioU	Ohio University
ORNL	Oak Ridge National Laboratory
SBU	Stony Brook University
SLAC	SLAC National Accelerator Laboratory
SU	Shandong University

https://indico.bnl.gov/event/8552/contributions/43221/

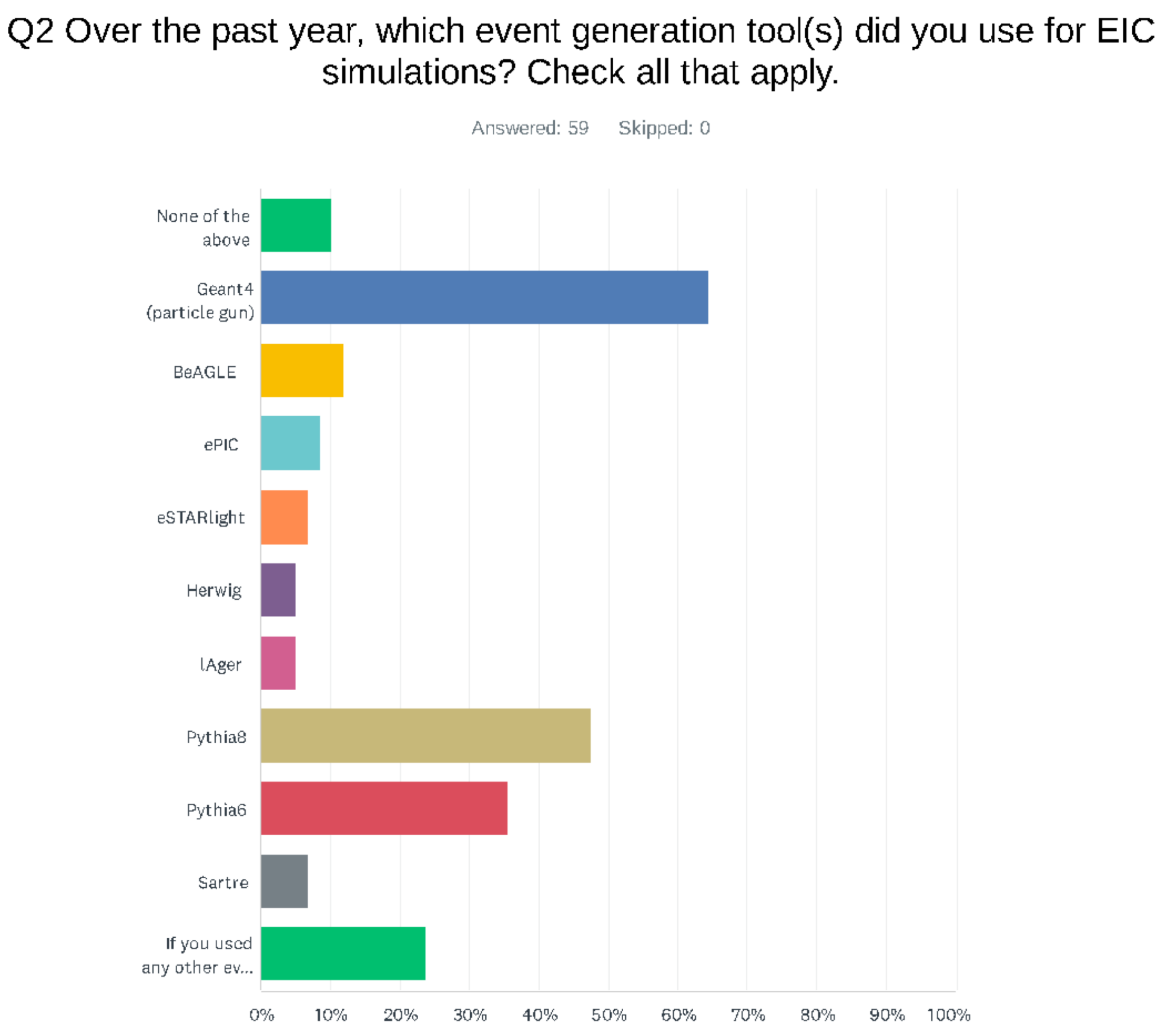
Common Projects

- **Software Tools for Simulations and Reconstruction**
  - Monte Carlo Event Generators
  - Detector Simulations
  - Reconstruction
- **Middleware and Preservation**
  - Workflows
  - Data and Analysis Preservation
- **Interaction with the Software Tools**
  - Explore User-Centered Design
  - Discoverable Software
  - Data Model

Future Technologies

- Artificial Intelligence
- Heterogeneous computing
- New languages and tools
- Collaborative software

Monte Carlo Event Generators



Other DEMPGen, Djangoh, elSpectro, TopHEG

We have successfully established a HEP standards: **HepMC3**

And have started with **Rivet**

for **MCEG validation** for the EIC.

We understand how to handle accelerator and beam effects in (and after) the event generation.

We have a vibrant community:

and are part of a community white paper on Event Generators for HEP Experiments, with EIC as part of cross-cutting aspects.

Detector Simulations

- **Detailed detector simulations based on Geant4**:
  - Various versions being used: ATHENA 10.7 (compatible with 11.0) and ECCE 10.6
  - EIC physics list from Project eAST, validation being started.
  - Lack of test beam data for the validation of the detector simulations.
  - Challenge: Reconstruction (**!**) of Cherenkov detectors (Geant4 describes Cherenkov radiation and optical physics very well).

- **Work focused on successful geometry integration**:
  - YR: Full simulation of detector components. → ATHENA and ECCE: Full simulation of detector concepts.
- **Geometry description and exchange**:
  - ATHENA: DD4hep. Geometry browser in the cloud (jsROOT).
  - ECCE: Pure G4 geometry for simulation -> TGeo for reconstruction.

- **Accelerate simulations**:
  - eAST: Full and fast simulations in Geant4. Sub-event level parallelism for heterogeneous computing.
  - Open question: What are the most promising applications for AI/ML?

- **ECCE**:
  - Detector design optimization using AI/ML.

Reconstruction

- **Enormous progress in reconstruction**:
  - Mainly for central detector though.
  - A lot of work needed towards 4D reconstruction for central detector and far-forward detectors, in particular for integrated reconstruction and PID.

- **Plethora of reconstruction algorithms being used**:
  - Reusing existing approaches:
    - Some lessons learned are not applied, e.g., not reusing event-level based reconstruction for dRICH.
  - But also new ones based on AI/ML.

- **ATHENA and ECCE**:
  - A lot of progress with ACTS, including EIC-specific contributions from ATHENA.

Distributed Workflows and Data Management

- **ATHENA deployed successfully automated workflows on eicweb**:
  - Workflows based on either slurm (Compute Canada, JLab) or HT-Condor (OSG).

- **ECCE used successfully git-based production system on batch systems**:
  - Notes that they could used fully developed solutions instead, e.g., PanDA.

- Both ATHENA and ECCE successfully used distributed computing resources.
- Both ATHENA and ECCE worked with the host labs on computing resources.

- **Scientific data management has been timesink during simulation campaigns**:
  - Rucio is used increasingly widely in HEP and drawing interest in NP (cf. recent round table).
  - Rucio has been discussed within EIC community. Requires support from host labs.

Data and Analysis Preservation

- **The ATHENA and ECCE workflows allow to reproduce results**:
  - Here, a key aspect is containerization.

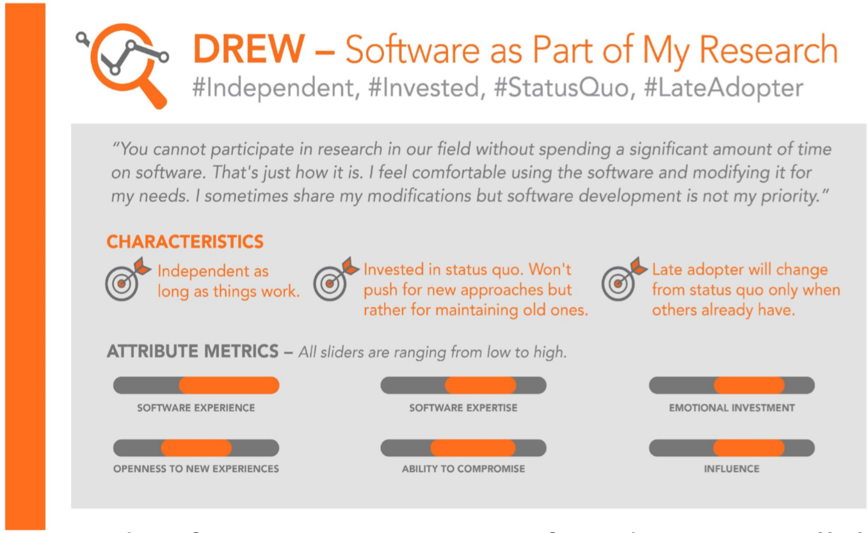
- **It pays to start early**:
  - e.g. create websites and documentation repositories for the long haul
  - There has been major progress on that by ATHENA and ECCE.

- **Beyond that, the EIC community needs to develop a strategy for data and analysis preservation**:
  - We need to work with the user community: Data and analysis preservation can only work with the majority of the community.

- **CERN, DPHEP, and collaborators have developed a suite of DAP tools that can contribute to a DAP strategy**:
  - HEPData, OpenData, Zenodo/InvenioRDM, REANA, InspireHEP, …

User-Centered Design: Listen to Users, and/then Develop Software

- **State of Software Survey**: Collected information on software tools and practices during the Yellow Report Initiative.
- As part of the State of Software Survey, we asked for volunteers for focus-group discussions:
  - Students (2f, 2m), Junior Postdocs (2f, 3m), Senior Postdocs (2f, 3m), Staff Scientists (2f, 3m), Industry (2f, 2m)
- **Results from the six focus-group discussions**:
  - Extremely valuable feedback, documented many suggestions and ideas.
  - Developed user archetypes with Communication Office at Jefferson Lab and UX Design Consultant:



User Archetypes: Input to software developers as to which users they are writing software for:

- Software is not my strong suit.
- Software as a necessary tool.
- **Software as part of my research.**
- Software is a social activity.
- Software emperors.

- We repeated Software Survey now after detector collaboration proposals:
  - The regular software census will be essential to better understand and quantify software usage throughout the EIC community..

Discoverable Software

- **Both ATHENA and ECCE have setup GitLab and GitHub organization for their software stack**:
  - (Major?) part of their repositories is available on the GitHub organization for the EIC.

- **ATHENA provided containers both on eicweb and Docker Hub and singularity via cvmfs**:
  - Spack environment to handle environment.

- **ECCE provided singularity containers**.

- **Both ATHENA and ECCE put an emphasis on the education of their user base**:
  - Various tutorials that have been well received.
  - Documentation for users.
  - Documentation for developers, e.g., Doxygen.

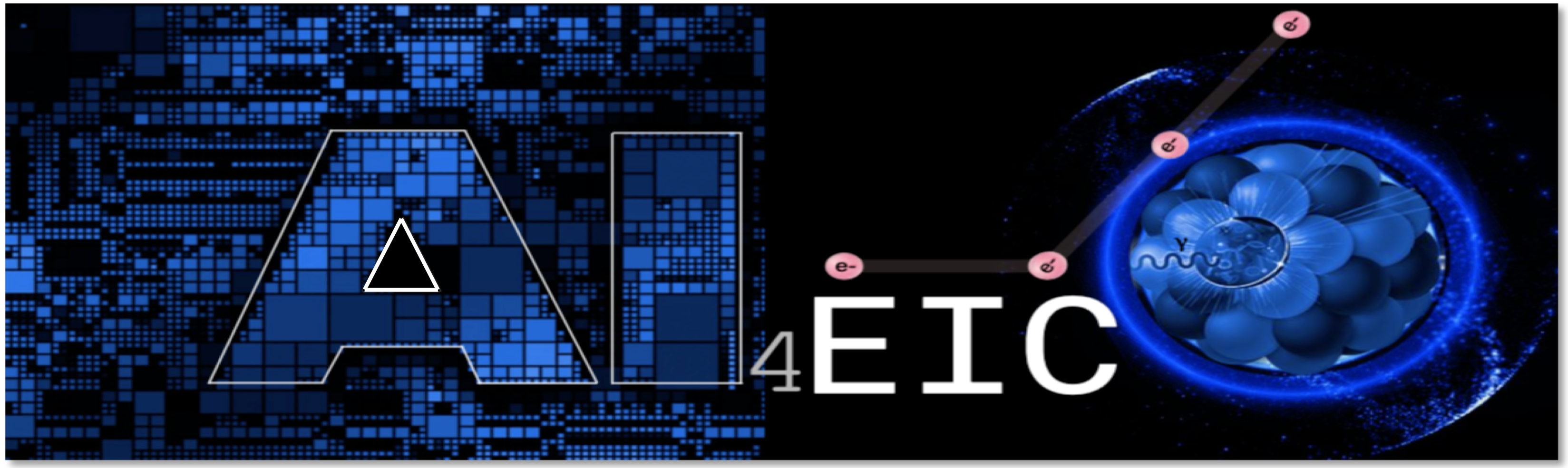
Data Model

- **Major progress**:
  - Both ATHENA and ECCE developed standardized ROOT files for their simulation campaigns.
  - Working with flat data structures paid off for development of physics analyses and first AI/ML algorithms.
  - ATHENA data model in eicd inspired from key4hep/EDM4hep (generic event data model that has been developed for future HEP collider experiments):
    - ATHENA uses now key4hep/EDM4hep directly.

- **Promising**:
  - ATHENA formulized the creation of flat data models using key4hep/EDM4hep.

Future Technologies

- **Artificial Intelligence**



https://eic.ai

- **Heterogeneous computing**

- **New languages and tools**
  - Both ATHENA and ECCE successfully deploy continuous integration.

- **Collaborative software**

From “Lessons Learned” to One Software Stack

Proceeding with work on one software stack:

- ATHENA and ECCE created very impressive and successful software efforts that delivered for their proposals.
- While ongoing efforts will naturally mean their existing software will continue to be used within their communities for a time, both the proto-collaboration software communities agree – together with the EICUG SWG – **on proceeding with work on one software stack**:
    - as expeditiously as the still evolving situation permits,
    - drawing from the good work and ideas of both proto-collaborations,
    - drawing on the common interest in common software and unified efforts,
    - looking towards and preparing for the software needs of the EIC.
- The attitudes and approaches of the ATHENA and ECCE teams make this a very attainable goal.

EIC Single Software Stack 2022

Link

Overview

Preparations of proposals for the DPAP decision resulted in multiple software stacks being developed. In the post-DPAP era the community is converging behind the "Detector-1" effort. As such, a Single Software Stack is needed that has input and endorsement from the entire Detector-1 community. In order to move quickly to this, a series of discussions/decisions are scheduled where the key pillars of the stack will be determined.

A procedure has been formalized so that these decisions can be made in a timely manner. Presentations can be found at the links below, but the summary is:

1. *Publicize schedule of topics with dates of discussion and decision*
2. *Assign chair for each topic. Chair will be POC for the topic. Responsibilities are:*
    1. *Organize discussion session agenda*
    2. *Publish draft list of requirements for the software being discussed at least 1 week in advance.*
    3. *Form list with at least one choice for the software to adopt to address the topic*
    4. *Collect suggestions for modifications to the requirements list and/or the software choices list*
    5. *Lead discussion on topic, starting with requirements list and the list of options*
3. *Presentations may be made regarding a specific decision topic, but should be communicated to discussion lead in advance for purposes of scheduling.*
4. *Use guiding principles from the EIC Software Statement of Principles*
5. *Discussion is required for all topics (formal presentations only as necessary).*
6. *Based on the meeting, the joint CompSW and SimQA WG conveners will propose a single option, which will be open for comments and endorsement for one week.*

n.b. Not every decision made will need to follow this procedure!. This is here only for the major decisions needed to get the single stack off the ground so the earlier stacks can be retired and the community united.

External Links

- Critical Path Procedure proposal (May 11, 2022 EICUG/CompSW meeting) https://indico.bnl.gov/event/15637
- Critical Path Procedure (May 13, 2022 General Detector-1 meeting): https://indico.bnl.gov/event/15784
- Critical Path Decision/Discussion Schedule (Google Spreadsheet)
- Software Decision Documents:
    1. Template
    2. Code Repository : Live Notes : Indico Meeting
    3. Geometry Package : Live Notes : Indico Meeting
    4. Data Model : Live Notes : Indico Meeting (day 1), (day 2)
    5. Reconstruction Framework : Live Notes : Indico Meeting (day 1), (day 2)
    6. Data and Analysis Preservation : Live Notes : Indico Meeting

Lessons Learned About EIC Software

2016 – 2020 **EIC Software Consortium (ESC)**

2018 – now **EICUG Software Working Group (SWG)**

2019 – 2021 **Yellow Report Initiative**

2021 – 2022 **Detector Collaboration Proposals**

2016 – now **Software & Computing Round Table**

2016 – now **Future Trends in Nuclear Physics Computing**

Statement of Software Principles

Link

We as a community have now an enormous amount of expertise and experience:

EIC Software: Statement of Principles

We aim to develop a diverse workforce, while also cultivating an environment of inclusivity and a culture of belonging.

We will have an unprecedented compute-detector integration:

- We will have a common software stack for online and offline software.
- We aim for autonomous alignment and calibration.
- We aim for a rapid turnaround of the raw data to online and offline physics analyses and do this in near real time.

Currently **being developed** by EICUG SWG.

Guiding principles to frame the discussion about requirements and resulting approaches and solutions.

Will be shared for further input with EICUG.

Planned Discussions:

- How to organize geometry and the exchange of geometry between detector simulation and reconstruction?
- Community goals and tools for software development.
- Workflow tools / frameworks for heterogeneous computing.

Many more to follow (planning ongoing).

Building a Software Community

Lessons Learned from workshop on "Future Trends in Nuclear Physics Computing":

- People are most important, not the software. Setting up an organization to create the right incentives to create and maintain the software.
- A strain repeated throughout workshop: career support!
- A way of supporting developers and their careers: software citations.
- Common software projects create a pool of highly valuable, valued developers who can carry expertise on a key tool to other experiments and communities, cf. career path.
- Management support up the hierarchy is important for successful open source project:
  - Acceptance of objectives wider than those of the home experiment.
  - Recognition of the value of the wider investment.
- Developers need the time and space to develop something new, not something just a little better.

Lessons Learned [3]

We have shared ”Lessons Learned”:

- From ATHENA and ECCE,
- From the EIC Software effort of the last years,
- including our workshop on ”Future Trends in Nuclear Physics Computing”.

We are now proceeding with work on one software stack:

- **“Statement of Software Principles”:**
  - Discuss in the next part of the lecture:
  - https://eic.github.io/activities/principles.html