EPIC SOFTWARE STACK

Sylvester Joosten
on behalf of
EPIC CompSW & SimQA

EICUG Meeting
July 28, 2022

EPIC CompSW
Software and Computing Conveners

Andrea Bressan
andrea.bressan@ts.infn.it

Cristiano Fanelli
cfanelli@mit.edu

Sylvester Joosten
sjoosten@anl.gov

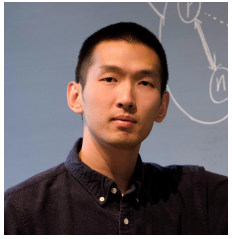
David Lawrence
davidl@jlab.org

EPIC SimQA
Simulation, Production, and Quality Assurance Conveners

Wouter Deconinck
wouter.deconinck@umanitoba.ca

Wenliang (Bill) Li
wenliang.li@stonybrook.edu

Joe Osborn
osbornjd@ornl.gov



Zhoudunming (Kong) Tu
zhoudunming@bnl.gov

CompSW Charge

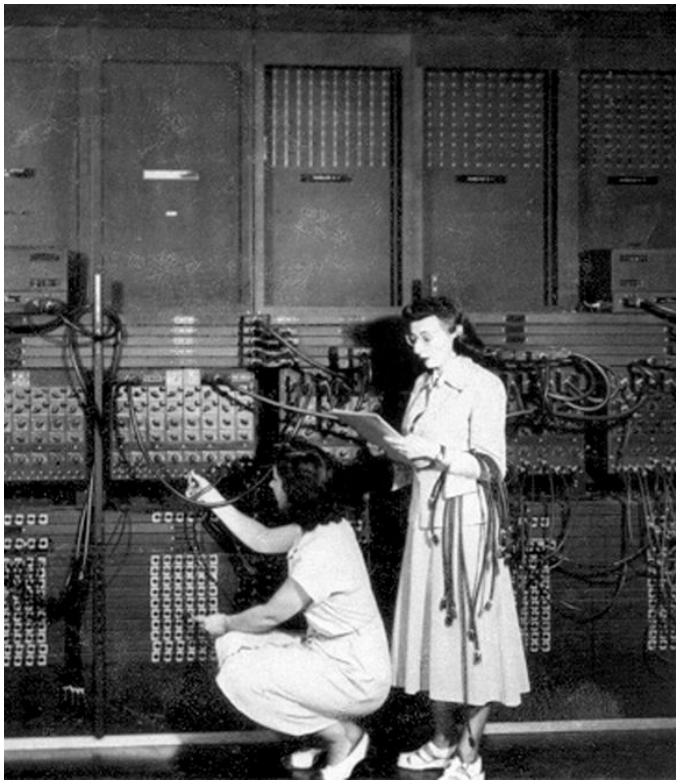
- Develop and curate a single **software** stack that will serve all aspects of the detector design and implementation regarding simulation, reconstruction, analysis, and the scientific program.
- Identify **computing** resource requirements to support the EIC program through detector development, construction, and production.
- **Foster** the development of new approaches to software and computing.
- **Support** all software and computing needs of the collaboration.

Detailed draft available here: https://docs.google.com/presentation/d/1aim32XWy6TpDpflK_YWnQhMkIgblOqv2Bh86k7278MM/edit?usp=sharing

SimQA Charge

We act as a **bridge** between the detector and physics working groups and the computing/software working group:

- **Simulation**: We provide technical guidance to PWG/DWG for simulation needs, help them i) with small-scale development/debugging, and ii) to be ready for large-scale productions.
- **Production**: We run/manage/coordinate large-scale simulation campaigns, bridging the demand and supply between PWG/DWG and the computing/software group.
- **QA**: We ensure communication and consistency among PWG/DWG on quality control of large-scale productions (e.g., 1 & 2).

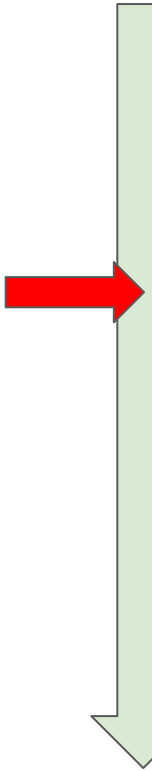


The need for a unified EPIC software strategy

- The proposal period saw a fragmented approach including different major frameworks and many smaller standalone projects.
- We need to unify our efforts to make the EIC detector a success, starting today throughout all CD milestones and into operation.
- We strongly believe in the EIC Software Statement of Principles, an effort of the entire EIC community under the umbrella of the EICUG.
- We will embrace these practices today to avoid starting our journey to EIC with technical debt.
- **We are writing software for the future, not the lowest common denominator of the past!**

How do we get there?

A critical path towards a unified software approach for the EIC



1. Assessment on the software solutions (pro & con list) together with the SimQA and DAQ working groups, guided by the EIC Software Statement of Principles.
2. Propose conclusion and recommendation to collaboration management and Project **by the Summer EICUG meeting**.
3. Software choice treated as any other technology choice? Optional independent review in the Summer**.**
4. Once decision is made, all new development should go in the official framework.
5. **Aim to have fully transitioned to the official software by October.**

Procedure to Travel the Critical Path

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:
   a. Organize discussion session agenda
   b. Publish draft list of requirements for the software being discussed at least 1 week in advance.
   c. Form list with at least one choice for the software to adopt to address the topic
   d. Collect suggestions for modifications to the requirements list and/or the software choices list
   e. 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.**

		Discussion topic(s)	Decision topic(s)	comments	Point of Contact
May	4	*AIWG*			
	11	Transition Period	Present procedure. Decide on list and order of decision topics		
	18	*No meeting (Streaming Readout X Workshop)*			
	25	Code Repository	Repository: - Location (GitHub, GitLab+Host) - Admins - Access		David Lawrence
Jun	1	Discussion Schedule	Schedule: - Decide most critical decisions to make before July 27th EICUG meeting - Schedule of topic discussions		
	8	Geometry	Geometry: - Package (e.g. DD4HEP)		Markus Diefenthaler
	15	Data Model	Data format - Generated events - Simulated data - Processed data (e.g. ROOT w/ specific tree format)		Whitney Armstrong
	22	Data Model			
	29	Reconstruction Framework	Reconstruction Framework - Package		Wouter Deconinck
Jul	6	Reconstruction Framework			
	13	Data and Analysis preservation	Data Preservation - What is preserved (simulated, DSTs, ...) - Location(s) - Access (S3, xrootd, rucio, ...)		Kolja Kauder
	20	Documentation	Documentation: - Location of User documentation (wiki, repository,...) - Who will set up skeleton with list of topics (e.g. "Getting Started")		
	27	*EICUG Meeting*			
Aug	3	Continuous Integration	Continuous Integration		
	10	Containerization Official builds	Containerization - platform (Singularity, Docker, multi, ...) - Supported OSes - Location of images (e.g. cvmfs) Official builds - Location (e.g. cvmfs, container image, ...)		
	17	Calibration DB Conditions DB	Calibration / Conditions DBs - Package - Server/Host - Access		

Code Repository

Convener Summary:

- We will implement a hybrid solution that uses GitHub as the primary repository, while using the eicweb GitLab instance for CI/CD.
- An ad hoc committee of eicweb experts will investigate the best option for leveraging CI/CD at ANL using GitHub (e.g. GitHub runners, mirrors, webhooks, etc.
- The existing "eic" organization at GitHub established by the EICUG Software WG will be used.
    - Some admin privileges will need to be shared with the Detector-1 WG conveners
- The best practices model for the repository will include:
    - Repositories will be open and public unless there is a specific reason to make them private
    - External packages will not be forked/cloned to the eic organization and modified unless under extremely exceptional circumstances.

Decision Document:
https://docs.google.com/document/d/1jT8CXj1cS9FEa0MbpJBV5jBA0T_vu2UdNDX93IJSo9c/edit?usp=sharing

Live Notes:
https://docs.google.com/document/d/1eOG5UnxFrBZajkucGgjq2ryPukI2yE5BPSmQOZ-GpGM/edit?usp=sharing

Indico:
https://indico.bnl.gov/event/15640/

Geometry Description and Detector Interface

Summary

We will implement the geometry description and detector interface using DD4hep.

Decision Document:
https://docs.google.com/document/d/16dQ-u2u5CdJIN3_slvcI79vTWJYnQytoQclMu2e-TpY/edit?usp=sharing

Live Notes:
https://docs.google.com/document/d/1C3KuUzRC6nXhCFlvjR2NV1fgqmt6MSuZKjqY-NtffM0/edit?usp=sharing

Indico:
https://indico.bnl.gov/event/16154/

Data Model

Convener Summary:

- We will adopt PODIO as the tool for managing the EDM.
- We will adopt the EDM4hep Data model as the initial Data Model

Decision Document:
https://docs.google.com/document/d/1sddpfySwvkRSLgpUq4NgnFlw7tLdHc6ofT4trwV45xA/edit?usp=sharing

Live Notes:
https://docs.google.com/document/d/1seWDXQr570Tv_yJijUCKgXhla6u5HAV-ibO82PR43Xk/edit?usp=sharing

Indico:
day1: https://indico.bnl.gov/event/15642/
day2: https://indico.bnl.gov/event/15643/

Framework

Convener Summary:

The working group conveners recommend JANA2 as the reconstruction framework.

Although both Gaudi and JANA2 are technically able to meet the requirements, there is too much risk in depending on Gaudi which is not focused on a community outside of LHCb. The efforts invested in the already written juggler algorithms will be able to be reused with relatively minor effort in JANA2 algorithms. Likewise, the efforts invested in fun4all algorithms may be able to be reused.

The translation of the relevant Gaudi/juggler and fun4all algorithms will be completed by the Jefferson Lab EPSCI group by October 2022, aided by the 1 FTE-year per year committed by Jefferson Lab to the support of JANA2 for EIC.

The working group conveners point out that continued engagement with the key4HEP project through the development of modular reconstruction algorithms and functional programming approaches is desirable.

The Jefferson Lab EPSCI group is encouraged to develop JANA2 into a community project where developers from outside Jefferson Lab are valued at all stages of software development.

Decision Document:
https://docs.google.com/document/d/1lomak02ztchkwQB2d_f-58gabBOQF9WaPaQhf8kTvfY/edit

Live Notes:
https://docs.google.com/document/d/1IdlQ63PxfIDsGdOlkik0OE76EzIpHCt00Br583hOJxI/edit?usp=sharing

Indico:
day1: https://indico.bnl.gov/event/15644/
day2: https://indico.bnl.gov/event/15645/

Framework transition timeline

Yesterday: ● Simulations either in fun4all or DD4hep+Gaudi/Juggler or standalone
Today: ● Implementation of Podio and DD4hep services in Jana2 ● Short-term simulation still served by DD4hep+Gaudi/Juggler, or fun4all. ● Focus short-term algorithm development on DD4hep+Gaudi. ● Stop major developments in Juggler unless there is a clear transition strategy
August: ● first training sessions on the DD4hep+Jana2 stack? ● Start transitioning and validating algorithms (pulling from fun4all, Juggler, and many standalone algorithms)
October: ● full switch to Jana2

October is coming together, but what about today?

- ✅ 🚧 Implementation of EPIC detector in DD4hep
- ✅ Data model based on (somewhat extended EDM4hep)
  - 🚧 Needs further fleshing out for more advanced simulations, and need to define pathway to collaborate with Key4hep
- ✅ Short-term critical simulations to be done in either fun4all stack or DD4hep+Gaudi stack (e.g. background embedding).
  - 🚧 Algorithm porting will be supported by JLab crew to facilitate the transition.
- 🚧 Porting of CI infrastructure and core repos from eicweb to GitHub.com ongoing
- 🚧 JANA2 transition:
  - Base components to integrate with detector geometry (DD4hep) and data model (PODIO/EDM4hep) being written, based on the infrastructure available in Gaudi/Juggler.
  - Comprehensive suite of reconstruction algorithms in Gaudi and fun4all being ported to Jana2.
  - Aim to keep some level of algorithm interoperability with Key4hep.
- 🚧 Tutorials for the “final stack” starting in August.

(Personal) thoughts on the role of the EICUG SWG

Different boundary conditions:

- Collaboration has strict timeline with well-defined deliverables from the short to the long term.
- EICUG SWG more room for creative open-ended efforts with a long horizon.

Ideal platform for cross-collaborations

- Experiment-theory collaboration.
- Inter-experiment coordination.
- Interdisciplinary collaborations (data scientists, high-energy physics, …)

Summary

- Journey to our unified EPIC software stack well underway.
- All major topics have been addressed!
- Preparing for software review by the Project late this Summer.
- Aim to fully transition to the unified stack by October, in time for the second large simulation campaign.


EPIC **Questions?**