EPIC Computing and Software

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

*"**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.

A. Bressan, W. Deconinck, M. Diefenthaler,

C. Fanelli, S. Joosten, W. Li, J. Osborn, and Z. Tu

on behalf of the EPIC Computing and Software WG.

Working Groups Structure

Based on discussions in the WGs and the collaboration.

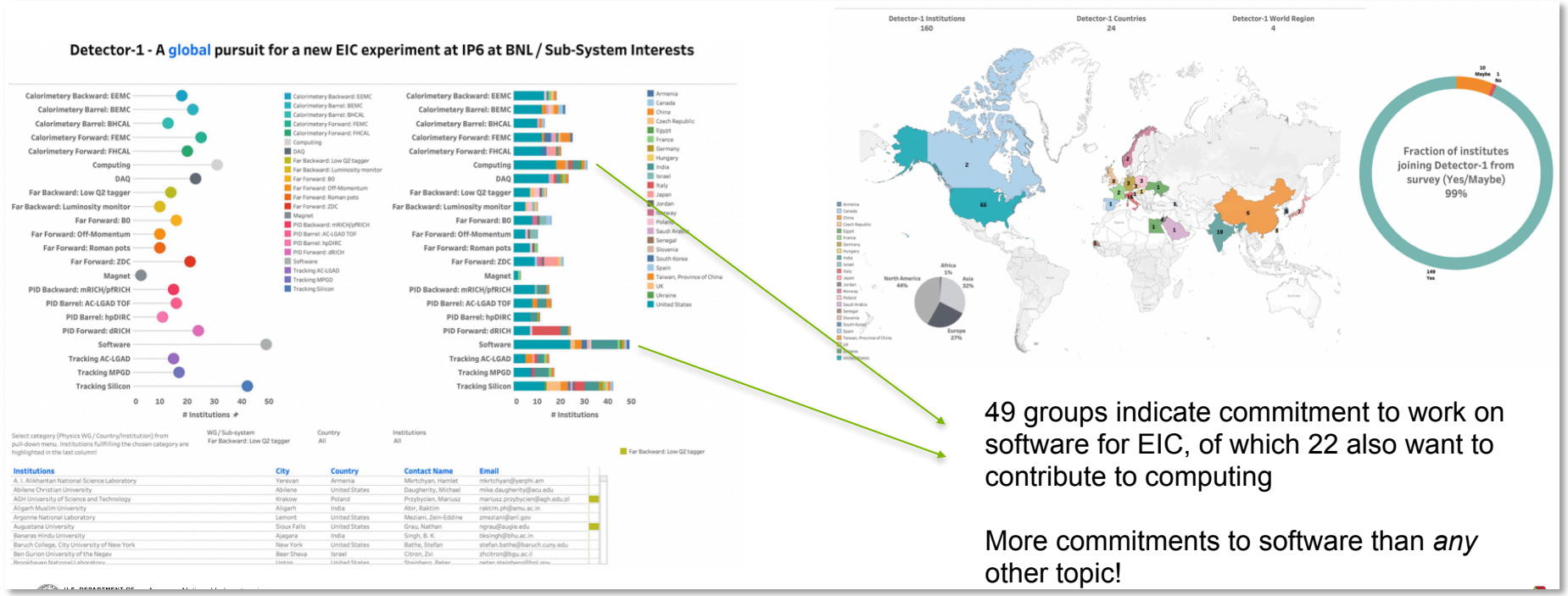
We are the **EPIC Computing and Software** WG!

Why have we merged the EPIC CompSW and SimQA WGs:

- Overlap in topics between the WGs.
- Our goal of not separating software development from software use and support. This is motivated by:
  - Lessons learned from successful scientific software projects.
  - The EIC Software: Statement of Principles:

> 8 **We will provide a production-ready software stack throughout the development:**
>
> - We will not separate software development from software use and support.

Involvement from EPIC Institutions



- In Q1, we will offer online tutorials for onboarding (**in preparation for Indian / Asian timezone**).
- We will reach out to EPIC Institutions about their software and computing interests and will point to opportunities for shared leadership and responsibilities (**ongoing**).
- Of course, we welcome any institution that is proactive and reaches out to us as, e.g., the institutions from India.

Subgroups: Shared Leadership and Responsibilities

- Containers/infrastructure/dependencies/spack
- MCEGs
- Detector Simulations
- Digitization / Streaming Readout Simulation
- Reconstruction
- Physics Algorithms
- Framework
- Large-Scale Simulations / Simulation Campaigns
- Workflow Tools and Environment

- AI/ML
- Heterogeneous Computing

- Training and Documentation
- User Support

- Data and Analysis Preservation

List of possible subgroups

Keep one working group:

- **For coordination and communication**
- **But also for agile organization**

Priorities for 2023

Based on WG discussions.

Priorities for the working group itself: Next Priorities
- Build a large and active working group with shared responsibilities and leadership
- Portal to get starting with using simulations and analysis as well as software development

Priorities for simulations and analysis: Next Priorities
- Reconstruction for full integrated detector for holistic understanding of the detector and its capabilities
  - Including PID and backgrounds
  - Enforcing modularity for clear separation of development of reconstruction algorithms and development of the framework and its services
- Reproducible workflows for simulation and analysis
  - Building up on the work on continuous integration with tests and benchmarks
  - Start including handling of metadata (e.g., conditions database)
- Simulations of eA in addition to ep
- Simulations of streaming readout

Priorities to advance our science:
- Start incorporating AI/ML methods and approaches in our software stack
- Use heterogeneous nodes at BNL, Jefferson Lab, and other facilities as testbeds for start rolling out solutions for heterogeneous computing challenges
- Work with the collaborations towards fully reproducible, re-usable, and re-interpretable analyses as a collaboration standard

Priorities for software development:
- Debugging in containers
- Support Jupyter notebooks for analyses through documentation and examples

Next Simulation Campaign and Related Task Forces

- **Next simulation campaign**:
  - **Target date**: End of March.
  - **Goal**: Improved software stack for the reconstruction, including benchmarks.

tbd

- **Task forces** to reach our goal:
  - **Clustering**
  - **Jet Reconstruction**
  - **PID**
  - **Tracking**
  - **EICRecon** for urgent fixes to the reconstruction software stack
  - **Modular Reconstruction** for substantial improvements of the reconstruction software stack
  - **Simulation Campaign Preparation**

Important note: Task forces != subgroups

- Task forces solve specific issues.
- Subgroups based around interest in a general area, e.g., reconstruction. Will form later.

- Each task force will be responsible for leading the effort for a particular topic with the following goals:
  - The development is accessible to the whole collaboration in our main repository.
  - We can evaluate the reconstruction quality using a set of well-defined plots that we can easily reproduce.
  - Kickoff meeting with task force leaders tomorrow, January 27.

What are we changing the reconstruction software stack?

- **Starting point**:
  - Software decision process end of August 2022, only 5 months ago.
  - Many components developed from scratch.
- **Main issues during first simulation campaign**:
  - Inconsistent treatment of units between Python and C++ layer in EICrecon.
    - E.g.: `100 MeV` specified in Python, passed as '`100`' to C++ and interpreted as `100 GeV`.
    - Python layer has been removed to avoid inconsistency.
  - Imaging calorimeter reconstruction was considered out of scope.
  - Unclear boundary between roles of DWGs (development of reconstruction algorithms) and SWGs (development of reconstruction framework).
- **Lessons learned**:
  - Enforce modularity; separate the development of:
    - Framework and its services (driven by computing needs(, major contributions from software experts,
    - Reconstruction algorithms (driven by physics needs), major contributions from detector experts.
  - Validation of reconstruction algorithms with detector experts needed.
  - Developing policies for software development to make it easier to develop software as a collaboration.

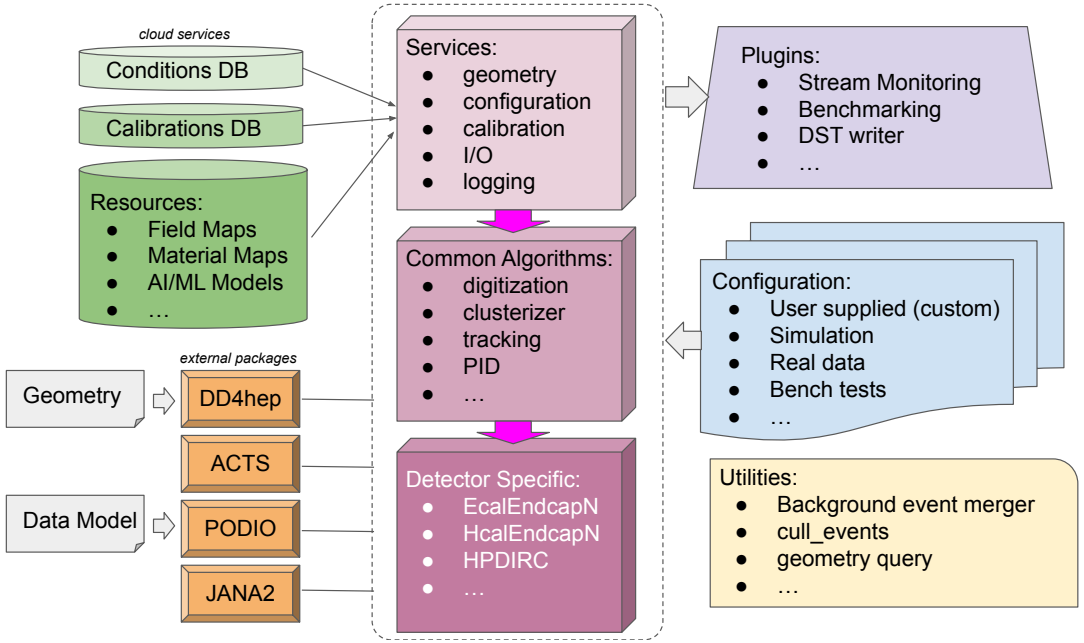
EPIC Reconstruction: Next Steps

Dec. 7 Discussion on reconstruction Software Stack (Indico, Live Notes)

Feb. 1 Follow up meeting with presentation on prototype

Two presentations

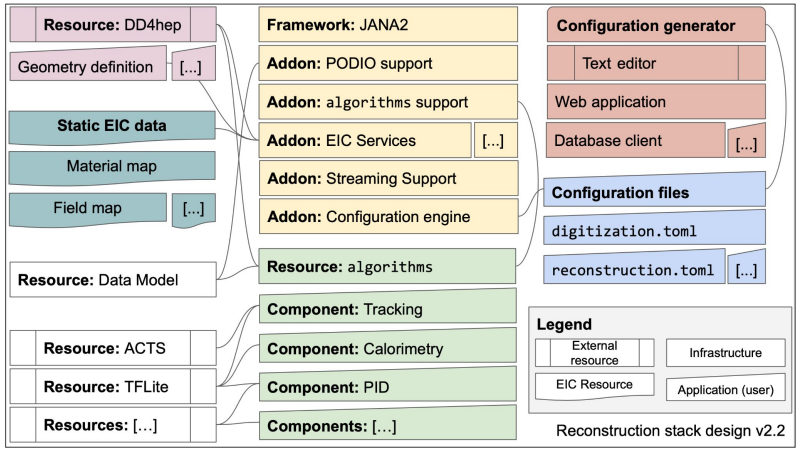
David – EICrecon Refactoring



Sylvester – `algorithms`: The case for framework- and experiment-independent algorithms at EPIC

A strawman re-design of the EPIC reconstruction stack

- Strictly modular approach reduces scope of each component
- Easier to onboard new users in any singular piece of the stack
- Every user can find their place based on experience and needs
- Better maintainability and more resilient against changing software needs
- Baked-in reproducibility by enforcing configuration files in every workflow



6

Statement of Software Principles (SP) on Reconstruction Algorithms

- **SP3** We will have a modular software design with structures robust against changes in the computing environment so that changes in underlying code can be handled without an entire overhaul of the structure.

- **SP4** We aim for a modular development paradigm for algorithms and tools without the need for users to interface with the entire software environment.

- **SP7** We will ensure that mission critical software components are not dependent on the expertise of a single developer, but managed and maintained by a core group.

- **SP8** We will provide a production-ready software stack throughout the development.

We will support the current version of EICRecon while re-factoring is in progress:

- We aim to make transition smooth: the changes are in the backend and infrastructure, not in algorithms themselves.
- Simulations for the BEMCal and bRICH reviews will be based on the current version of EICRecon (with needed fixes).
- The next simulation campaign might already use the the re-factored version with major improvements.

Summary and Discussion

- We have a **modular simulation, reconstruction, and analysis toolkit** for the development of the our detector and science program.
- The toolkit is based on the '**Statement of Software Principles**' and a **decision-making process involving the whole community**.
- We have **merged** the **EPIC CompSW** and **SimQA WGs into** the **EPIC Computing and Software WG** to not separate software development from software use and support.

- We are **restructuring the WG** to build a large and active WG with shared leadership and responsibilities:
  - Tutorials for onboarding in preparation.
  - We have started to reach out to institutions.
- We have set our **priorities for 2023** and have started on work on an improved software stack for reconstruction:
  - Task forces on clustering, jet reconstruction, PID, and tracking algorithms. Also on EICRecon, modular reconstruction, and simulation preparation.
- We propose **end of March** as target date for the **next simulation campaign**.