# EPIC Computing & Software and Simulation, Production, QA

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**

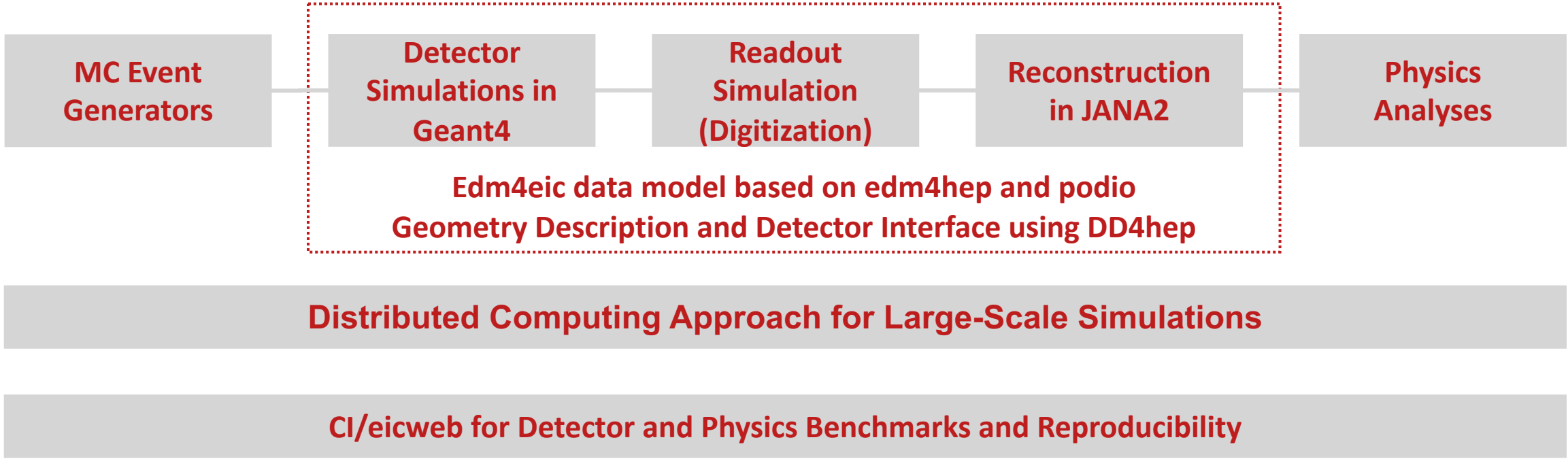
**for the EPIC Computing & Software and Simulation, Production, QA WGs.**

# EPIC Software & Computing

Clear mandate:  Commonality; one software stack!

The design of the **modular simulation, reconstruction, and analysis toolkit** for the development of the EPIC detector and science program is based on the **Statement of Software Principles** and a **decision-making process** involving the whole community. The decisions have been well reviewed in the **EIC Software Infrastructure Review**.

## Simulation, Reconstruction, and Analysis



# EIC SOFTWARE:
# Statement of Principles

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

# Principle 2: Compute-Detector Integration

**2 We will have an unprecedented compute-detector integration:**

- We will have a common software stack for online and offline software, including the processing of streamed data and its time-ordered structure.
- We aim for autonomous alignment and calibration.
- We aim for a rapid, near-real-time turnaround of the raw data to online and offline productions.

# Principle 3: Heterogeneous Computing

**3 We will leverage heterogeneous computing:**

- We will enable distributed workflows on the computing resources of the worldwide EIC community, leveraging not only HTC but also HPC systems.
- EIC software should be able to run on as many systems as possible, while supporting specific system characteristics, e.g., accelerators such as GPUs, where beneficial.
- 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.

# Principle 4: User-Centered Design

**4 We will aim for user-centered design:**

- We will enable scientists of all levels worldwide to actively participate in the science program of the EIC, keeping the barriers low for smaller teams.
- EIC software will run on the systems used by the community, easily.
- We aim for a modular development paradigm for algorithms and tools without the need for users to interface with the entire software environment.

# Principle 5: Open, Simple, and Self-Descriptive Data Formats

**5 Our data formats are open, simple and self-descriptive:**

- We will favor simple flat data structures and formats to encourage collaboration with computer, data, and other scientists outside of NP and HEP.
- We aim for access to the EIC data to be simple and straightforward.

# Principle 6: Reproducible Software

**6 We will have reproducible software:**

- Data and analysis preservation will be an integral part of EIC software and the workflows of the community.
- We aim for fully reproducible analyses that are based on reusable software and are amenable to adjustments and new interpretations.

# Principle 7: Community

**7 We will embrace our community:**

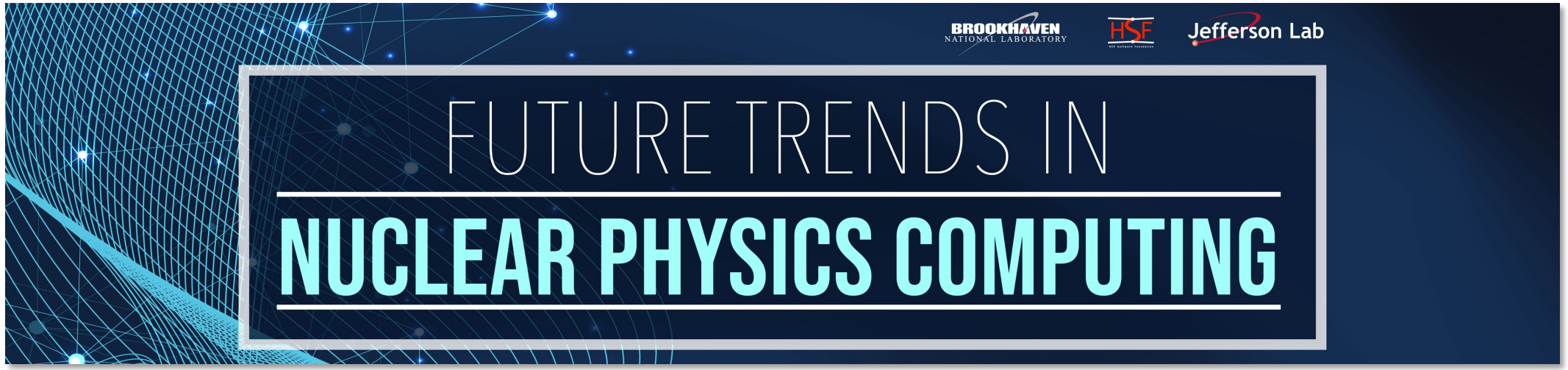
- EIC software will be open source with attribution to its contributors.
- We will use publicly available productivity tools.
- EIC software will be accessible by the whole community.
- We will ensure that mission critical software components are not dependent on the expertise of a single developer, but managed and maintained by a core group.
- We will not reinvent the wheel but rather aim to build on and extend existing efforts in the wider scientific community.
- We will support the community with active training and support sessions where experienced software developers and users interact with new users.
- We will support the careers of scientists who dedicate their time and effort towards software development.

# Principle 8: Development and Operation

**8 We will provide a production-ready software stack throughout the development:**

- We will not separate software development from software use and support.
- We are committed to providing a software stack for EIC science that continuously evolves and can be used to achieve all EIC milestones.
- We will deploy metrics to evaluate and improve the quality of our software.
- We aim to continuously evaluate, adapt/develop, validate, and integrate new software, workflow, and computing practices.

# Successful Scientific Software: Lessons Learned




**Common Scientific Software – Keys to Success**

Summary by Paul Laycock and Torre Wenaus

- **The team is the most important** Do not separate development and operations, both ACTS and Rucio benefited from experience with developing and operating a worse software package, crucial experience. Developers keen to use modern software paradigms, open-source and open-minded, proactively searching out best practice and adopting it.
- **The project** Clear, well-focused short-term goals are important, grounded in real-world deliverables. Aligned with the long-term plan of building something sustainable and designed to be used by outside collaborators.
- **The management** Accept that the long-view takes longer to deliver the short-term product, manage expectations of the collaboration and funders to ensure the team have sufficient time and space to succeed.

# EPIC CompSW and SimQA WGs

We have started to merge the EPIC CompSW and SimQA meetings.

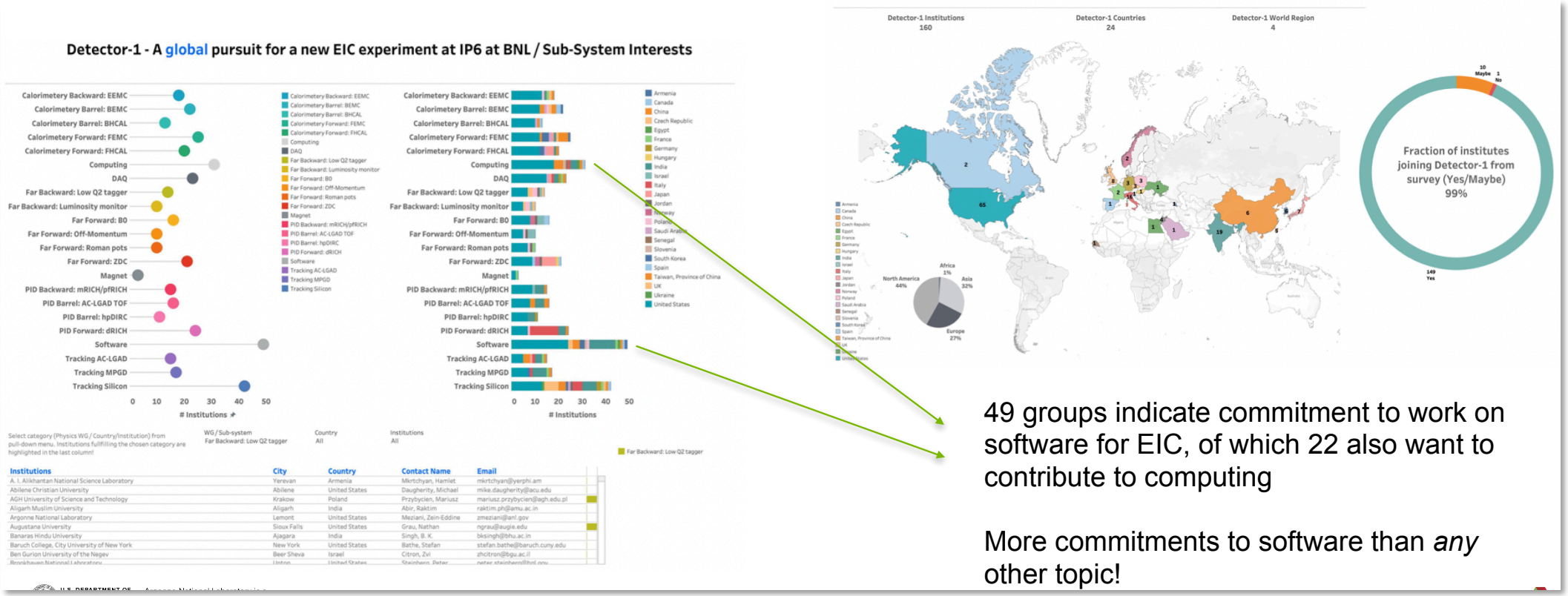
This reflects:

- The overlap in topics of the CompSW and SimQA WGs.
- Our goal of not separating software development from software use and support. This is motivated by:
  - Lessons learned from successful scientific software projects (previous slide).
  - The Statement of Software Principles:

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

We intend to **merge** the **EPIC CompSW and SimQA WGs**.

# Involvement from EPIC Institutions




- In Q1, we will offer online tutorials for onboarding.
- We will reach out to EPIC Institutions about their software and computing interests and will point to opportunities for shared leadership and responsibilities.
- 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 (that will evolve over time)

Based on WG discussions.

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

**Priorities for simulations and analysis:**
- 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

# Communication with GD/I

In an effort to improve horizontal communication,

Joe Osborn has taken on a convener role in both CompSW/SimQA and GD/I

**GD/I Charge**

- Work with the project and the joint working groups to develop a detailed, integrated technical design of the project detector. This includes the integration of various detector systems, the necessary supports and services, and the requirements imposed by the ability to service the detector between EIC running periods.
- Work with the detector and physics working groups, as well as project management, to ensure that the integrated project detector remains capable of the full science program outlined in the EIC white paper and NAS report. Where compromises need to be made in the integration of the project detector, ensure that the proper simulations studies are completed to ensure they do not unduly compromise the EIC science program.

**CompSW/SimQA Charges**

- Develop and curate a single software stack that will serve all aspects of the detector design and implementation regarding simulation, reconstruction, analysis, and the EIC scientific program
- …provide technical guidance to PWG/DWG for simulation needs…
- Run/manage/coordinate large-scale simulation campaigns…
- Ensure communication and consistency among PWG/DWG on quality control of large-scale productions

# Horizontal Communication Need

- GD/I needs guidance from simulation on issues that can only be addressed by simulation (e.g. integration)
- Simulation needs guidance from GD/I on implementation (e.g. two detector configurations in November simulation campaign)

# Horizontal Communication Need

- GD/I needs guidance from simulation on issues that can only be addressed by simulation (e.g. integration)
- Simulation needs guidance from GD/I on implementation (e.g. two detector configurations in November simulation campaign)
- We need horizontal communication to ensure simulation campaigns accurately reflect (any) global integration issues, and that integration issues from simulation are made aware to the broader collaboration

# CompSW/SimQA Liaisons

- We have a (mostly) complete list of liaisons to the subsystems and/or working groups
  - https://github.com/eic/epic/pull/64

- However, we have a less complete list of liaisons to the physics working groups
- We also (in general) do not have great attendance to the software meetings from these liaisons
  - How to improve? Better structure within our WG? Clearer definitions of roles/responsibilities?

# Lessons Learned from October

- There is poor communication between the detector menagerie and the simulation configuration
  - Need improved communication in both directions
    - Bare minimum: machine readable input/output to make direct comparisons between menagerie and simulation
- Need embedded machine background simulation capabilities in place
  - New task force meeting at 9:30 AM EST on Fridays to discuss background embedding progress (more on this from Elke Aschenauer this morning and Rey Cruz-Torres tomorrow)
- Need continued development of supports and services (ongoing discussions in GD/I)

# 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 intend to **merge** the **CompSW** and **SimQA WGs** 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.
- We have set our **priorities for 2023**.

- We aim for **better communication across working groups**, e.g., with:
  - Well defined liaison roles/responsibilities,
  - Better defined structure of CompSW/SimQA.
- Future integration related issues will require **continued development between CompSW/SimQA and GD/I**.