Embarking on a New Phase

Software & Computing

- **Software & Computing** play an ever-growing role in modern science, including NP, HEP, and related fields.

- As new experiments commence and data volumes rapidly increase, the NP community is exploring the **next generation of data processing and analysis workflows to optimize scientific output**:
  - This includes **streaming readout**, **AI/ML**, and **common scientific software**.

- To achieve our goals for next-generation software and computing for NP, we must **work together globally** and **across various fields**.

- The **EICUG Software Working Group** has played a key role in building an **active software community** for the EIC. Moreover, it has successfully fostered strong connections with the software communities in NP and HEP.

EIC Software Efforts

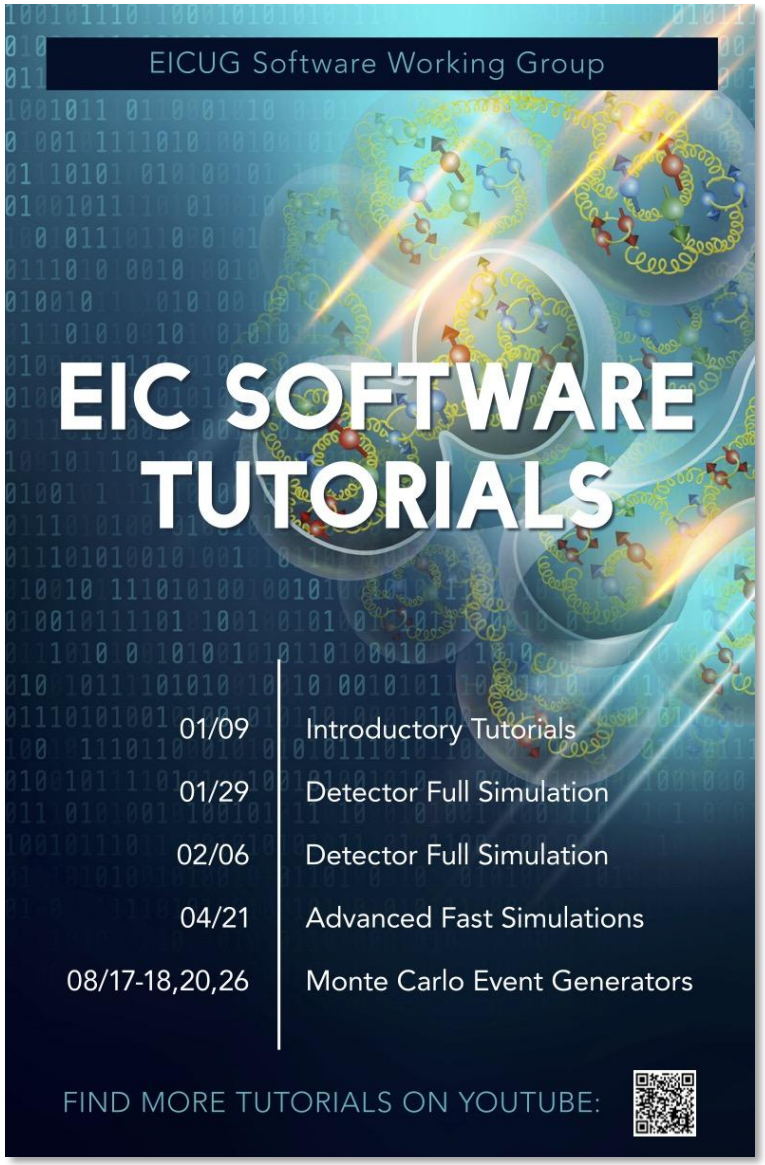
2016 – 2020	EIC Software Consortium (ESC)	
2018 – now	EICUG Software Working Group (SWG)	
2019 – 2021	Yellow Report Initiative	← Supported by SWG
2021 – 2022	Detector Collaboration Proposals	← Supported by SWG
2018 – now	ePIC Collaboration	← Formation supported by SWG
2016 – now	Software & Computing Round Table	
2016 – now	Future Trends in Nuclear Physics Computing	

EICUG SWG Phase I: Charge

The EICUG Software working group's **initial focus will be on simulations of physics processes and detector response** to enable quantitative assessment of measurement capabilities and their physics impact. This will be pursued in a manner that is **accessible, consistent, and reproducible to the EICUG as a whole.** It will embody simulations of all processes that make up the EIC science case as articulated in the White-paper. The Software working group is to engage with new major initiatives that aim to further develop the EIC science case, including for example the upcoming INT program(s), and is anticipated to play key roles also in the preparations for the EIC project(s) and its critical decisions.

The working group will build on the considerable progress made within the EIC Software Consortium (eRD20) and other efforts. The evaluation or development of experiment-specific technologies, e.g. mass storage, clusters or other, are outside the initial scope of this working group until the actual experiment collaborations are formed. **The working group will be open to all members of the EICUG to work on EICUG related software tasks**. It will communicate via a new mailing list and organize regular online and in-person meetings that enable broad and active participation from within the EICUG as a whole.

Common Scientific Software



Software Community Building

- The **EICUG SWG** has played a key role in building an **active software community** for the EIC.
- **Participation in HEP Software Foundation** Involvement in MC event generators, frameworks, reconstruction and software triggers, training.

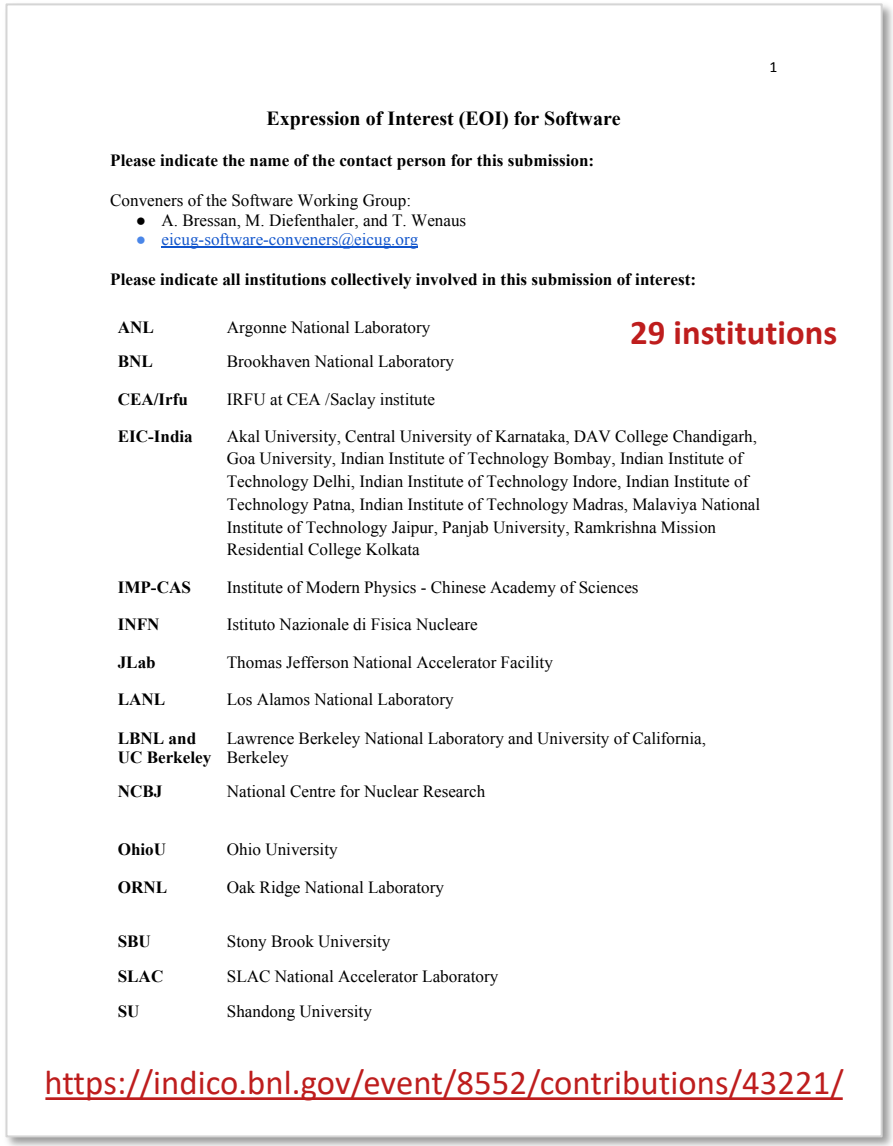
Common Scientific Software: Lessons Learned

- **The team is the most important**: Do not separate development and operations.
- **The project**: Clear, focused short-term goals should align with a sustainable long-term plan that accommodates external collaborators.
- **The management**: Manage expectation to allow the team enough time to achieve success.

Scientific Software Careers Need Support

- Support for education and training in software development.
- Provide career paths and funding that allow for and value software development.

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:

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

29 institutions

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

Common Projects

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

Future Technologies

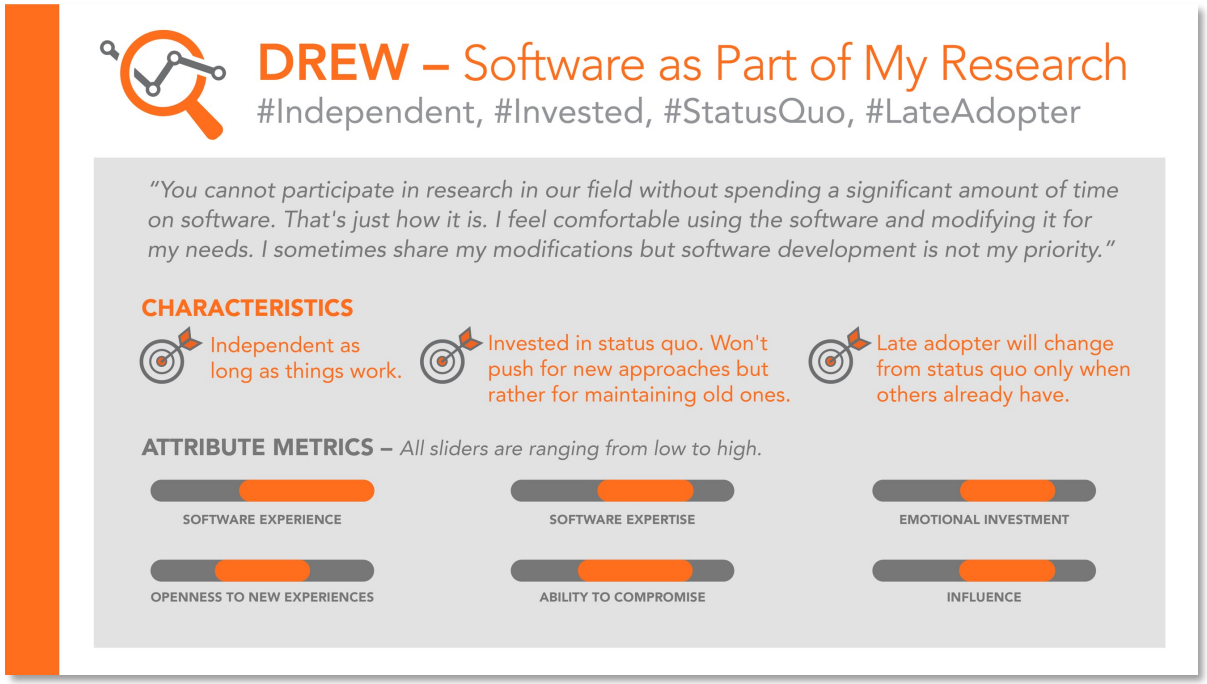
- Artificial Intelligence
- Heterogeneous computing
- New languages and tools
- Collaborative software

Gaining Insight Into the Community

- **Goal**: Enable active participation in physics analysis, regardless of career stage, beyond just students and postdocs.
- **Survey**: On average, 78% of students' and postdocs' research time is devoted to software and computing.

→ **User-Centered Design**: Engage community in development. Listen to users, then develop software.

- User archetypes developed on feedback from focus group discussions.
- Input to software developers as to which users they are writing software for:



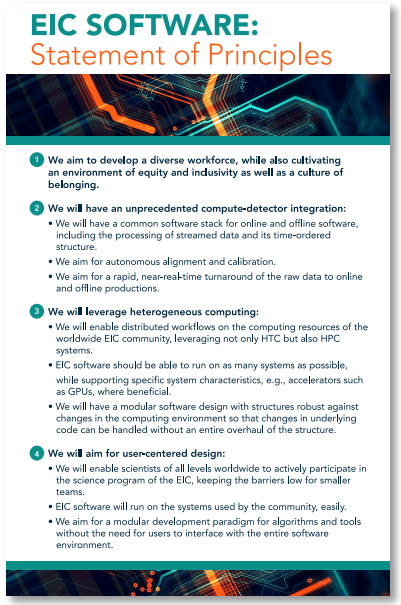
User Archetypes

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

One Software Stack for the EIC

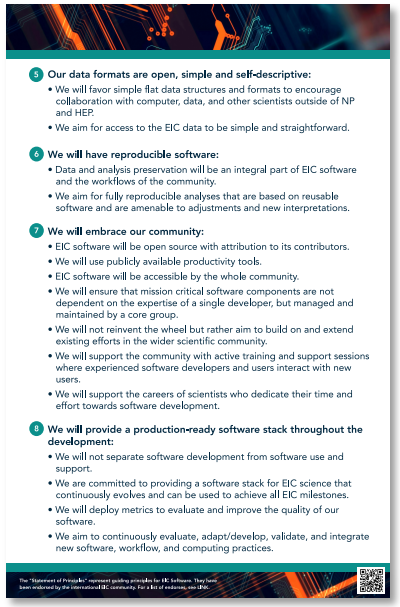
- **How to decide on our software stack?**
  - How do we ensure we work towards to our vision for EIC Software?
  - How do we ensure we meet the needs of the EIC community?

- **Solution: Statement of Principles**
  - Community process to define guiding principles for EIC Software.
  - Guiding principles define the requirements for EIC Software.
  - Endorsement by the international EIC community.



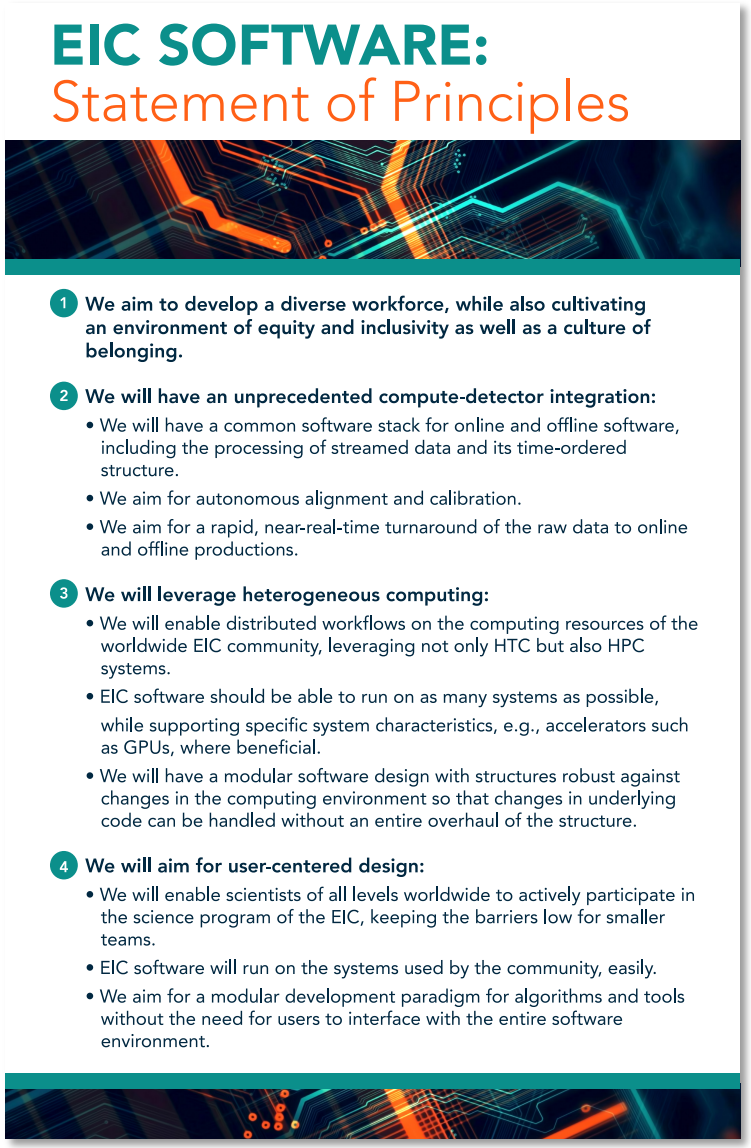
EIC SOFTWARE:
Statement of Principles

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



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.
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.
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.
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.

EIC Software: Statement of Principles



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.

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.

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.

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.

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.

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.

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.

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.

The "Statement of Principles" represent guiding principles for EIC Software. They have been endorsed by the international EIC community. For a list of endorses, see LINK.

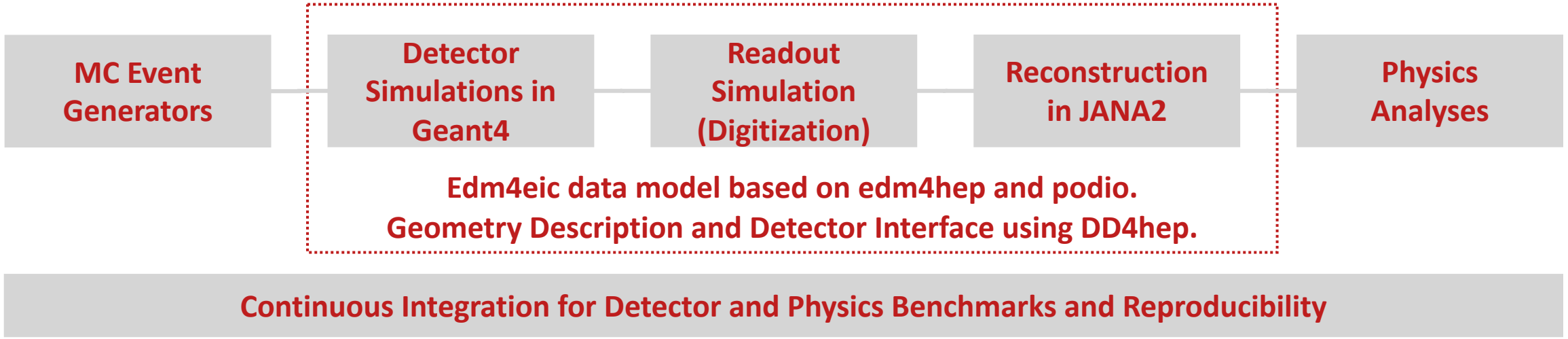
EIC Software is:

1. Diverse
2. Integrative
3. Heterogeneous
4. User-centered
5. Accessible
6. Reproducible
7. Collaborative
8. Agile

ePIC Software Stack

Our software design is based on **lessons learned in the worldwide NP and HEP community** and a **decision-making process** involving the whole community. We will continue to work with the worldwide NP and HEP community.

Modular Simulation, Reconstruction, and Analysis Toolkit using tools from the NP-HEP community



We are providing a production-ready software stack throughout the development:

- **Milestone**: Software enabled first large-scale simulation campaign for ePIC.

We have a good foundation to meet the near-term and long-term software needs for ePIC.

Phase Transition

ePIC Software & Computing

Sylvester's Picture of Salt & Pepper

EICUG SWG

- The formation of the first scientific collaboration for the EIC marks a phase transition.
- EIC Project and ePIC adhere to a strict timeline with well-defined deliverables spanning short to long term.
- **ePIC Software & Computing**:
  - **Drive the development of EIC Software**.
  - **Shape the computing model for ePIC**.

- EICUG plays a vital role in building and supporting the EIC community.
- **EICUG SWG**:
  - **Forum for Forward-Looking Projects**
  - **Cross-Collaboration Platform:** Encouraging collaboration across:
    - Experient and theory
    - ePIC and Detector II
    - Interdisciplinary fields (data scientists, HEP, ...)
  - **EICUG SWG will continue with a new charge.**

EICUG SWG Phase II: Charge

The mission of the EIC Software Working is to provide a forum for discussion and development of forward-looking software projects relevant to the entire EIC community. The working group will serve as a platform for cross-collaboration, between experiment and theory, the ePIC Collaboration and the Detector II/IP8 working group, as well as the nuclear physics, high- energy physics and data science communities. **The working group is charged with**:

1. **The continued development of the Monte Carlo Event Generators needed to realize the EIC science mission.** The working group will continue to coordinate with external theory and phenomenology, simulation and software groups, such as MCNet and the HEP Software Foundation, and will organize events to educate and assist the EIC community in utilizing these tools.

Topic of the upcoming discussion.

2. **Developing and maintaining connections to the data science community and engaging with the rapidly evolving AI/ML toolset that may impact the realization of the EIC science mission.** The working group will organize events to educate and assist the EIC community in utilizing AI/ML techniques in the areas of detector design and controls, simulations, data readout and analysis and theory and phenomenology.

Topic of the upcoming presentation by Cristiano Fanelli.

3. **Provide technical expertise and computing support for the development of the simulation, reconstruction and analysis software needed to support the Detector II working group.** Where possible, the working group should leverage existing frameworks and tools being developed within the ePIC collaboration.

EICUG SWG Conveners


Andrea Bressan


Cristiano Fanelli


Markus Diefenthaler


Tanja Horn


Torre Wenaus

Liaison to EICUG SC

Many thanks to EICUG SC for support.

From Andrea, Markus, and Torre: We will resign as conveners as soon their successor is known.
We would like to express our gratitude for your strong support over the past years. It has been an absolute pleasure working with you and striving for the success of the EICUG.