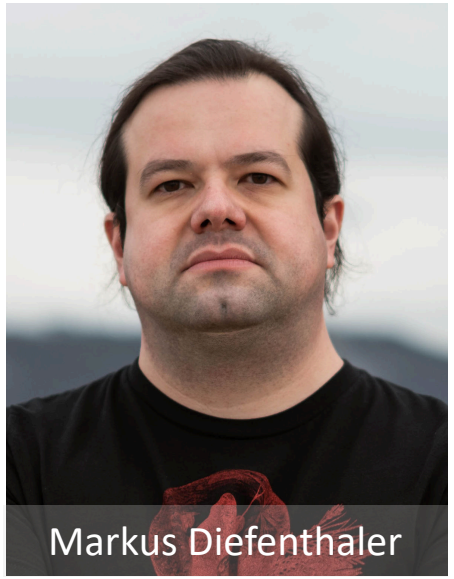
Update from EICUG Software Working Group

EICUG Software Working Group: Conveners


Andrea Bressan


Cristiano Fanelli


Markus Diefenthaler


Tanja Horn


Torre Wenaus

Liaison to EICUG SC

Many thanks to EICUG SC for support.

Mailing list (135 members): eicug-software@eicug.org

Meetings (22 meetings): https://indico.bnl.gov/category/301/

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**

Future

- Embrace software community in the EIC and beyond.
- Continue focus on common software projects.
- Explore and develop potential of AI / ML with **new** AIWG.
- Build on existing work with theory community and develop common projects.
- Support simulation efforts for second IR and detector.

Common discussion of the a) role of the EICUG SWG and b) interplay of EICUG-EPIC SWGs immediately after this and Sylvester's talk.

EIC Software Community

- Software in very early life stage.
- **Focus on common software tools**
  - Avoid duplication of efforts.
  - Team up on challenges, e.g., heterogeneous computing.
- Work with community standards, e.g., HepMC3 for event generation.

- **Engage with the wider NHEP community**
  - AI/ML
  - **Detector collaborations**
    - EPIC and 2nd detector efforts
  - Geant4 collaboration
  - HEP Software Foundation
  - **Theory community**
    - MC event generators (MCEGs)

Work with forming EPIC collaboration

Detector collaboration proposals

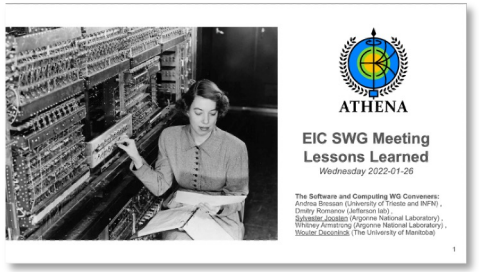
- Built upon work from Yellow Report Initiative.
- Very successful efforts on large-scale, detailed full detector simulations:
  - ATHENA have successfully developed a modular software stack based on common NHEP software.
  - ECCE have successfully leveraged familiar software.

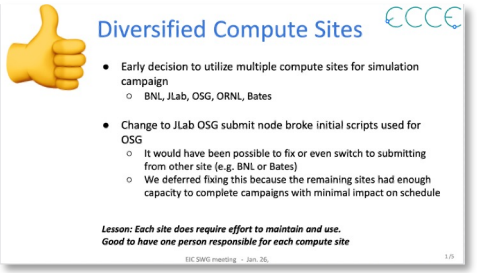
Lessons Learned

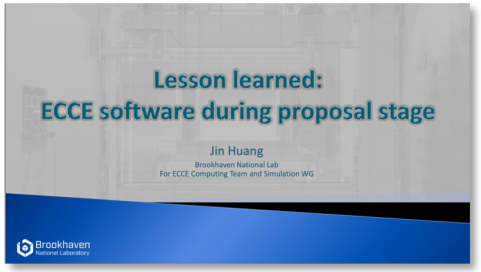
- **01/27** **Presentations from ATHENA and ECCE**
- **02/23** **Discussion with ATHENA and ECCE developers**, identifying commonality.
- **03/23** **Balanced summary by SWG** in the context of:
  - Common Software as laid out in Expression of Interest for Software.
  - Proceeding with work on one software stack.

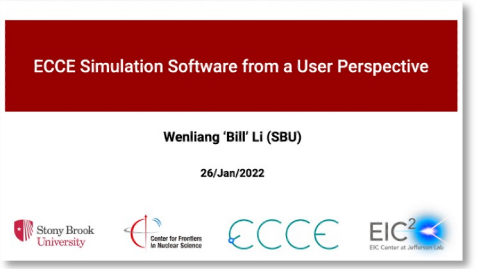
Joint meetings towards one EIC Software stack

- **May 11 – now** Software decision process
- EIC Software: Statement of Principles

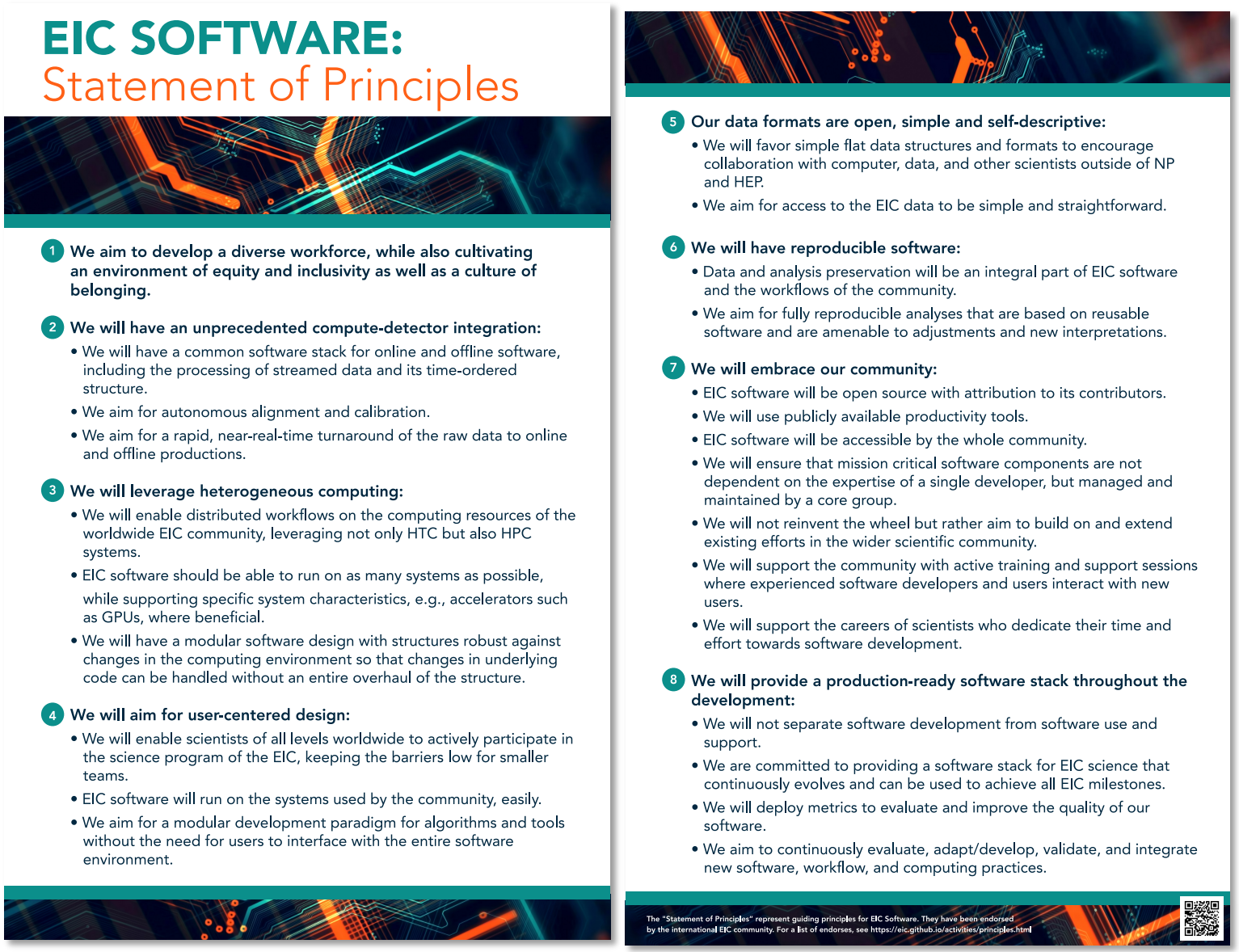








Guiding Discussion on EIC Software Stack



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 https://eic.github.io/activities/principles.html

- Used for **software decision process**.
- **Requirements** based on **Statement of Principles**.
- **PDF version** by Joanna Griffin (JLab).

Community Endorsement

- Endorse **here**.
- **List of Endorsers** (40).

Common Software Projects in the context of the Software EoI

Software Tools for Simulations and Reconstruction

- **Monte Carlo Event Generators**: Discussed on slide 13.
- **Detector Simulations**: Geant4 with DD4hep for geometry description and detector interface.
- **Reconstruction**: Modular reconstruction algorithm in the multi-threaded JANA2 framework.

Middleware and Preservation

- **Workflows**: Automated workflows on GitHub / GitLab.
- **Data and Analysis Preservation**: Task force for reproducible and reusable analyses.

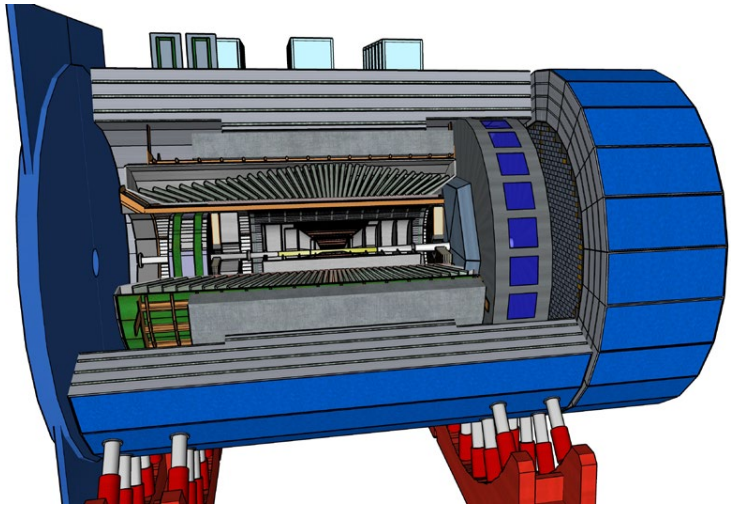
Interaction with the Software Tools

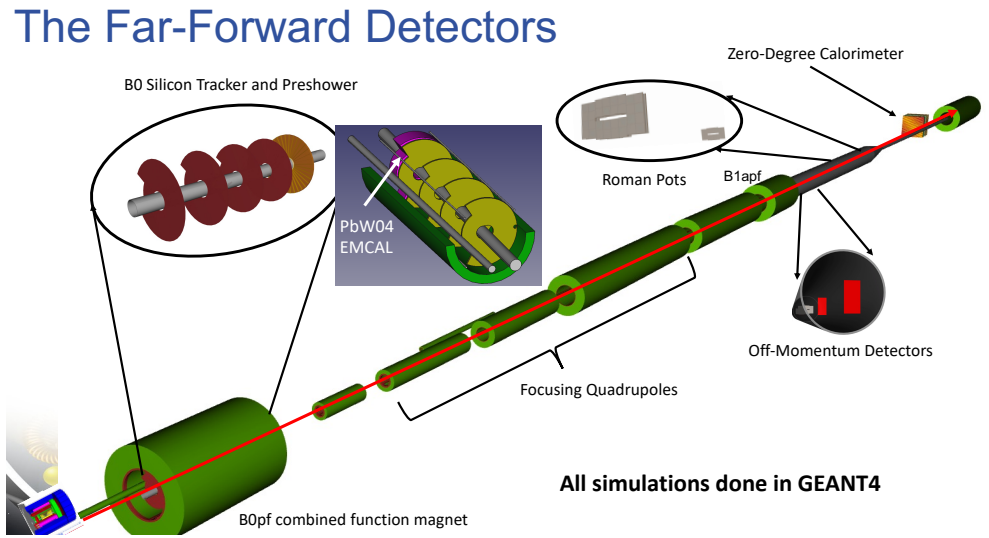
- **Explore User-Centered Design**: Discussed on slide 12.
- **Discoverable Software**: Software on GitHub, deployed in various ways (CVMFS, containers, spack).
- **Data Model**: Extend EDM4hep data model, use podio.

More Details in Sylvester's talk, directly after.

Priorities for Detector Design

- **Detector Simulations**
  - Validation of Geant4: Make test-beam setup and results available.
  - Detector design optimization using AI/ML.
  - Accelerate detector simulations:
    - Fast and accurate simulations using AI/ML, e.g., for simulation of calorimeters, Cherenkov detectors.
    - Fast simulations fully integrated into Geant4.
- **Reconstruction**
  - Accelerate reconstruction using AI/ML.
  - Reconstruction with far-forward detectors fully integrated.





Slide from Alex Jentsch

R&D Towards Next-Generation Detector Simulations

Detector Simulation

- **EIC focused project**
- **Turn-key application**
- **Built on top of Geant4 for full and fast simulations**
- **With library of potential detector option**

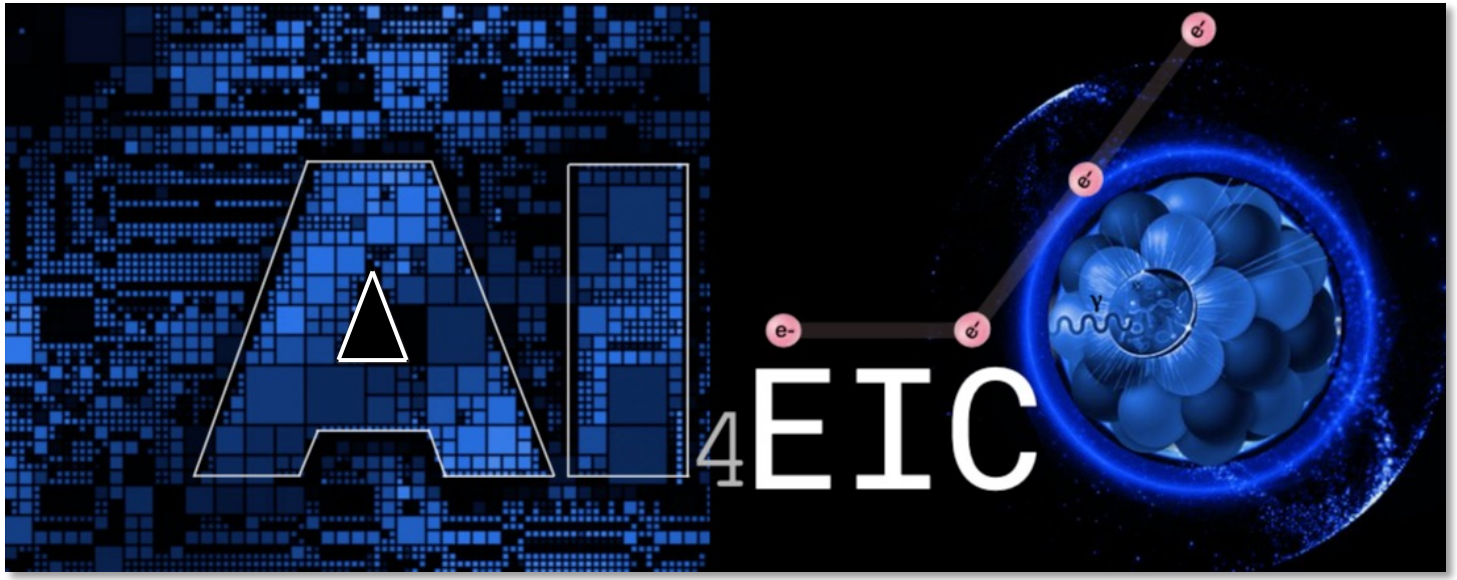
Requirements

- Ease of **leveraging new and rapidly evolving** technologies:
  - AI/ML to accelerate simulations
  - Heterogeneous architectures:
    - AI/ML is the best near term prospect for using LCF/Exascale effectively.
- Ease of switching detector options
- Ease of switching between detailed and coarse detector descriptions

Project

- Support for high concurrency heterogeneous architectures and fast simulations integrated with full detector simulations allows to leverage AI/ML in Geant4.
- Next phase in concurrent Geant4: Sub-event parallelism.

Explore and Develop Potential of AI / ML



AI Working Group (AIWG) as part of SWG

- **AI / ML research**
  - Scientific, systematic approach to applying AI / ML approaches to EIC problems.
  - Developing ML networks and techniques particular to EIC applications.
  - Uncertainty quantification.
- **Reference datasets for AI / ML development in EIC**
- **Data and analysis preservation for AI / ML approaches**

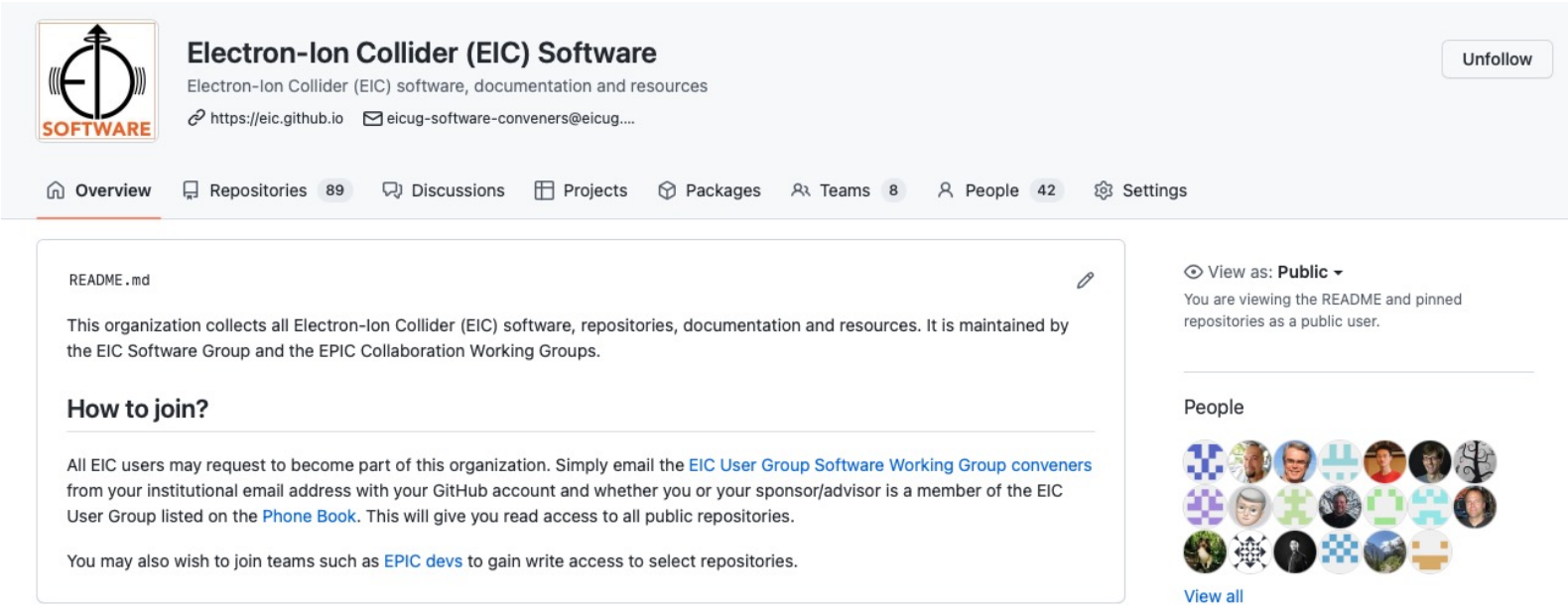
More details in Cristiano's talk, shortly after.

Building a Software Community

“Future Trends in Nuclear Physics Computing”: Discussion on common software and software community:

- 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.
- Developers need the time and space to develop something new, not something just a little better.

Community Resource https://github.com/eic, setup by SWG and maintained with collaboration(s)

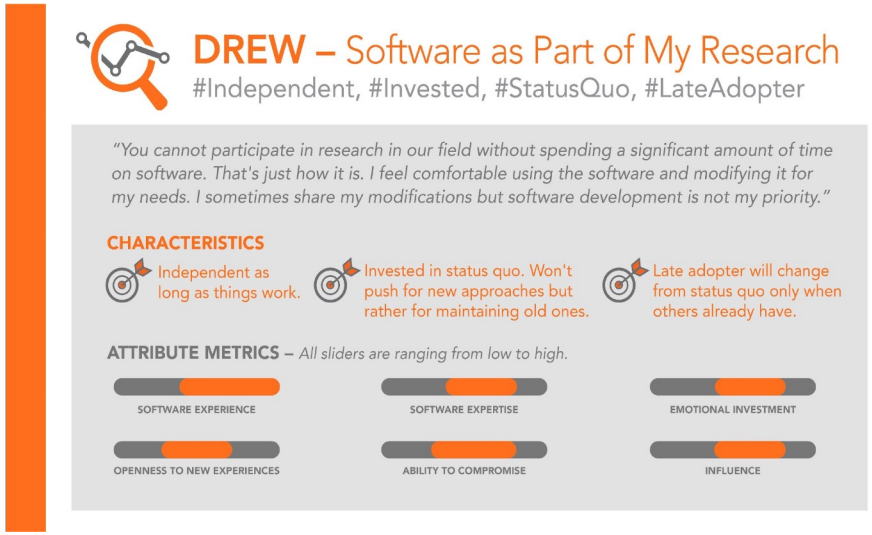


Working with Users: User-Centered Design in addition to training

- **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 five focus-group discussions**:
  - Extremely valuable feedback, documented many suggestions and ideas.
  - Developed user archetypes with Communication Office at Jefferson Lab and UX Design Consultant:

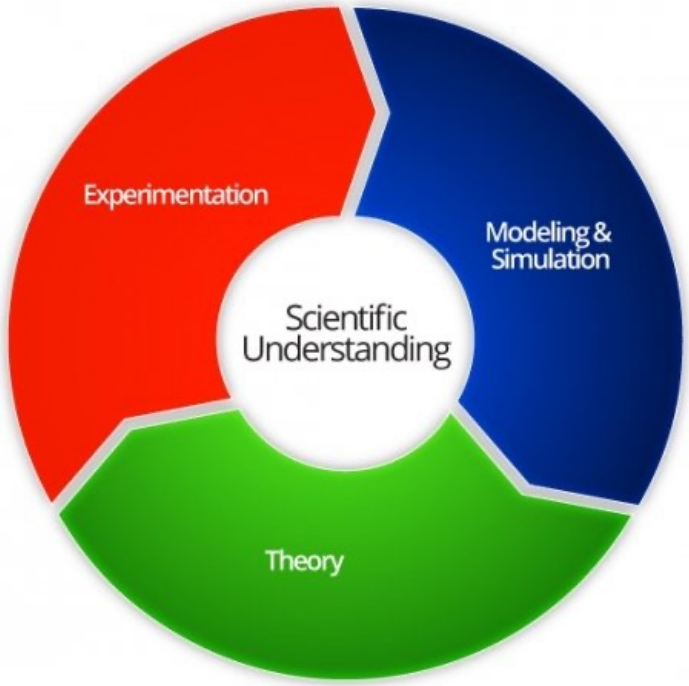
User Archetypes: Input to software developers as to which users they are writing software for:

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

Working with Theory Community on MCEGs



Indico, Summary (25 pages)

Common challenges in NHEP

High-precision QCD measurements require high-precision simulations.

- Community white paper on Event Generators for HEP Experiments, with EIC as part of cross-cutting aspects.

Monte Carlo Simulation for EIC:

- electron-proton (ep) collisions,
- electron-ion (eA) collisions, both light and heavy ions,
- including higher order QED and QCD effects,
- including a plethora of spin-dependent effects.

Unique challenges for EIC/NP

MCEGs for electron-**ion** collisions and **spin-dependent** measurements, including novel QCD phenomena (e.g., GPDs or TMDs).

Ongoing projects

- **MC4EIC** as forum for many open theory questions.
- Validation of general-purpose MCEGs and DIS tune.

Work with HEP Software Foundation

HSF general announcements
hsf-forum@googlegroups.com

Active involvement from EIC community in HSF (and vice versa)

Markus Diefenthaler appointed as

Convener of Physics Generators WG

Common forum for:

- **Discussion** on the physics event generators used by **N**HEP experiments
- **Technical work** on these physics event generators

Promotes collaboration among experiments and experiment and theory.

Jin Huang appointed as

Convener of Reconstruction and Software Triggers WG

- Considers approaches and solutions to common challenges in the area of event reconstruction and software triggering
- Algorithms and data structures that perform the heavy lifting data processing linking real/simulated data → analysis

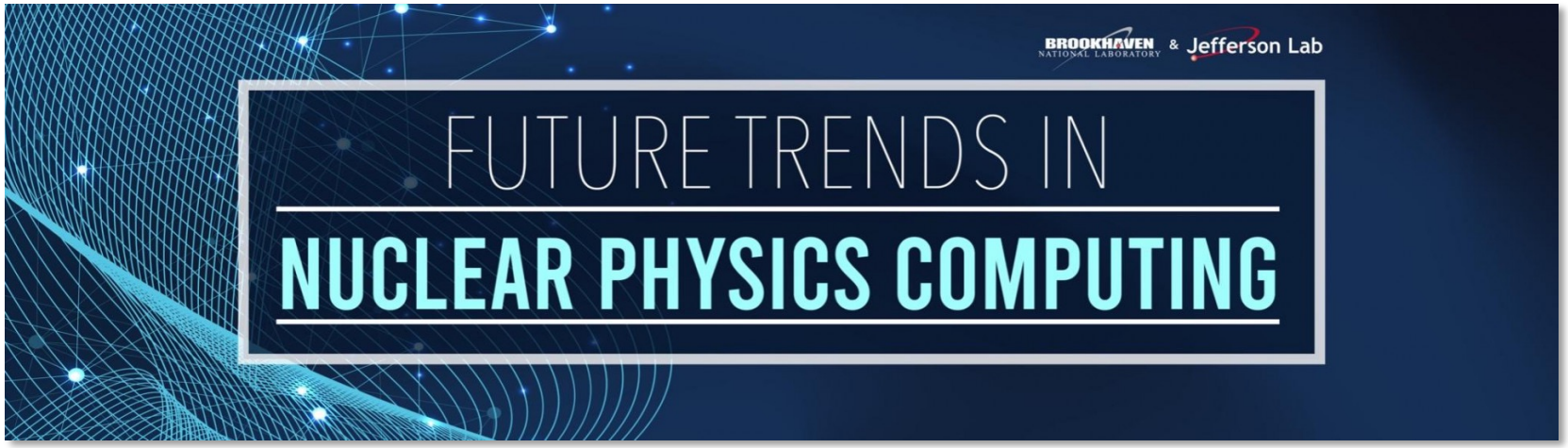
Wouter Deconinck appointed as

Convener of Training WG

- Provide training in the computing skills needed for researchers to produce high quality and sustainable software.
- Offer and expand a training program, ranging from basic core software skills needed by everyone to the advanced training required by specialists in software and computing.
- Organize periodic software carpentry workshops.

See also **EIC Software meeting on Feb. 16**

Future Trends in Nuclear Physics (2016 – now)



- **Fourth workshop**
  - September 28–30, hosted by CFNS.
  - Jointly organized by BNL, HSF, and JLab.
- **Goal** Discuss priorities for design and development as input for a community white paper to inform the next Long Rang Plan for Nuclear Science.
- **Key questions**
  - Where are we as a community?
  - How can we make analysis easier?
  - How can we scale up and down computing?

Future Directions

- Continue focus on common software projects:
  - Work with EPIC collaboration and wider community.
- Explore and develop potential of AI / ML with dedicated AIWG.
- Build on existing work with theory community towards common projects.
- Support simulation efforts for second IR and detector.