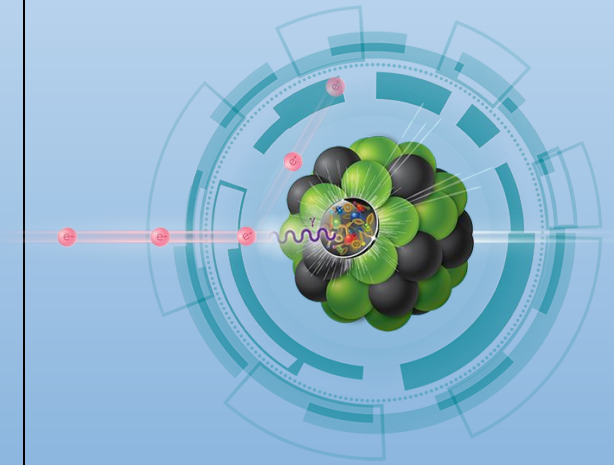
# Collaborative tools, EIC software EoIs and select EIC computing topics

Maxim Potekhin
*Nuclear and Particle Physics Software Group*

BROOKHAVEN NATIONAL LABORATORY



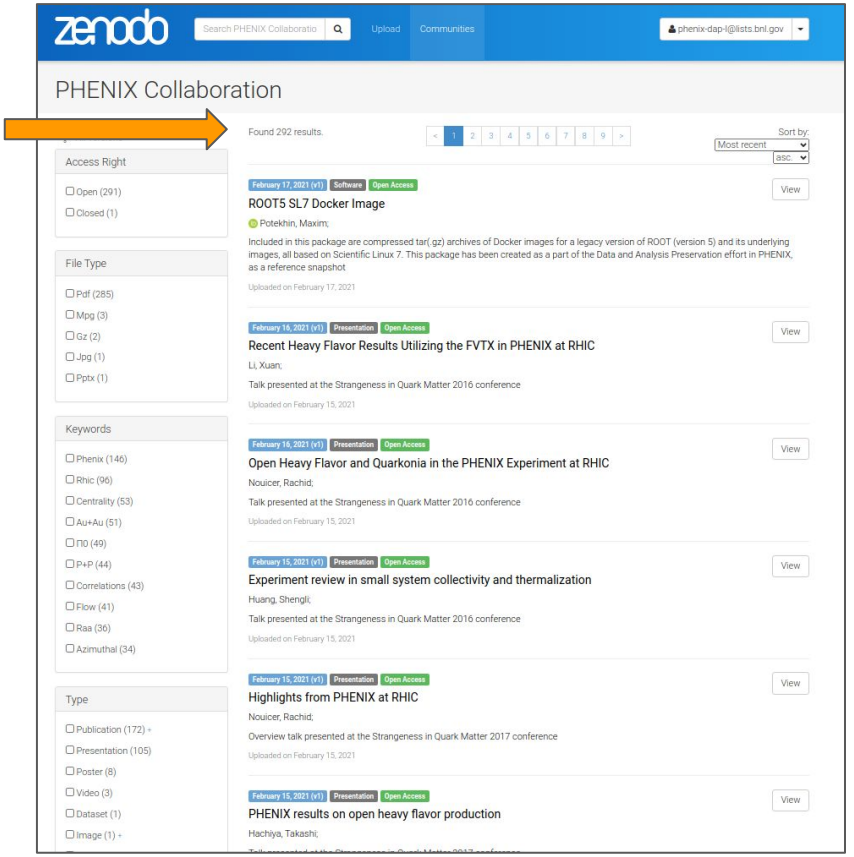
***BNL EIC Working Group Weekly Meeting***
02/18/2021

# Overview

- Zenodo
- Indico
- EIC S&C - the EoI process
- CI/CD
- REANA
- Storage
- The EIC Software Website

# Zenodo

- Example on the right - PHENIX uploads to the CERN instance, close to 300 items and growing
  - Theses, presentations, images, software, also in any combination thereof
  - Consistently tagged with a curated set of keywords
- The local Zenodo instance at BNL has been online for a while but doesn't see much traffic from the group - why is that?
  - Is the MFA a hurdle?
  - What's our document preservation strategy going forward?
- There will be a rollout of the *InvenioRDM* product which will replace the current version of Zenodo sometime around Summer'21, some migration may be necessary

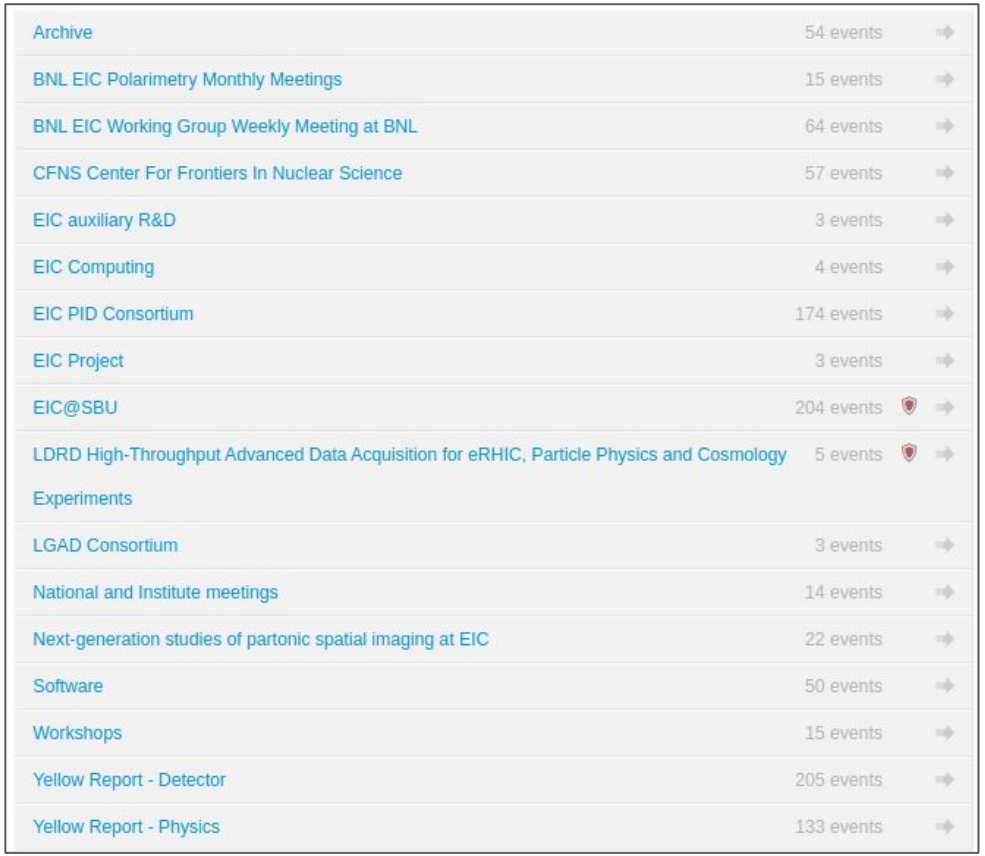


# Indico

- Over time, the head “EIC” category became overgrown and hard to navigate
- Creation of new categories was not coordinated across managers or groups
  - Can we change this going forward i.e. send announcements about proposed additions and changes?
  - Any proposals about the structure of our Indico area are welcome
- What do we do with the YR categories now that YR is close to final?
  - Rename, retire, archive?
- Dormant categories have been recently moved into the newly created “Archive”
- The EIC Software section relabeled and moved
- Moving categories does not break links (fortunately), so we can rearrange older and current materials to improve the quality of the site, without disruption

# EIC Indico

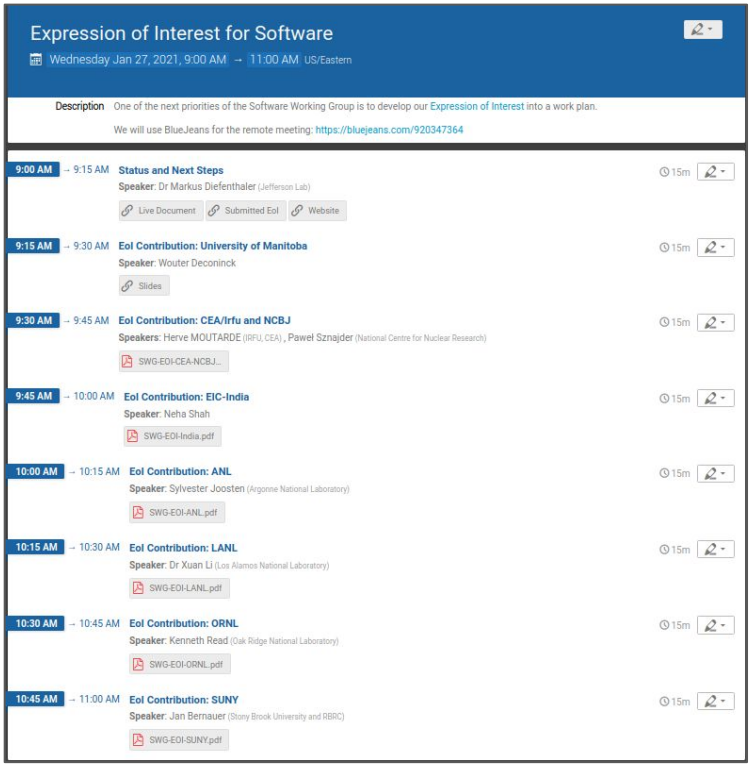
The current (updated) layout



# Software EoIs - kick-off on January 27th

https://indico.bnl.gov/event/10382/

- Please review if interested, a couple of points here -
- ANL: workflow-centric approach, CI/CD, GitLab (?)
- UManitoba: user-centric design, analysis preservation, workflows, code registry, tutorials, **REANA** (!) - more on this in subsequent slides
- ORNL: common software, containerization, Fun4All, MCGEN
- India: Fun4All, Escalate, QA, MC-data, EIC-smear
- ...more materials, please take a look



# Software EoI: the process

- This is work in progress
- Analysis preservation is a prominent item
  - Well aligned with software QA and validation
- Ties into CI/CD - experience with GitHub and GitLab tools in the community
- Current Software Group outlook presented at https://eic.github.io/activities/eoi.html
- NB. A large work item was not mentioned in the submitted EoI presentations:
  - Metadata
- Will keep you posted, this will be discussed in future meetings of the Software working group
- Last but not least, a new common simulation development is close to starting, leveraging previous efforts at BNL, JLab and other groups

# CI/CD as a validation tool

- Automation of testing, validation and deployment
  - flexible in how to define what each step is
- Many technologies can be used, including GitHub
  - offers CI/CD as a service (no local deployment needed)
- The concept of “GitHub actions” allows automation of *workflows*
  - It’s up to the user/developer to define what the workflow is
  - Operations are triggered by certain events, as configured by the operators
  - e.g. can automatically produce a set of reference histograms after a pull request
- Containers are the core of implementation
- Happening this week: a workshop/tutorial launched by HSF/IRIS:
  - https://indico.cern.ch/event/1001128/overview
- Very useful material, if time permits worth a look at the tutorial pages

# REANA

- https://reanahub.io/
- *“Reproducible research data analysis platform”*
- Substantial investment of effort by major experiments into adopting this platform
- Allows the user to capture both the software environment AND the workflow
  - Containers for the software
  - Formal YAML descriptions for the workflow
  - Result - reproducible workflows ready for analysis preservation and software validation
- We now have a test instance of REANA service at BNL
  - It “just works”
  - You are most welcome to try it!

# Storage

- We expect increasing needs for storage and various modes of access to it in the post-YR era
- BNLbox has been available for a while
  - Stable; CLI client available; can be thought of as Dropbox with extended storage, great for sharing docs and other files within a group
  - Not great for analysis/production (was not designed for this, weak client)
- Physical storage for EIC has been procured, and can be used in different ways
- Options for interfacing the scaled-up storage: 0(100TB)
  - XRootD
  - S3 (the BNL instance)

# XRootD

- Proven, scalable solution utilized by the LHC experiments
- Mature product with much effort invested in optimization
- Flexible auth/auth mechanisms
- Streaming option (i.e. remote files can be read transparently by ROOT)
  - Can be quite useful for final stages of analysis
- Organically supports storage federation (via redirectors)
  - Fully distributed data
- A rich set of client applications
- FUSE mount possible (i.e. a POSIX-like access method), cf. the EOS facility at CERN which is central to many experiments

# S3

- Familiar to many from its use in the Amazon cloud
- Object store as opposed to file store
- API supported by many modern platforms
- Experience in ATLAS
- S3 Storage Element supported by Rucio
- An open source product currently deployed at BNL on a test basis
- Initial testing underway with a simple CLI client - Kolja, Maxim
  - Some caveats/quirks discovered e.g. X509 cert setup details, wildcard interpretation etc
- Will evaluate its usefulness in real-life scenarios relevant for the EIC community
- Need to develop use cases

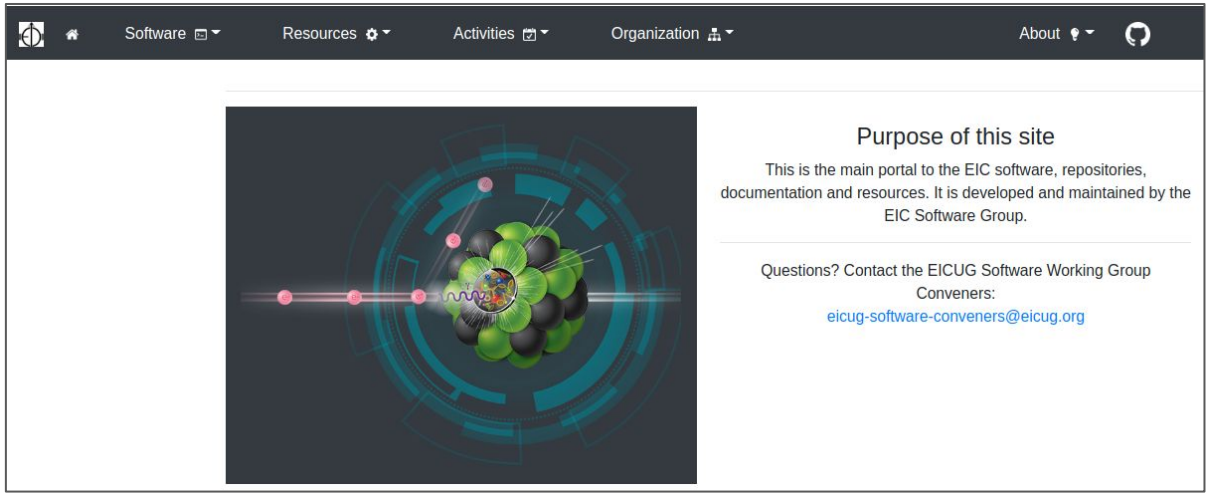
# S3: the “mc” client, basic testing

```
maxim@ferocity:~$ mc ls eic
mc: <ERROR> Unable to list folder. Get "https://dtn01.sdcc.bnl.gov:9000/": x509: certificate si
maxim@ferocity:~$ mc --insecure ls eic
[2020-11-04 22:37:23 EST]     0B eictest/
maxim@ferocity:~$ mc --insecure ls eic/eictest
[2021-02-17 15:09:23 EST]     0B TestFromWindows/
[2021-02-17 15:09:23 EST]     0B bla/
maxim@ferocity:~$ mc --insecure ls eic/eictest/bla
[2020-12-08 16:00:12 EST] 2.2GiB st_physics_adc_21021047_raw_6000002.daq
[2020-12-09 13:41:14 EST] 2.5GiB st_physics_adc_21023014_raw_6500002.daq
maxim@ferocity:~$ mc --insecure ls eic/eictest/bla
[2020-12-08 16:00:12 EST] 2.2GiB st_physics_adc_21021047_raw_6000002.daq
[2020-12-09 13:41:14 EST] 2.5GiB st_physics_adc_21023014_raw_6500002.daq
maxim@ferocity:~$ mc --insecure ls eic/eictest/
[2021-02-17 16:52:28 EST]     0B TestFromWindows/
[2021-02-17 16:52:28 EST]     0B bla/
maxim@ferocity:~$ mc --insecure ls eic/eictest/TestFromWindows
[2020-12-10 14:43:27 EST] 226KiB CAD_operation_current_001.jpg
[2020-12-10 11:13:36 EST] 3.2MiB s3browser-9-2-1.exe
maxim@ferocity:~$ mc --insecure cp eic/eictest/TestFromWindows .
mc: <ERROR> To copy a folder requires --recursive flag. Invalid arguments provided, please refe
maxim@ferocity:~$ mc --insecure cp eic/eictest/TestFromWindows/CAD_operation_current_001.jpg .
...CAD_operation_current_001.jpg:  225.92 KiB / 225.92 KiB |
maxim@ferocity:~$ ls -ltr
```

# Storage: a quick summary

- Opinions? SDCC currently leaning to S3 for reasons of available support but open to community feedback and requirements
- Need volunteers for evaluation
  - This can be useful for real work
  - Anyone can acquire an account
- Will report and revisit this topic at a later date i.e. experience with the S3 client
- Any interest in replicating data between BNL and JLab?
- At this time, not much progress with design of the Metadata system for EIC which will be necessary regardless of the future storage technology choices

# The EIC Software Website: https://eic.github.io/




- Highly functional framework, simple to maintain and contribute to
- Structure and navigation updated
- We need more content to better document software activities in the EIC community
- Please consider contributions - help needed!