eAST Documentation

Maxim Potekhin
Nuclear and Particle Physics Software Group

Brookhaven National Laboratory

Project eAST Meeting
09/28/2021

Motivation

- eAST is a complex product
- Short term - need to ensure uptake and wide use by the EIC community
- Long term - stewardship and maintenance
- Solid web-based documentation is crucial on any timescale
- “READMEs” in repositories are very useful but may be cumbersome to navigate

Solutions (1): Content Management Systems

- Content Management Systems: Wiki(s), Drupal, Wordpress etc
- Easy to create and populate
- Sites tend to overgrow with obsolete or irrelevant material
- Hard to get a bird-eye view of the content
- Data content is mixed with rendering/presentation
- DB and PHP-based, hence long-term maintenance and updates are required - some recent experience in EIC and STAR confirm that

Solutions (2): Static Website Generators

- Static website generators - Jekyll, Hugo etc
  - Full HTML content of a site is *generated automatically* based on Data+Templates
  - Speed and security (content is static), low long-term maintenance
  - Highly customizable **processing** and rendering of structured data is possible, with DB-like capabilities at build time
  - Macros
  - GitHub integration is organic i.e. included with GitHub
- Solid templating capability and a template language similar to Django, Jinja etc
  - Iterators, conditional statements, filters, sort etc
- Data and text content are decoupled from rendering (e.g. new pages can be created using same underlying data or the whole site can be re-skinned).
- Sphinx - another good option to consider

Website Generators: Recent Experience in NP/HEP

- HEP Software Foundation
- NPPS
- PHENIX
- ECCE
- EIC Software Group
- Experience is overwhelmingly positive

Markdown and GitHub pages

- If we want to host on GitHub, we must follow the GitHub requirement that the site materials are located in the folder named “docs” in the repository
- GitHub runs Jekyll and rebuilds the site automatically at every commit/push
- There are small differences between the vanilla MD and the dialect on GitHub
- On top of the page there is the “front matter” section which is a snippet of YAML enclosed in ‘---’ lines, which can be customized as needed to keep page data if necessary
- Importantly, the template for the page is also specified in the front matter section, and so is the name by which it can be found and referenced on the site
- See example on the next slide

Ordinary page example (the proposal)

```
---
title: "eAST proposal"
layout: base
name: proposal
---

# Motivation

Large-scale detector systems for the EIC are designed by larger
communities. The simulation efforts typically start as standalone
exercises for each detector component with various levels of maturity,
analytical calculations, simplified Monte Carlo exercises (fast
simulations), Geant4-based Monte Carlo approaches (full simulations),
and are later being extended in several different frameworks.

It is critical in the longer term to perform studies where the
information from various detector elements and also the interaction
region, support structures and other dead material is taken into
account, which is only possible in a comprehensive simulation. This is
essential to understand, e.g., the performance at the edges of the
detector system or the effect of combining different technologies for
the electromagnetic calorimeter.
```

A more complex Front Matter example

```
---
title: "Installation"
layout: base
name: installation

# What's below is optional, in this case it contains contents of a table

tables:
  g4options:
    headers: [Option, Comment]
    rows:
      - ["GEANT4_BUILD_MULTITHREADED", "-" ]
      - ["GEANT4_USE_GDML", "Requires <em>Xerces-C++</em> pre-installed" ]
      - ["GEANT4_USE_OPENGL_X11", "Needed if you use the event display. Requires <em>X11</em> and <em>OpenGL</em> libraries"]
      - ["GEANT4_USE_QT","Needed if you use the Qt window. Requires <em>Qt5</em> and <em>OpenGL</em> libraries"]
---
```

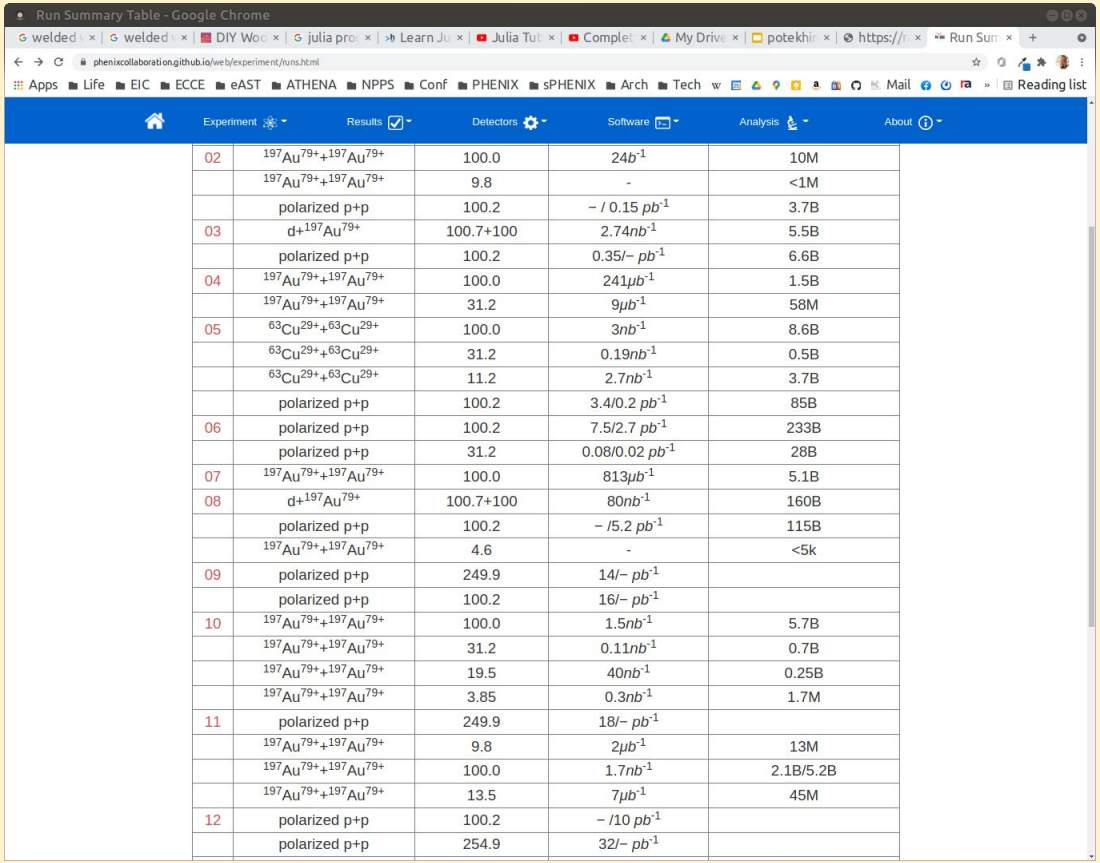
Data content

- Data are stored as YAML files which can be simple or as complex as the developer needs them to be: lists, dictionaries and any combinations thereof
- CSV is also supported transparently (so if needed import from Excel is trivial)
- Use can be trivial e.g. a list of publications with DOI references or a phone book, or quite technical/quantitative i.e. metrics of binaries run on different platforms, effects of optimization, **physics validation** results etc
- Can be ingested by other applications in other languages e.g. Python, C++ etc
- Geant 4 can read these files if necessary

Example of data records on a Jekyll-based site

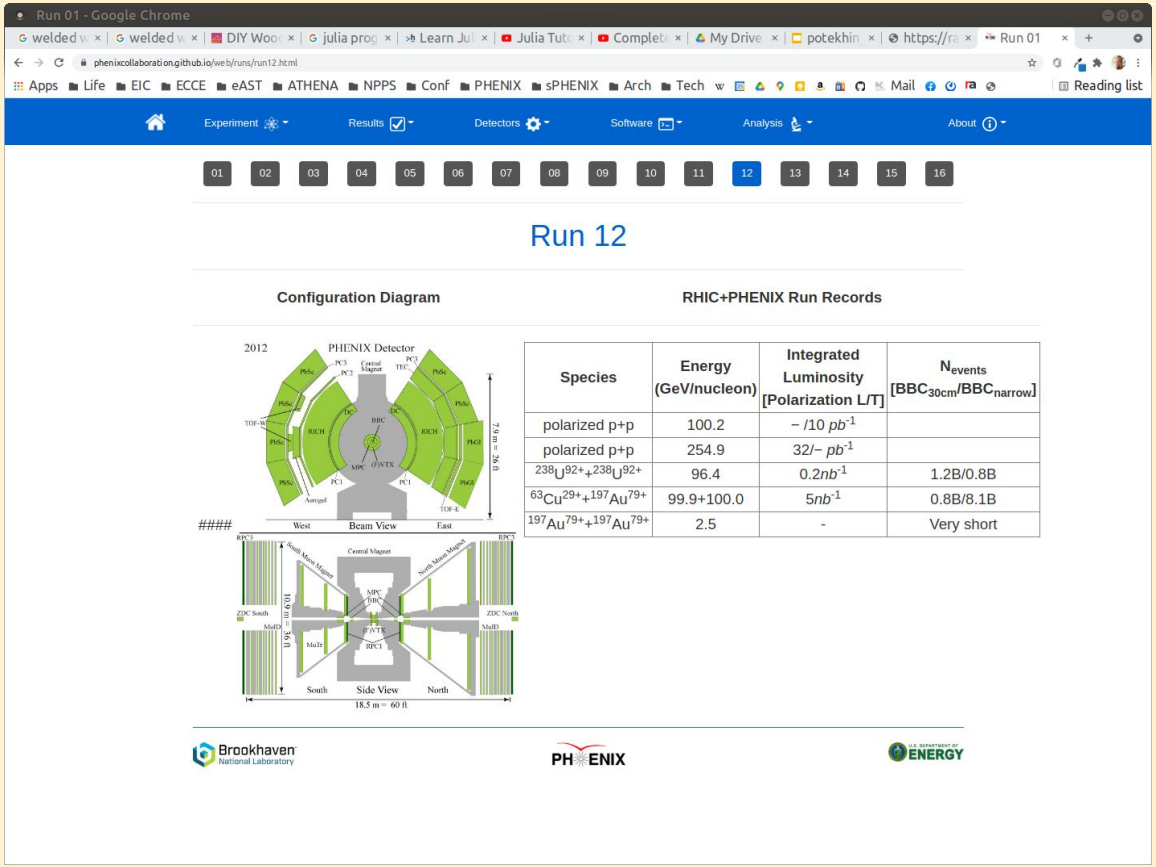
```
- run: run13
  title: Run 13
  period: 2013
  coordinator: Hubert van Hecke, LANL.
  rhic:
    - {
       species: 'polarized p+p',
       energy: 254.9,
       lumi: '155/− <i>pb</i><sup>-1</sup>',
       Nevents:
      }
  ert_comment: Summary of thresholds (DAC values). Values in parentheses are for the PbGl.
  ert_thresholds:
    - '03/10/13, 368358, 30(29), 31(30), 29(29), 34(25), 920, Run13pp510 - EMCal dynamic range ~50GeV'
  ert_notes: |
    ~50GeV dynamic range (&#8730;s=510GeV)<br/>
    this is basically taken from run12, and need to be updated<br/>
    2x2 DAC-values correspond approximately to [energy/tile]:    39=2.2 GeV (The analysis was done with 39(37))<br/>
    4x4 DAC-values correspond approximately to [energy/tile]:    29=3.7 GeV, 30=4.7 GeV , and 31=5.6 GeV<br/>
    RICH4x5 ADC-values correspond approximately to: 920=3p.e.<br/>

    These values are estimated from Run 9 ADC sum analysis. The real thresholds have to be determined offline.
    They depend on the gains/HV settings of the EMCal modules. The EMCal values in parentheses correspond to the
    threshold of the PbGl.
  ert_masks:
    - '03/10/13 17:30, 386830, ERT_MASK:addmask.txt_20130310, yes, Used ERT online monitoring. CHC add in 4x4C E3 SM10.'
```

Example of a rendered page



02	$^{197}Au^{79+}+^{197}Au^{79+}$	100.0	$24 b^{-1}$	10M
	$^{197}Au^{79+}+^{197}Au^{79+}$	9.8	-	<1M
	polarized p+p	100.2	$- / 0.15\ pb^{-1}$	3.7B
03	$d+^{197}Au^{79+}$	100.7+100	$2.74 nb^{-1}$	5.5B
	polarized p+p	100.2	$0.35/- pb^{-1}$	6.6B
04	$^{197}Au^{79+}+^{197}Au^{79+}$	100.0	$241 \mu b^{-1}$	1.5B
	$^{197}Au^{79+}+^{197}Au^{79+}$	31.2	$9 \mu b^{-1}$	58M
05	$^{63}Cu^{29+}+^{63}Cu^{29+}$	100.0	$3 nb^{-1}$	8.6B
	$^{63}Cu^{29+}+^{63}Cu^{29+}$	31.2	$0.19 nb^{-1}$	0.5B
	$^{63}Cu^{29+}+^{63}Cu^{29+}$	11.2	$2.7 nb^{-1}$	3.7B
	polarized p+p	100.2	$3.4/0.2\ pb^{-1}$	85B
06	polarized p+p	100.2	$7.5/2.7\ pb^{-1}$	233B
	polarized p+p	31.2	$0.08/0.02\ pb^{-1}$	28B
07	$^{197}Au^{79+}+^{197}Au^{79+}$	100.0	$813 \mu b^{-1}$	5.1B
08	$d+^{197}Au^{79+}$	100.7+100	$80 nb^{-1}$	160B
	polarized p+p	100.2	$- /5.2\ pb^{-1}$	115B
	$^{197}Au^{79+}+^{197}Au^{79+}$	4.6	-	<5k
09	polarized p+p	249.9	$14/- pb^{-1}$	
	polarized p+p	100.2	$16/- pb^{-1}$	
10	$^{197}Au^{79+}+^{197}Au^{79+}$	100.0	$1.5 nb^{-1}$	5.7B
	$^{197}Au^{79+}+^{197}Au^{79+}$	31.2	$0.11 nb^{-1}$	0.7B
	$^{197}Au^{79+}+^{197}Au^{79+}$	19.5	$40 nb^{-1}$	0.25B
	$^{197}Au^{79+}+^{197}Au^{79+}$	3.85	$0.3 nb^{-1}$	1.7M
11	polarized p+p	249.9	$18/- pb^{-1}$	
	$^{197}Au^{79+}+^{197}Au^{79+}$	9.8	$2 \mu b^{-1}$	13M
	$^{197}Au^{79+}+^{197}Au^{79+}$	100.0	$1.7 nb^{-1}$	2.1B/5.2B
	$^{197}Au^{79+}+^{197}Au^{79+}$	13.5	$7 \mu b^{-1}$	45M
12	polarized p+p	100.2	$- /10\ pb^{-1}$	
	polarized p+p	254.9	$32/- pb^{-1}$	

Same data rendered on a different page

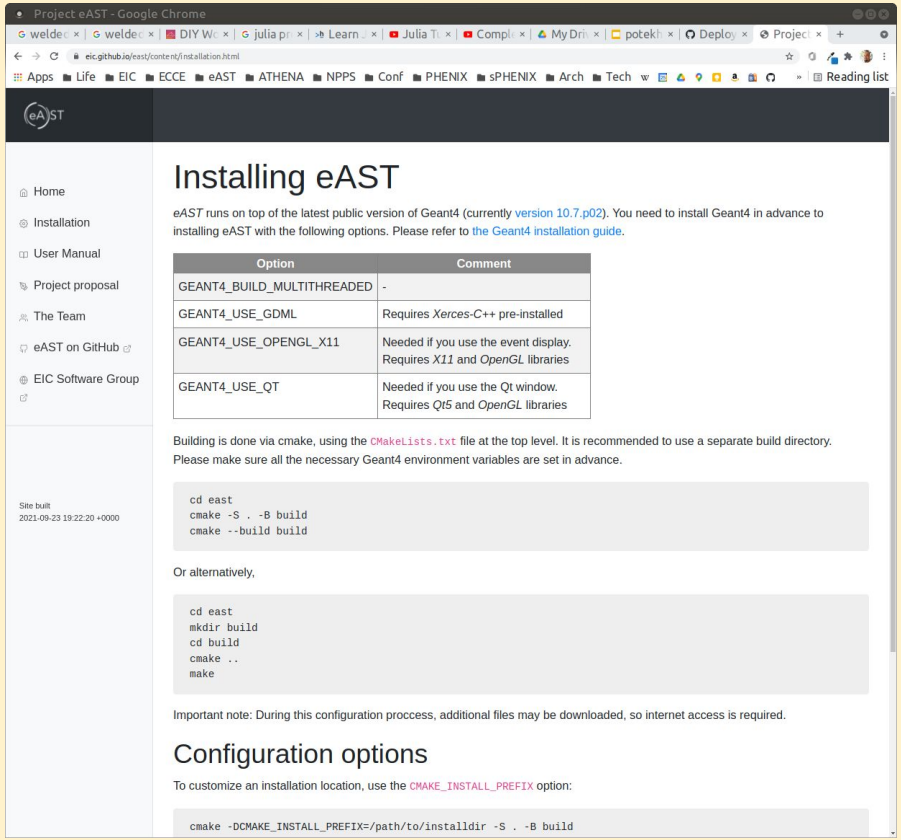


Species	Energy (GeV/nucleon)	Integrated Luminosity [Polarization L/T]	N_{events} [BBC_{30cm}/BBC_{narrow}]
polarized p+p	100.2	– /10 pb^{-1}	
polarized p+p	254.9	32/– pb^{-1}	
$^{238}U^{92+}+^{238}U^{92+}$	96.4	0.2nb^{-1}	1.2B/0.8B
$^{63}Cu^{29+}+^{197}Au^{79+}$	99.9+100.0	5nb^{-1}	0.8B/8.1B
$^{197}Au^{79+}+^{197}Au^{79+}$	2.5	-	Very short

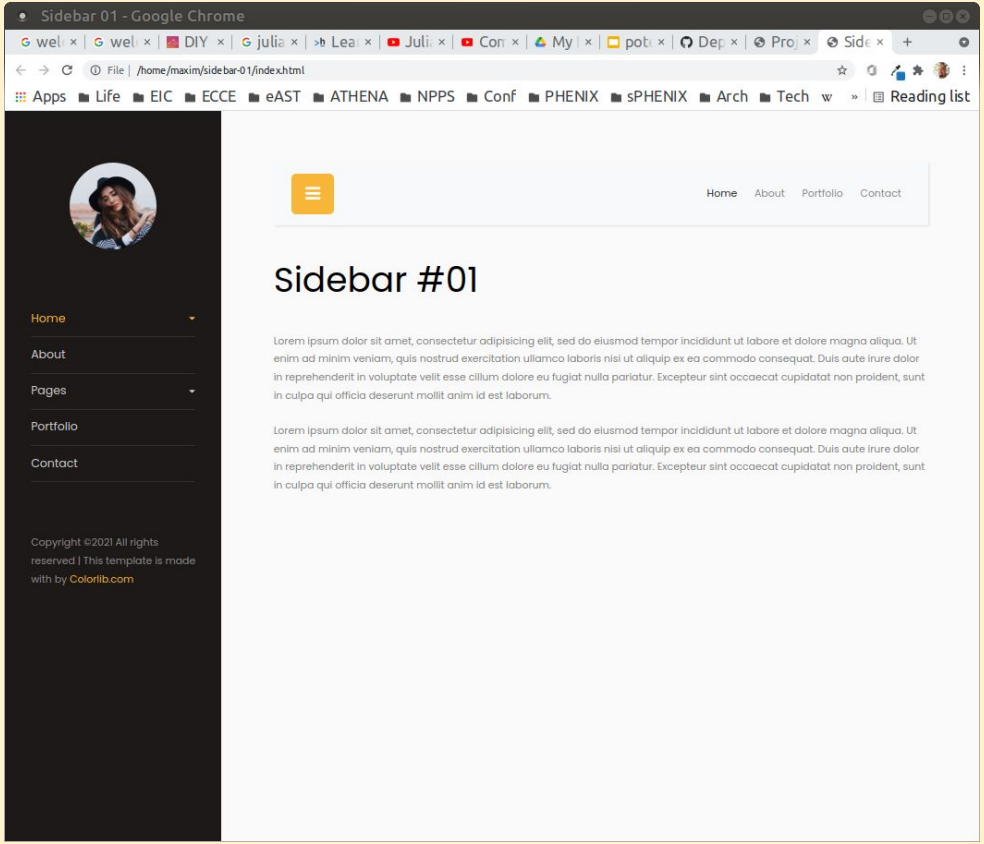
Rendering

- Pretty much any sort of template and styles can be created or acquired
- The Bootstrap-based pages are quite popular these days, and used on the EIC Software, NPPS and PHENIX sites in particular
- A site can be completely redesigned while the content is entirely intact
- It’s not difficult to add CSS in order to customized certain aspects of rendering (like inline code blocks, look and feel of tables etc)

The eAST doc site prototype: https://eic.github.io/east/



Alternative skin example (mobile-friendly)



Development, hosting and maintenance

- Markdown is a very basic text format and a comfortable medium to work with
- Intermixable with HTML if and when necessary
- GitHub recently added native (web-based) VS code capability meaning developers can use a very powerful editor in their browser and the changes are committed directly to the repo with instant effect, making it quite similar to content management systems
- Alternatively, the usual git workflow can be used, complete with branches, pull requests etc - collaboration is easy
- GitHub pages work well, but it's also easy to host content on any web server if necessary - HTML is trivial to export and plant on a web server

Beyond the Web

- As useful as the static site approach is, it's not ideal for generic document storage
- Aside from technical manuals the project is likely to produce white papers, plots, data tables etc which sometimes are best preserved as standalone items (pdf etc)
- Project eAST would benefit from durable and searchable storage
- Systems like DocDB are obsolete and hard to maintain in the long term
- Wikis, Indico, Dropbox etc don't quite cut it
- A proper digital repository is preferred

Zenodo/InvenioRDM

- Invenio “Research Data Management” is a CERN project
- Zenodo is a digital repository running on top of Invenio
- PHENIX has positive experience with the CERN instance of Zenodo which is based on the InvenioRDM backend
  - Durable links permanent (minted DOIs),
  - Powerful search function including elastic search (full text)
  - No restrictions on types of files to be stored, full access control
- However EIC will rely on different Invenio-based applications hosted at BNL
  - This is now under discussion by the management, stay tuned
  - This will take time

A possible interim solution

- Storing various plots, PDFs and white papers right in the “east” repository itself is not optimal as we need to keep it clean and trim
- A possible solution - create a sister repository “east-documents” where such materials would go
- An advantage - can migrate to an Invenio-based or other platform later once it becomes available

Summary

- eAST needs a solid and easy to maintain documentation platform
- Static website generation with native GitHub hosting is a popular solution, already used in EIC Software, PHENIX, ECCE and other projects
- A prototype of the eAST documentation website has been created
- If chosen as the solution, its usefulness will depend on engagement with, and contributions from developers. We can start today.
- Pages are not more difficult to edit than a README but the site has many other capabilities in addition to rendering Markdown