

A review of the legacy external PHENIX website and considerations for a possible migration/upgrade

Maxim Potekhin
(BNL, NPPS)
08/20/2020

Overview

- The lasting Web presence of PHENIX is an important part of its DAP
 - Perhaps the most important - DAP needs to be discoverable
- The topic of today's discussion is the [external](#) website (internal left for later)
 - It's a hub for a variety of links and a data source in and by itself
- The current technology platform consists of
 - PostgreSQL
 - PHP
 - File system
- It has grown “organically” over many years
- In the original situation (before the outage) one server was used for “everything”, from managing production to notes to theses
 - Also co-hosting the external and internal sites

Why are we here?

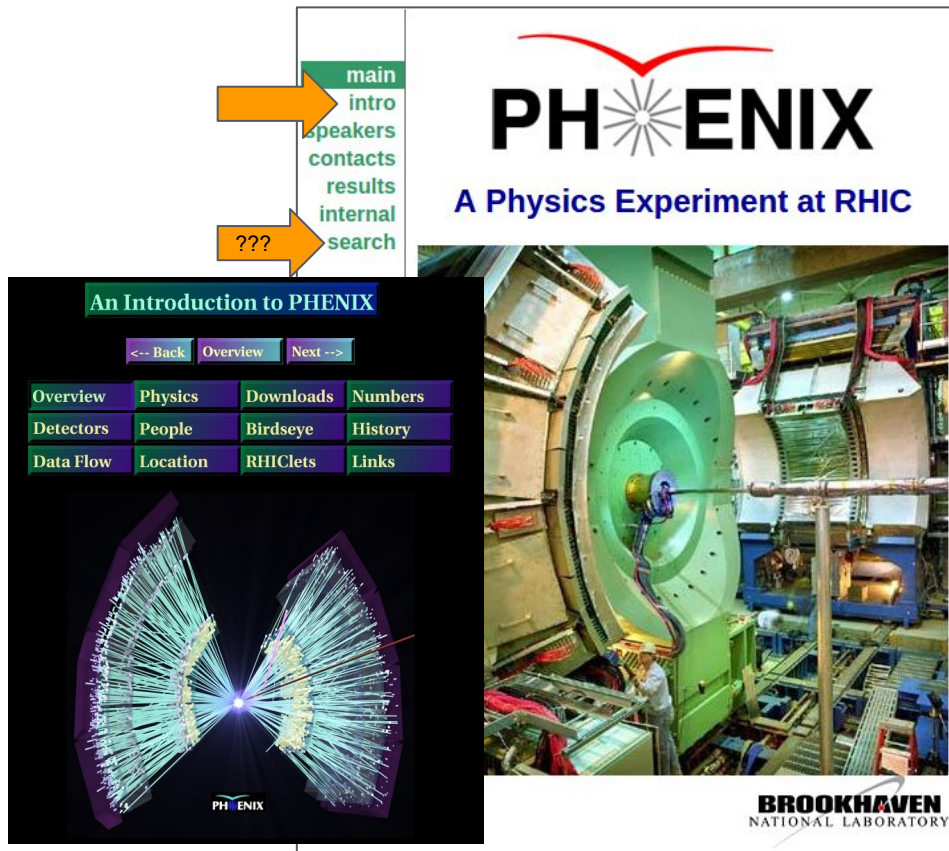
- Motivations for this discussion
 - Recent problems with security and robustness of the site - it is hard to properly fix
 - Anticipation of *operational risk* going forward in view of potential PHP version upgrades and evolution of the available OS installations at SDCC
 - Separation of external and internal parts of the service would be helpful
 - Obsolescence of certain links and material
 - A degree of overlap and duplication with the new DAP website
 - Perhaps we need to improve the look and feel
- Let's consider possible ways to mitigate these issues
 - ...and to estimate costs and benefits associated with this potential mitigation
 - Realistically the options are (a) fixing the old site (b) migration to DAP
 - In (a) it would still be hard to achieve proper look and feel

Requirements

- The central question in how to manage the website are the requirements
- What purposes does the website serve?
- It's the face of the Collaboration and its main PR instrument
 - Needs to look modern and sleek
 - Must have useful information both for collaborators and visitors
- What purposes will it need to serve in 3 years time?
 - General information about the experiment and the Collaboration, contact information
 - Pointers to research materials, publications, and **DAP**
- What's the acceptable level of upkeep and maintenance in the long term?
 - Close to zero would be a good answer
- Let's take a look at the current PHENIX page from the functional perspective

The PHENIX landing page - a critical look

- The “Intro” page - generic intro to PHENIX and its physics
 - A requirement
- But it looks decidedly dated and needs to be updated or migrated under any scenario
 - 80% links irrelevant
 - The event display gallery is easy to preserve if we decide to migrate
 - Detector information is already better laid out on the DAP site
- The “Search” link is a dud and can be ignored
- “Internal” is a subject of a separate discussion
- So, what’s left?



Relevant items on the external page

- Speakers Bureau
 - Very portable
 - Appearance needs to be improved
- Contacts
 - What functionality do we need?
 - Do we need to list all 1700 people who participated in PHENIX over time? (easy to do)
 - Quite portable (see next slides)
- Results
 - A requirement
 - Plots DB
 - Fairly complex metadata, needs thinking (see next slides)
 - List of publications
 - Seems portable, also Inspire, Zenodo and Google are your friends
 - cf. that all theses will eventually end up on Zenodo

A quick demo project - port contacts to the DAP site

- PHENIX is using multiple database tables to maintain the contacts, institutions and MGS information etc
 - Keeping track of relations is not too easy but doable
- Facilitating factors
 - Exporting data from PostgreSQL to CSV format (comma-separated values) is a one-liner: `copy people to 'people.csv' delimiter ',' header csv`
 - Jekyll (the platform on which the DAP site is built) reads CSV `natively`
 - Data relationships can be built in Jekyll by adding some code in its internal language
- A basic demo page has been created (by Maxim) and it can be improved/made much prettier if required
- In this example, a “JOIN” was implemented using the data from three tables
 - **people** (for first and last name)
 - **mgs20** (to filter active collaborators)
 - **emailaddr** (for e-mail addresses)

The phonebook/MGS20 demo on the DAP site

https://raw.githubusercontent.com/PhenixCollaboration/web/master/_experiment/collaboration.md

- Emulate the database “JOIN” operation with iterators and comparators

```
{% assign email=" %}  
{% for item in site.data.db.phenix_collab.emailladdr %}  
{% if item.person_id==person.id %}  
{% assign email=item.email %}  
{% break %}  
{% endif %}  
{% endfor %}
```



Reference to a file exported from Postgres

```
{% assign inst_name=" %}  
{% for item in site.data.db.phenix_collab.institutional_affiliation %}  
{% if item.person==person.id %}  
{% for institution in site.data.db.phenix_collab.institutions %}  
{% if institution.id==item.institute %}  
{% assign inst_name=institution.name %}  
{% break %}  
{% endif %}  
{% endfor %}  
{% endfor %}
```



Reference to a file exported from Postgres

The result: demo screen on the DAP site

<https://phenixcollaboration.github.io/web/experiment/collaboration.html>

Experiment	Detectors	Software	Analysis	Resources	About
Collaboration					
MGS20					
Family Name	First Name	e-mail	Institution		
Aidala	Christine	caidala@bnl.gov	Department of Physics, University of Michigan, Ann Arbor, Michigan 48109-1040, USA		
Akiba	Yasuyuki	akiba@bnl.gov	RIKEN BNL Research Center (RBRC), Brookhaven National Laboratory (BNL), Upton, New York 11973-5000, USA		
Bazilevsky	Alexander	shura@bnl.gov	Physics Department, Brookhaven National Laboratory (BNL), Upton, New York 11973-5000, USA		
Belmont	Ronald	belmonrj@gmail.com	University of Colorado, Boulder, Colorado 80309, USA		
Berdnikov	Yaroslav	berdnikov@spbstu.ru	Saint Petersburg State Polytechnic University, St. Petersburg, 195251 Russia		
Berdnikov	Alexander	alexber@phmf.spbstu.ru	Saint Petersburg State Polytechnic University, St. Petersburg, 195251 Russia		
Blankenship	Brandon	brandon.t.blankenship@vanderbilt.edu	Vanderbilt University, Nashville, Tennessee 37235, USA		
Chiu	Mickey	chiu@bnl.gov	Physics Department, Brookhaven National Laboratory (BNL), Upton, New York 11973-5000, USA		
Connors	Megan	meganEconnors@gmail.com	Georgia State University, Atlanta, Georgia 30303, USA		
Csanad	Mate	csanad@bnl.gov	ELTE, Eötvös Loránd University, H - 1117 Budapest, Pázmány P. s. 1/A, Hungary		
David	Gabor	david@bnl.gov	Physics Department, Brookhaven National Laboratory (BNL), Upton, New York 11973-5000, USA		

The demo - comments

- Same technique can be used for other data stored in the PHENIX DB
- Question: what requirements do we have for the “phone book”? Virtually anything is possible with the exception of a dynamic query (however the content can be paginated alphabetically and grouped per institution etc, and search by tags is available)
- Export from PostgreSQL can be done periodically to keep the DAP site up to date until *we decide to make it the primary source of information*
- MGS process in its entirety can be ported to the DAP site with little effort
 - People would fill a YAML template and then a “git push” is done
- With a few helper scripts, the whole phone book functionality can be ported to the DAP site

The difficult part - the plots

- About 1600 plots with attached attributes and metadata
- For the attribute matrix please see one of the next slides
- Where and how to migrate? The DAP site vs Zenodo (or HEPData)
 - Migration to the DAP site will amount to ~200MB of graphics which is doable
 - Needs automation (manual option not feasible), requires development
- Zenodo - “single plot” migration path (brute force, manual)
 - In practice, not doable by one person, a huge chore and a distraction from other work
 - Feasible if a group of 8-10 students committed to this task over a period of weeks
 - Within the group, ~16 hours of hard work per person
- Zenodo - “single plot” migration path (automated)
 - Doable but highly involved
 - Need to learn the API and automate migration - quite a project
- Will likely need manual adjustment/cleanup in either case
- Benefit: **metadata for each plot can be corrected/tuned easily**

Migrating the plots, the batch option

- Two steps
 - Automated extraction from the database
 - Manual commit (automation would be an extra cost and unnecessary)
- HEPData as an option (instead of Zenodo)
 - Requires a collaboration note to be published (on arXiv?) and registered on Inspire
- To migrate in batches we need to define **groupings** which can carry keywords/attributes facilitating queries; this requires thought process
- If successful the migration would be a win
 - Durability (cf. we may be less lucky next time a BNL server outage happens)
 - Visibility and PR for PHENIX

Plots: how to group them for archival? Batch option.

PHENIX collaborators can use PHENIX only [Data Plot Search Form](#) and add a new data plot to the database using [Data Plot Entry Form](#).

Plot Status:

X-axis:

Y-axis:

Species :

Collision Species/Energy/Run Data : Multiple selections will do Logical AND operation of the selected keywords unless you check here ☐ for OR operation.

Run/Spec	p+p	Au+Au	d+Au	Cu+Cu	Cu+Au	U+U	He3+Au	p+Au	p+Al
Run-16		<input type="checkbox"/> 200GeV	<input type="checkbox"/> 200GeV <input type="checkbox"/> 62GeV <input type="checkbox"/> 39GeV <input type="checkbox"/> 20GeV						
Run-15	<input type="checkbox"/> 200GeV							<input type="checkbox"/> 200GeV	<input type="checkbox"/> 200GeV
Run-14		<input type="checkbox"/> 200GeV <input type="checkbox"/> 14.6GeV					<input type="checkbox"/> 200GeV		
Run-13	<input type="checkbox"/> 510GeV								
Run-12	<input type="checkbox"/> 510GeV <input type="checkbox"/> 200GeV				<input type="checkbox"/> 200GeV	<input type="checkbox"/> 193GeV			
Run-11	<input type="checkbox"/> 500GeV	<input type="checkbox"/> 200GeV <input type="checkbox"/> 27GeV <input type="checkbox"/> 19.6GeV							
Run-10		<input type="checkbox"/> 200GeV <input type="checkbox"/> 62.4GeV <input type="checkbox"/> 39GeV <input type="checkbox"/> 7.7GeV							
Run-9	<input type="checkbox"/> 500GeV <input type="checkbox"/> 200GeV								
Run-8	<input type="checkbox"/> 200GeV	<input type="checkbox"/> 9.2GeV	<input type="checkbox"/> 200GeV						
Run-7		<input type="checkbox"/> 200GeV							
Run-6	<input type="checkbox"/> 200GeV <input type="checkbox"/> 62.4GeV								
Run-5	<input type="checkbox"/> 410GeV <input type="checkbox"/> 200GeV			<input type="checkbox"/> 200GeV <input type="checkbox"/> 62.4GeV <input type="checkbox"/> 22.4GeV					
Run-4	<input type="checkbox"/> 200GeV	<input type="checkbox"/> 200GeV <input type="checkbox"/> 62.4GeV							

Plots: searchable fields and other metadata

J/psi cross-section vs. collision energy

Detailed description

J/psi differential cross-section vs. pT

Other comments and remarks

Searchable fields

- X-axis: momentum
- Y-axis: yield
- Species: jpsi
- Runs: p+p_510GeV_Run-13
- Theoretical curves: None
- Non-PHENIX data: None

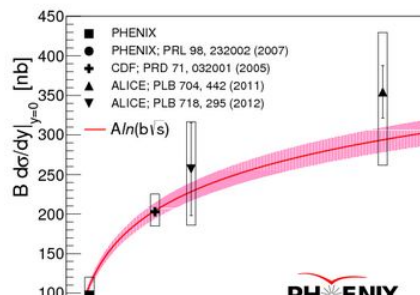
Supplementary information

- Preliminary was granted to Alexandre Lebedev on 2020-02-21
- PWG: Spin
- Analysis note(s) and PPG number(s) supporting this plot: an1365
- Published plot link:
- This plot is to be (was) shown at the conference: by Alexandre Lebedev at "[Winter Workshop on Nuclear Dynamics](#)"

Total Number of Figures: 1

NOTE: There may be an initial delay of up to 30 minutes for the plot files to be archived and the .jpg and .gif versions genera

Figure 1: J/psi cross-section vs. collision energy
linear horizontal scale



Summary

- Discussion of the internal site is left until later date
 - Many of similar arguments can be presented and similar strategies pursued
- It appears that we do need to improve the external site with regards to
 - Appearance (many obsolete looking parts)
 - Functionality (irrelevant material reduces usefulness)
 - Durability (currently subject to PHP whims and security holes)
- Solution: migration of relevant materials to the DAP website
 - To reap the benefits, **the legacy site needs to be deactivated after the migration**
 - **Static HTML** can be hosted at BNL if needed (in addition to GitHub) with zero concerns from Cybersecurity and/or support personnel
- There is an associated **cost** - weeks of development and validation time
- **Benefit** - a much better site requiring **zero maintenance** in perpetuity
- Can we make a decision? What are other priorities?