

# PHENIX Experience with Zenodo and the EIC perspective

Maxim Potekhin

BNL

*Nuclear and Particle Physics Software Group*

04/30/2020

# Overview

- Will keep this short (see backup slides for some detail)
- We observe synergy and common requirements in a few projects/experiments
  - I'm involved in PHENIX and EIC so have some perspective in these areas
  - Previous experience in DUNE
- Will briefly consider history, current situation and outlook
- EIC, PHENIX and other projects need to define action items for short and medium term
  - Not in a good position to wait for a solution for much longer

# 2019

- PHENIX starts its Data and Analysis Preservation (DAP) effort
  - Taking stock - fragmented and hard to navigate websites, overgrown Wikis, obsolete documentation in software and other areas, loss of records of know-how necessary for analysis, diminishing manpower
  - Custom apps (PHP) used to keep notes, papers and other publications
  - **Knowledge management** recognized as the key work area for both medium and long-term goals of PHENIX - maintaining quality of existing analyses as well as fulfilling the DAP mandate by the Lab and the funding agency
  - Participation in the DAP workshop at CERN in Fall'19
- BNL EIC group
  - White papers and other such materials uploaded to the Wiki, Drupal and even Indico
  - An instance of DocDB, older version, not much traffic
  - The need for a document management system recognized but not considered as pressing
  - Change of landscape with the start of the Yellow Report process

# 2020

- PHENIX

- Development of a permanent DAP Website with community involvement, started with run configuration, detector information etc, software tutorials and intros are in the works
  - Not a document management system
- Explored Invenio “a la sPHENIX” (a custom app), held meetings - TBD
- Converged on the decision to use the Zenodo instance at CERN, created a “PHENIX Community” on Zenodo, started with a low-hanging fruit - centralized storage and tagging of the PhD theses produced by the Collaboration (totaling 194)
- Decision to make the analysis software and analysis notes public deferred for later

- EIC

- An EIC GitHub organization established with a few dozen repositories and growing
- An EIC Software Group Website under construction (using GitHub pages)
- General use of collaborative tools has greatly expanded
- Yellow Report as a top-priority deliverable with more complex documents on the horizon
- The need to generate and systematize complex information has increased

# EIC vs PHENIX

- While PHENIX finished taking data a while ago and the EIC is still pretty far in future, they both have deadlines that drive document management requirements
  - Yellow Report and CDR for EIC
  - PHENIX working at reduced personnel levels and needing to preserve know-how both in near term (analysis) and long term (DAP) - this is not optional

# Platforms

- While helpful, the Wikis, Drupal, Dropbox etc not the right solutions
- DocDB has been popular in the past, heavily used in DUNE and still provides a baseline for evaluating features in other solutions
- In zeroth approximation, a document management application is a union of
  - Storage
  - Metadata
  - User Interface
- In that approximation, Zenodo is a drop-in replacement for DocDB
  - Plus many additional features and API
  - Minus agenda management which is no longer used anyway since taken over by Indico
- Policies: public vs restricted vs private

# First experience with Zenodo

- Meets the definition of a good system:
  - Simple things are “easy”
  - Complex things are “possible”
- Initial learning curve is quick and painless, the keyword function transparent
- Versioning
- ORCID and DOI capability off the bat
- Tiers of access
  - Private (locked in)
  - Restricted (by individual request)
  - Public
- “Communities”
- GitHub integration
- Search capabilities (many!)

# DOI, versions, keywords, conference-awareness

EXAMPLE

March 2, 2020

Poster Open Access

Edit

## Evolution of the Data Quality Monitoring and Prompt Processing System in the protoDUNE-SP experiment

Maxim Potekhin

The DUNE Collaboration has successfully implemented and currently operates an experimental program based at CERN which includes a beam test and an extended cosmic ray run of two large-scale prototypes of the DUNE Far Detector. The volume of data already collected by the protoDUNE-SP (the single-phase Liquid Argon TPC prototype) amounts to approximately 3PB and the sustained rate of data sent to mass storage is of the order of 0(100) MB/s. In addition to this data being committed to mass storage and processed in the Grid environment a small fraction of it is captured by the Prompt Processing System which is optimized for continuous low-latency calculation of the vital detector metrics and parameters as well as the output rendered as event display images. This system is the platform for Data Quality Monitoring in protoDUNE-SP and has served a crucial role starting from the commissioning of the apparatus and throughout its operation in 2018-2019, which continues at the time of writing. We present our experience in operating the system in the CERN environment, as well as work currently underway to make the system more scalable, resilient and to simplify system recovery procedures in preparation for the second run of protoDUNE-SP foreseen after the Long Shutdown of the LHC in the Fall of 2019.

The screenshot shows a document viewer interface with a blue header bar containing navigation icons and a 'Page: 1 of 1' indicator. The main content area displays a technical document with several sections:

- 3D event display based on raw data**
- TPC monitoring application with an extensive set of histograms at various levels of channel aggregation, FFT charts etc**
- Health monitor for the Front End Electronics motherboard**
- Event reconstruction with metrics such as track candidate count, hit count etc**
- Traxo candidate based Argon purity estimations with time series stored in the DB**
- Signal-to-noise ratio monitoring**
- Data preparation for the 3D event display implemented on a separate server**
- Out-of-band feeds from the hardware Argon purity monitor**

Below the text, there are several diagrams and charts:

- A 3D wireframe diagram of a detector structure.
- A '3D event display' showing particle tracks.
- A 'Dataflow Dashboard' with various status indicators.
- A grid of 12 heatmaps or histograms, with one labeled 'FFT'.

New version

18

views

11

downloads

[See more details...](#)

Indexed in

OpenAIRE

Publication date:

March 2, 2020

DOI:

DOI [10.5281/zenodo.3693788](https://doi.org/10.5281/zenodo.3693788)

Keyword(s):

[dqm](#) [prompt processing](#) [data quality monitoring](#) [DUNE](#)  
[neutrino](#) [Liquid Argon](#)

Meeting:

[24th International Conference on Computing in High Energy and Nuclear Physics \(CHEP 2019\)](#), Adelaide, Australia, 4-8 November 2019

License (for files):

[Creative Commons Attribution 4.0 International](#)

Versions

Version 1

[10.5281/zenodo.3693788](https://doi.org/10.5281/zenodo.3693788)

Mar 2, 2020

**Cite all versions?** You can cite all versions by using the DOI [10.5281/zenodo.3693787](https://doi.org/10.5281/zenodo.3693787). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)



# Recent Zenodo activities in PHENIX

- Approved for PHENIX Data and Analysis Preservation (EC, conveners)
  - Location of the host not considered an issue
- “Zenodo Communities” - see next slide - functional testing started
  - A “PHENIX Collaboration” community created, **started populating it with materials**
  - DAP site links to Zenodo
- Communication with the developers, looking for guidance regarding
  - Possible future data migration from Zenodo to Invenio RDM
  - Feature requests for community management
  - Storage allocation and use pattern discussion
- GitHub integration - “nice to have” but not core - initial testing done
  - Additional cloud replica of your GitHub release tagged with arbitrary metadata (discoverability)
  - Citeable via DOI

# Zenodo Community (another way to tag material)

- A way to organize material, and to consistently attribute materials to a collaboration/project/experiment - keeping a consistent brand
  - No need for multiple Zenodo instances?
- An improvement in visibility/discoverability/PR
  - An addition to the already existing metadata query aids in discovery of materials
- Anyone can upload a material to the community which is subject to **curation**
  - The curator gets notified and inspects the submission
    - If accepted, it becomes posted under the community umbrella
    - If rejected, it still remains on Zenodo site but is not officially owned/acknowledged by the community, this is an accordance to the “open access” platform
  - *There is currently one curator per community and there is no easy way to transfer this duty to a different account (something few people expected) but a fix is on the way according to the lead developer and other team members. Unofficial ETA is late 2020.*

# PHENIX Community on Zenodo

The screenshot shows the Zenodo interface for the PHENIX Collaboration Community. The top navigation bar is blue with the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. The user profile 'potekhin@bnl.gov' is visible in the top right. The main heading is 'The PHENIX Collaboration Community'. Below this is a 'Recent uploads' section with a search bar and a 'View' button. Three upload entries are listed, each with a date, version, and 'Open Access' tag. The first entry is 'PhenixCollaboration/web: First release of the PHENIX DAP site' by Maxim Potekhin, Ron Belmont, and amolhj, uploaded on April 21, 2020. The second entry is 'Transverse, Single-Spin Asymmetries for Charged Hadrons and for Muons from Open-Heavy-Flavor Decays in Polarized Proton-Proton and Proton-Nucleus Collisions in PHENIX' by Bok, Jeongsu, uploaded on April 20, 2020. The third entry is 'Measurements of  $\mu\mu$  pairs from  $cc$ ,  $b$  and Drell-Yan in p+p and p+Au collisions at  $\sqrt{s_{NN}} = 200$  GeV with PHENIX at RHIC' by Leung, Yue Hang, also uploaded on April 20, 2020. A 'More' button is at the bottom of the list. On the right, there is a 'New upload' button and a community profile for 'The PHENIX Collaboration Community' featuring the PHENIX logo. The profile text describes the community's mission to preserve and curate PHENIX data. It lists the curator as Maxim Potekhin and the curation policy as being done by members of the PHENIX DAP Task Force. The creation date is April 20, 2020, and the harvesting API is the OA-PMH interface. At the bottom right, there is a section titled 'Want your upload to appear in this community?' with a button to click.

zenodo Search Upload Communities potekhin@bnl.gov

## The PHENIX Collaboration Community

### Recent uploads

Search The PHENIX Collaboration Community

April 21, 2020 (v1.0) Software Open Access View

**PhenixCollaboration/web: First release of the PHENIX DAP site**

Maxim Potekhin, Ron Belmont, amolhj

This is the first release of the PHENIX DAP website

Uploaded on April 21, 2020

May 1, 2019 (v1) Thesis Open Access View

**Transverse, Single-Spin Asymmetries for Charged Hadrons and for Muons from Open-Heavy-Flavor Decays in Polarized Proton-Proton and Proton-Nucleus Collisions in PHENIX**

Bok, Jeongsu

Transverse single-spin asymmetry (TSSA) phenomena have gained substantial attention in several decades because they provide valuable information on the spin structure of the nucleon. Production of heavy flavor is dominated by gluon-gluon fusion in the leading order perturbative Quantum Chromodynamic

Uploaded on April 20, 2020

May 1, 2019 (v1) Thesis Open Access View

**Measurements of  $\mu\mu$  pairs from  $cc$ ,  $b$  and Drell-Yan in p+p and p+Au collisions at  $\sqrt{s_{NN}} = 200$  GeV with PHENIX at RHIC**

Leung, Yue Hang


Dilepton spectra are a classic probe to study ultra-relativistic heavy ion collisions. At  $\sqrt{s_{NN}} = 200$  GeV, the dimuon continuum is dominated by correlated pairs from semi-leptonic decays of charm and bottom hadrons and the Drell-Yan process. Measuring the azimuthal correlations of heavy flavor

Uploaded on April 20, 2020

More

New upload

Community



### The PHENIX Collaboration Community

The PHENIX Collaboration has initiated a Data and Analysis Preservation (DAP) effort in 2019. Within this scope there are a few areas of activity such as curating the available information, development of a new website to support DAP and to systematize and store available document for the long term. The latter is the main reason the PHENIX Zenodo community has been created.

**Curated by:**  
MaximPotekhin

**Curation policy:**  
Curation is done by members of the PHENIX DAP Task Force and contributors who join individual projects within the scope of DAP.

**Created:**  
April 20, 2020

**Harvesting API:**  
OA-PMH interface

Want your upload to appear in this community?

Click the button above to upload straight to

# Advanced search capabilities

By default, all searches are sorted according to an internal ranking algorithm that scores each match against your query. In both the user interface and REST API, it's possible to sort the results by:

- Most recent
- Publication date
- Title
- Conference session
- Journal
- Version

## Regular expressions

Regular expressions are a powerful pattern matching language that allow to search for specific patterns in a field. For instance if we wanted to find all records with a DOI-prefix 10.5281 we could use a regular expression search:

**Example:** `doi:10\.\d{4}\.*/`

Careful, the regular expression must match the *entire* field value. See the [regular expression syntax](#) for further details.

## Missing values

It is possible to search for records that either are missing a value or have a value in a specific field using the `_exists_` and `_missing_` field names.

**Example:** `_missing_:notes` (all records without notes)

**Example:** `_exists_:notes` (all records with notes)

## Advanced concepts

### Boosting

You can use the boost operator `^` when one term is more relevant than another. For instance, you can search for all records with the phrase *open science* in either *title* or *description* field, but rank records with the phrase in the *title* field higher:

**Example:** `title:"open science"~5 description:"open science"`

### Fuzziness

You can search for terms similar to but not exactly like your search term using the fuzzy operator `~`.

**Example:** `oepr~`

Results will match records with terms similar to `oeprn`, which would e.g. also match `open`.

### Proximity searches

A phrase search like `"open science"` by default expect all terms in exactly the same order, and thus for instance would not match a record containing the phrase *open access and science*. A proximity search allows that the terms are not in the exact order and may include other terms inbetween. The degree of flexibility is specified by an integer afterwards:

**Example:** `"open science"~5`

### Wildcards

You can use wildcards in search terms to replace a single character (using `?` operator) or zero or more characters (using `*` operator).

**Example:** `ope? scien*`

Wildcard searches can be slow and should normally be avoided if possible.

## Fields reference

The table below lists the data type of each field. Below is a quick description of what each data type means and what is possible.

- **string** Field does not require exact match (example field: `title`).

# Policy issues

- Zenodo defines itself as an open science platform i.e. for the most part public
- It does have access tiers: private, restricted and public
  - “Restricted” means that a request for access is forwarded to the owner
  - Not designed to handle “roles” for large groups of people
- Consider the fact that DocDB instances are often protected
  - In reality I would say 95%+ of materials don’t need to be protected
  - However in some cases (preliminary analysis etc) this may be important

# GitHub/Zenodo mechanics (see backup slides)

- A snapshot of a GitHub repo can be included in Zenodo organically+DOI
  - Integration/app link is in place: prepares and preserves tarballs of your releases
  - Makes your code easy to find (using the metadata) and to reference by a unique ID
  - Nice GUI
  - DOI reference to the code - becomes citeable
- Easy to use
  - Well-developed interface, I tested this functionality and it was quite simple
  - DOIs take some time  $O(10\text{min})$  to propagate to the DOI.org system, but this is not a problem

# Summary

- Convergence of requirements in EIC, PHENIX (and potentially sPHENIX)
- Interest in the EIC Software Group (including GitHub integration)
- This is urgent (cf. PHENIX started using the CERN instance but it's not too late to reconsider, and EIC needs this ASAP)
- Is Zenodo at BNL production-grade? Does it have full functionality?
- When can PHENIX migrate from CERN to BNL?
- Will there be multiple instances (and if yes, why)?
- Does it have same open policy wrt account creation and access?
- How much storage can we count on immediately, and in the long term?
- Quotas?
- What is the migration plan to Invenio RDM and do we need one?





# Terminology

- **Zenodo** is an open science data repository at CERN
  - In a nutshell, storage+metadata
  - Any data within the set limits
- **Invenio** is a toolkit used to in a number of CERN systems *including* Zenodo
  - A complex and capable framework.
  - Framework, not a system. *An application is needed to make use of its functionality.*
  - *cf. Zenodo is an Invenio-based application.*
- **Invenio RDM** (“research data management”) is a new product aiming to achieve
  - Portability (currently installing and configuring Invenio requires a high level of expertise)
  - Configurability i.e. eliminating the need for a custom app - a turnkey solution
  - ETA: late 2020

<https://zenodo.org/> - named after Ζηνόδοτος, inventor of metadata in 280 BC

The screenshot shows the Zenodo website interface. At the top, there is a blue header with the Zenodo logo, a search bar, and navigation links for 'Upload' and 'Communities'. A user profile dropdown menu is visible on the right, showing the email 'potekhin@bnl.gov'. Below the header, a notification banner states: 'Zenodo is continuing normal operation during the COVID-19 outbreak. All Zenodo staff are working remotely in accordance with preventive measures taken by CERN.' The main content area is divided into several sections:

- COVID-19 related communities:** This section features a card for 'Chicago COVID-19 Response'. It includes a thumbnail image of a virus, a description: 'This repository community collects research outputs and information objects relevant to the COVID-19 / SARS-CoV-2 efforts in Chicago. Users are encouraged to upload their research objects in this collection to facilitate sharing and discovery of information. Although Open Access articles and...', and a 'Curated by: saragon' note. There are 'Browse' and 'New Upload' buttons.
- Featured uploads related to COVID-19:** This section displays three featured items:
  - BIP4COVID19: Impact metrics and indicators for coronavirus related publications** (April 19, 2020, v1.1, Dataset, Open Access). Authors: Thanasis Vergoulis, Ilias Kanellos, Serafeim Chatzopoulos, Danae Pla Karidi, Theodore Dalamagas. Description: 'This dataset contains impact metrics and indicators for a set of publications that are related to the COVID-19 infectious disease and the coronavirus that causes it. It is based on: The CORD-19...'. Uploaded on April 24, 2020.
  - A Twitter Dataset of 179+ million tweets related to COVID-19 for open research** (April 12, 2020, v5.0, Dataset, Open Access). Authors: Banda, Juan M., Tekumalla, Ramya, Wang, Guanyu; Yu, Jingyuan; Liu, Tuo; Ding, Yuning; Chowell, Gerardo. Description: 'Due to the relevance of the COVID-19 global pandemic, we are releasing our dataset of tweets acquired from the Twitter Stream related to COVID-19 chatter. Since our first release we have received...'. Uploaded on April 24, 2020.
  - Code for Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing** (March 25, 2020, Software, Open Access). Authors: Ferretti, Luca; Wymant, Chris; Fraser, Christophe. Description: 'This code implements the COVID-19 mathematical analyses of Ferretti, Wymant et al. Science 2020. Namely, inference of the generation time interval for transmission pairs, solving the...'. Uploaded on April 24, 2020.

At the bottom of the featured uploads section, there is a button labeled 'Browse COVID-19 related research'.

**Recent uploads:** This section shows a card for 'Rosalyn Moran/Covid-19: Covid-19' (April 24, 2020, v1.0, Software, Open Access). Authors: Rosalyn Moran, Erik D. Fagerholm, Maell Cullen, Jean Daunizeau, Mark P. Richardson, Steven Williams, Federico Turkheimer, Rob Leech, Karl J. Friston. Description: 'Initial code release for 'Estimating required 'lockdown' cycles before immunity to SARS-CoV-2: Model-based analyses of susceptible population sizes, 'S0', in seven European countries including the UK and Ireland.' There is a 'View' button.

**Need help?:** A sidebar box with a 'Contact us' input field and the text: 'Zenodo prioritizes all requested related to the COVID-19 outbreak. We can help with: • Uploading your research data, software.'

# Zenodo “in a nutshell”

- General purpose digital repository
- Version control
- Data (storage space) + Metadata (DB)
- Extensive query capabilities
  - Full-text search is in the works
- DOI management (**doi.org** integration)
- ORCID-aware
- Gateway to other repositories
- GitHub integration (citeable code)
- Currently a service instance at CERN, being transformed into a more portable system under the “Invenio RDM” brand

## Zenodo in a nutshell

- **Research. Shared.** – all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
- **Citeable. Discoverable.** – uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
- **Communities** – create and curate your own community for a workshop, project, department, journal, into which you can accept or reject uploads. Your own complete digital repository!
- **Funding** – identify grants, integrated in reporting lines for research funded by the European Commission via OpenAIRE.
- **Flexible licensing** – because not everything is under Creative Commons.
- **Safe** – your research output is stored safely for the future in the same cloud infrastructure as CERN's own LHC research data.

# Zenodo: durability

Safe

**– more than just a drop box!**

Your research output is stored safely for the future in same cloud infrastructure as research data from CERN's [Large Hadron Collider](#) and using CERN's battle-tested repository software [Invenio](#), which is used by some of the world's largest repositories such as [INSPIRE HEP](#) and [CERN Document Server](#).

# Motivations

- Managing documents and other materials is a universal necessity in the field
  - Consider the needs of the Yellow Report working groups (papers, presentations, tables etc)
  - Not a replacement of the Wiki (which is not a document handling system in the first place)
- Not too many products exist in that area
  - DocDB is used at FNAL, BNL and a few other places, it's an aging product, no clear API
  - CERN CDS is not portable (NB shares the Invenio back-end with Zenodo)
- In EICUG here is currently not a single accepted solution or a policy
- The new EIC Software website is not designed as a general purpose document store (scalability, lack of proper metadata etc)
- Zenodo is an obvious contender

# Zenodo - GitHub panel - repo selection

The screenshot displays the Zenodo user interface for the GitHub panel. At the top, there is a blue navigation bar with the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. The user's email 'potekhin@bnl.gov' is visible in the top right corner. Below the navigation bar, the breadcrumb trail shows 'Home / Account / GitHub'. On the left side, there is a 'Settings' sidebar with options: Profile, Change password, Security, Linked accounts, Applications, Shared links, and GitHub (which is currently selected). The main content area is titled 'GitHub Repositories' and includes a '(updated now)' status and a 'Sync now...' button. A 'Get started' section contains three numbered steps: 1. Flip the switch (with a description and an 'ON' toggle), 2. Create a release (with a description and a 'create a release' link), and 3. Get the badge (with a description and a 'DOI 10.5281/zenodo.8475 (example)' badge). Below this, a 'Repositories' section lists several repositories with their respective GitHub links and 'OFF' toggle switches: BNLNPPS/BNLNPPS.github.io, BNLNPPS/BirdView, BNLNPPS/tpc-rs, DUNE/FNALCore, DUNE/Sandbox-TDR, and DUNE/SpaceCharge.

# Zenodo - GitHub panel - published release

The screenshot shows the Zenodo interface for a GitHub repository. The top navigation bar is blue with the Zenodo logo on the left, a search bar, and links for 'Upload' and 'Communities'. A user profile dropdown for 'potekhin@bnl.gov' is on the right. Below the navigation bar is a breadcrumb trail: 'Home / Account / GitHub / Repository'. On the left is a 'Settings' sidebar with options: Profile, Change password, Security, Linked accounts, Applications, Shared links, and GitHub (highlighted in blue). The main content area shows the repository 'PhenixCollaboration/web' with a blue 'ON' toggle. Below this is a 'DOI 10.5281/zenodo.3759876' label. A 'GitHub / Releases' section contains a table with one entry: 'v1.0 PhenixCollaboration/web: First release of the PHENIX DAP site', which is marked as 'Published' and was released '11 minutes ago'. A 'Create release ...' button is in the top right of the releases section.

zenodo Search Upload Communities potekhin@bnl.gov

Home / Account / GitHub / Repository

Settings

- Profile
- Change password
- Security
- Linked accounts
- Applications
- Shared links
- GitHub

PhenixCollaboration/web ON

DOI 10.5281/zenodo.3759876

GitHub / Releases Create release ...

v1.0 PhenixCollaboration/web: First release of the PHENIX DAP site	Published
DOI: 10.5281/zenodo.3759876	11 minutes ago
First release of the PHENIX DAP site	

# Zenodo - GitHub panel - published release browser

The screenshot shows the Zenodo interface for a software release. At the top, there is a blue header with the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. Below the header, the date 'April 21, 2020' is displayed on the left, and 'Software' and 'Open Access' tags are on the right. The main title is 'PhenixCollaboration/web: First release of the PHENIX DAP site'. Below the title, the authors 'Maxim Potekhin; Ron Belmont; amolhj' are listed, followed by the description 'This is the first release of the PHENIX DAP website'. A 'Preview' section is open, showing a file tree for 'web-v1.0.zip'. The tree includes a root folder 'PhenixCollaboration-web-c9d991e' with subfolders like '\_about', '\_analysis', and '\_data', and various files such as '.gitignore', 'Gemfile', 'LICENSE', 'README.md', 'contact.md', 'dap.md', 'howto.md', 'site.md', 'overview.md', '\_config.yml', 'vars.yml', 'detectors.yml', 'documents.yml', and 'gallery.yml'. File sizes are listed to the right of each item.

zenodo Search Upload Communities

April 21, 2020 Software Open Access

## PhenixCollaboration/web: First release of the PHENIX DAP site

Maxim Potekhin; Ron Belmont; amolhj

This is the first release of the PHENIX DAP website

Preview

web-v1.0.zip

- PhenixCollaboration-web-c9d991e
  - .gitignore 216 Bytes
  - Gemfile 285 Bytes
  - LICENSE 11.4 kB
  - README.md 1.6 kB
  - \_about
    - contact.md 324 Bytes
    - dap.md 1.3 kB
    - howto.md 6.9 kB
    - site.md 2.5 kB
  - \_analysis
    - overview.md 114 Bytes
  - \_config.yml 1.3 kB
  - \_data
    - acc
      - vars.yml 1.4 kB
    - detectors.yml 2.6 kB
    - documents.yml 3.5 kB
    - gallery.yml 4.5 kB



# DOIs are an increasingly popular way to reference software

Persistent, durable link to archived software, can be nicely embedded in any page.



README.md updated the README yesterday

\_config.yml Reworked references to "assets", added flexlink to assets in the tab... 8 days ago

index.html Added includes, corrected README 3 months ago

## The PHENIX Collaboration Data and Analysis Preservation Website

This website is designed to further the goals of the long-term Data and Analysis Preservation (DAP) of the PHENIX Collaboration.

The site is under heavy development and is not officially in production. At present, and until further notice, it is not expected to be 100% functional or have immediately useful content.

Release archival on Zenodo: [DOI 10.5281/zenodo.3759876](https://doi.org/10.5281/zenodo.3759876)

## SCOPE

We use the Jekyll static site generator.

The site is intended to preserve curated documentation for the PHENIX experiment, including technical write-ups on the PHENIX software and its infrastructure. It is not a document server although it does host a limited number of documents (primarily in PDF formats) and as well as some diagrams.

Please note that the static nature of the site also implies lack of common database query functions at runtime, authentication and authorization etc. Where needed, such services will be hosted separately and links will be provided.

## TECHNICAL

### For the Developer

Please see the "how-to" section in the "About" menu of the site for the information being constantly updated.

### Gems

Pay attention to the following dependencies (need to be installed and also included in the Gemfile in this folder):

```
gem "jekyll", "~> 4.0"
gem "jekyll-mentions", "~> 1.5", ">= 1.5.1"
gem "jekyll-sitemap", "~> 1.4"
gem "jekyll-redirect-from", "~> 0.16.0"
```

# GitHub/Zenodo integration benefits

- Not a core functionality by a long shot, however...
- ...provides a uniform way to reference digital products using DOI
- ...metadata is a good thing to have - better discoverability!
- ...can leverage the Zenodo “community” feature to organize materials and increase visibility
  - Cf. simulated data and the code used to produce it can be kept under the same umbrella
- Longer term - Data and Analysis Preservation
- In general, an “EIC Software” community on Zenodo may be a useful thing to have (papers, conference presentations etc)