

# Experience in leveraging Zenodo as the ePIC Document Database

M.Potekhin

BNL Nuclear and Particle Physics Software Group (NPPS)

*EIC and RHIC experiments collaborative tools and technologies*

06/17/2028

# Overview

- © The need for an effective document management system was realized by the ePIC Collaboration early on.
- © The decision to leverage Zenodo for this purpose was made in 2024, after a review process and based on BNL experience in a few other projects including RHIC (cf. PHENIX Collaboration committed more than 750 records to Zenodo). It was determined that Zenodo meets the current ePIC requirements and has an extensive set of capabilities.
- © There has been a convergence between the original Zenodo, hosted as a service at CERN, and the evolved underlying project, the Invenio. This opens new possibilities for future hosting and migration of content.
- © These presentation summarizes our experience and outlook in this area.
- © Please see the backup section for supporting materials.

# Key facts about Zenodo use in ePIC

- ◎ Zenodo instance at CERN was chosen by ePIC for three main reasons:
  - A well established service with with a large user base and solid industry-level support; easy to manage, low overhead.
  - According to a thorough review, meets the ePIC requirements.
  - Minting of DOIs is free of charge for us (it's not cheap).
- ◎ Informed by prior experience at RHIC (PHENIX), we are using a curated collection of keywords to standardize indexing of the materials committed to Zenodo, create automated query links, and simplify queries
  - The policy of adding standard keywords to the ePIC items is currently not strictly enforced by the management team, and this could be improved.
- ◎ There are 147 items committed to Zenodo by ePIC, at the time of writing. The principal way of managing this is via the “ePIC Community” (next slide)

# Community management

- ◎ The Zenodo “community” is a collection of documents managed by a research organization via its appointed representatives. The ePIC Collaboration is actively using this feature in its activities.
- ◎ In the current version of Zenodo, access levels have been simplified, now having only two categories, “public” and “restricted”.
- ◎ Levels of management and access:
  - Owner: has all powers including appointing managers
  - Manager: manage members, curate records, view restricted records
  - Curator: curate records, view restricted records
  - Reader: view restricted records
- ◎ The management team is responsible for review and approval of the materials submitted to the community.

## Access control

- © As already mentioned, the ePIC Zenodo management team is responsible for reviewing and approving materials to be accepted into the community.
- © Another function of the management team is to approve access of ePIC members to restricted material. The entry bar for an ePIC member is extremely low, as it amounts to a creation of a Zenodo account which is simple and transparent. The procedure of member authorization in ePIC is well established, streamlined and takes place rather quickly.

## Zenodo: additional capabilities

- © Software citation: Zenodo has a level of integration with GitHub which allows to assign DOIs to software releases, making the software citeable in a durable and straightforward manner. This is now an established practice in the scientific community and it brings many benefits to both authors (proper credits and career support) and projects. Looking into that feature, for ePIC.
- © Ability to register datasets (due to overall flexibility regarding the formats and packaging). This maps well onto efforts such as datacard development, and data preservation and reuse in general (new work area, ramping up).
- © Comment and review capabilities: while still in development, these do exist and work well for ePIC, with recent enhancements such as:  
<https://blog.zenodo.org/2026/04/13/2026-04-13-latest-platform-updates/>
- © Conference references are handled organically.

# Zenodo: feasibility of content migration to InvenioRDM

- © Having the ability to migrate content committed to Zenodo to an alternative (compatible) platform, hosted elsewhere, was one of the key requirements that ePIC had put forth before adopting Zenodo – to ensure our decision was future-proof. The essential issue here is the compatibility of the metadata schema, and we have looked into this (see the backup section).
- © After the migration of the Zenodo service to the updated Invenio infrastructure, both share the same general JSON schema for metadata.  
Caveats:
  - Conference data (preserved in JSON) is not handled in the InvenioRDM UI, but this can be added by extending applications built with the RDM backend
  - Comments added in the review workflow are not exported to JSON, hence can't be migrated
- © Any future solution will need the software citation and dataset registration features (see above) to meet the needs of the ePIC Collaboration.

# Summary and outlook

- © Adoption of Zenodo as the main document management tool allowed ePIC to sustain fast-paced development in its many work areas over the past two years.
- © The choice of the hosting platform (i.e. one at CERN) was motivated by it meeting the ePIC requirements and its ample set of capabilities presented above, plus simplicity of operations.
- © Prior differences between Invenio and Zenodo have largely disappeared and specificity of Zenodo is properly handled in the API and the updated metadata schemas. This opens the way for future migration of the ePIC body of materials to a different hosting solution should this become necessary.
- © Use of DOIs has always been high on the requirements list and will need to be resolved accordingly, should such migration happen.

# Backup

## Experience in PHENIX – hosting and uptake

- ◎ Zenodo plays a crucial role in the successful Data and Analysis preservation effort in PHENIX. It replaced a legacy PHP/DB-based document catalog which had a limited feature set, limited visibility and was increasingly difficult to maintain.
- ◎ As an international collaboration, a strategic decision was made by PHENIX in 2020 to utilize the Zenodo instance at [CERN](#).
- ◎ The learning curve turned out to be exceptionally simple.
- ◎ At the time of writing, [757](#) items produced by PHENIX have been committed and indexed with [curated keywords](#), including more than 160 PhD thesis documents, a large number of the conference presentations, and some tutorials and technical write-ups.

## DOIs

- ◎ Using the DOIs facilitates the long-term persistence of the materials committed to various repositories as it is *conducive* to designing and implementing future migrations of the repository platforms. A DOI (included in a link) will be *resolved* to the same document, for posterity — if everything is done right. Example: [10.5281/zenodo.7052131](https://doi.org/10.5281/zenodo.7052131)
- ◎ Minting the DOI is done by external agencies and typically costs money. It's covered by CERN when the CERN Zenodo instance is used. For projects sponsored by DOE, a viable alternative is to use DOIs minted by OSTI.
- ◎ Use of DOIs is stipulated on all Invenio-based platforms.

## Experience in PHENIX – integration with the website

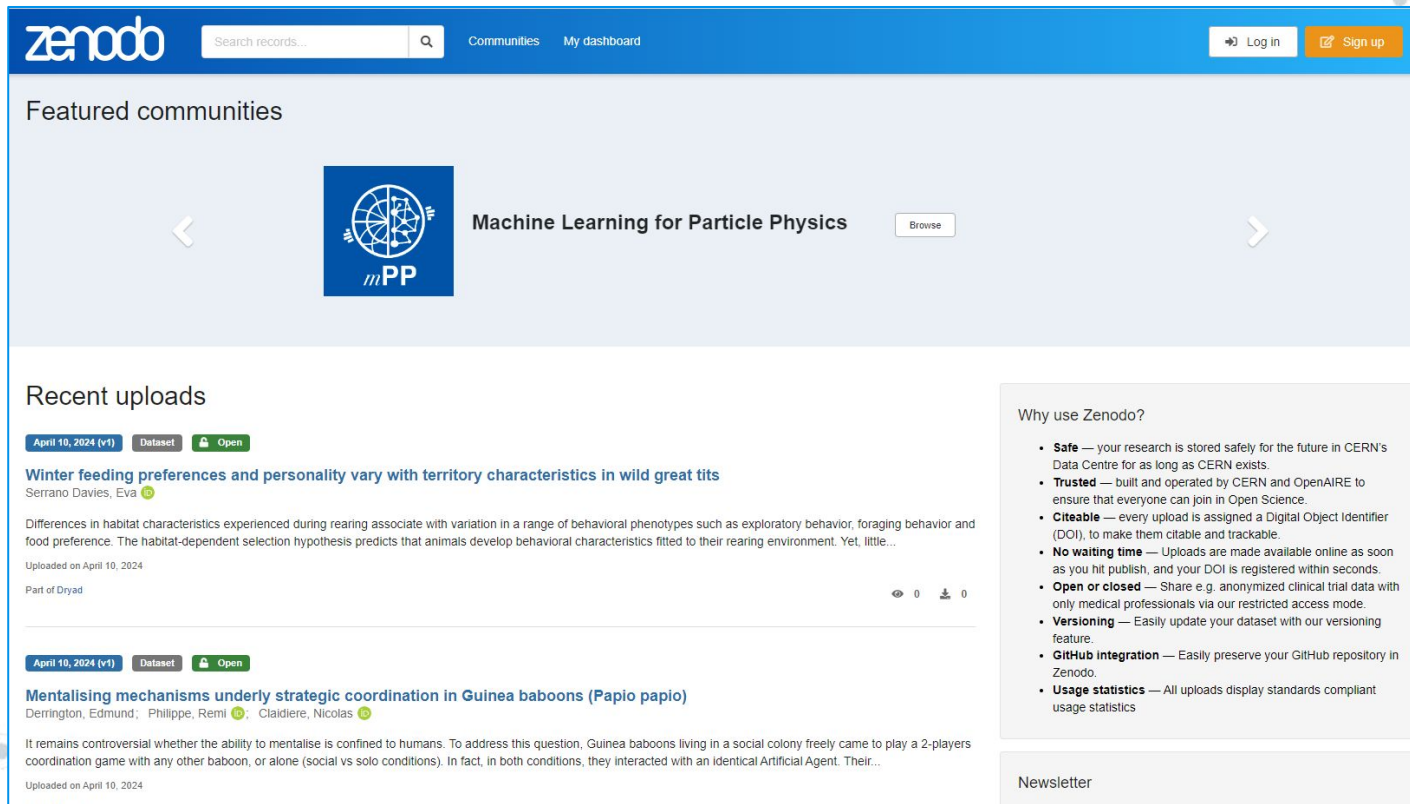
- © Starting in 2019, the PHENIX Collaboration migrated its website to a new platform based on Jekyll technology (cf. GitHub Pages)
- © Reliance on Zenodo for content storage allowed the Collaboration to create a lightweight, portable and easy to maintain website, since it by itself does not contain much data.
- © The website is used to locate and serve PHENIX content from Zenodo using a number of auto-generated links based on a *collection of curated keywords*.

## Zenodo use in other projects with BNL involvement

- © The EIC User Group (EICUG) solved its long-standing problem with document management (previously in ad-hoc storage in Drupal) by migrating to Zenodo in 2022.
- © The LuSEE-Night project (astrophysics) which has considerable leadership and science contributions from BNL, has also started using the CERN instance of Zenodo in 2023.
- © Successful use of Zenodo in these projects contributed to the motivation in ePIC to consider this platform.

# <https://zenodo.org/> – landing page

## Named after Ζηνόδοτος, the inventor of metadata (280 BC)



The screenshot shows the Zenodo landing page with a blue header. The header includes the Zenodo logo, a search bar, and navigation links for 'Communities' and 'My dashboard'. There are 'Log in' and 'Sign up' buttons in the top right.

**Featured communities**

The featured community is 'Machine Learning for Particle Physics' (mPP), represented by a blue square logo with a white circular diagram and the letters 'mPP' below it. A 'Browse' button is visible to the right of the community name.

**Recent uploads**

Two recent uploads are listed:

- Winter feeding preferences and personality vary with territory characteristics in wild great tits**  
Serrano Davies, Eva  
Differences in habitat characteristics experienced during rearing associate with variation in a range of behavioral phenotypes such as exploratory behavior, foraging behavior and food preference. The habitat-dependent selection hypothesis predicts that animals develop behavioral characteristics fitted to their rearing environment. Yet, little...  
Uploaded on April 10, 2024  
Part of Dryad
- Mentalising mechanisms underly strategic coordination in Guinea baboons (Papio papio)**  
Derrington, Edmund; Philippe, Remi; Claidiere, Nicolas  
It remains controversial whether the ability to mentalise is confined to humans. To address this question, Guinea baboons living in a social colony freely came to play a 2-players coordination game with any other baboon, or alone (social vs solo conditions). In fact, in both conditions, they interacted with an identical Artificial Agent. Their...  
Uploaded on April 10, 2024

**Why use Zenodo?**

- **Safe** — your research is stored safely for the future in CERN's Data Centre for as long as CERN exists.
- **Trusted** — built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science.
- **Citeable** — every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable.
- **No waiting time** — Uploads are made available online as soon as you hit publish, and your DOI is registered within seconds.
- **Open or closed** — Share e.g. anonymized clinical trial data with only medical professionals via our restricted access mode.
- **Versioning** — Easily update your dataset with our versioning feature.
- **GitHub integration** — Easily preserve your GitHub repository in Zenodo.
- **Usage statistics** — All uploads display standards compliant usage statistics

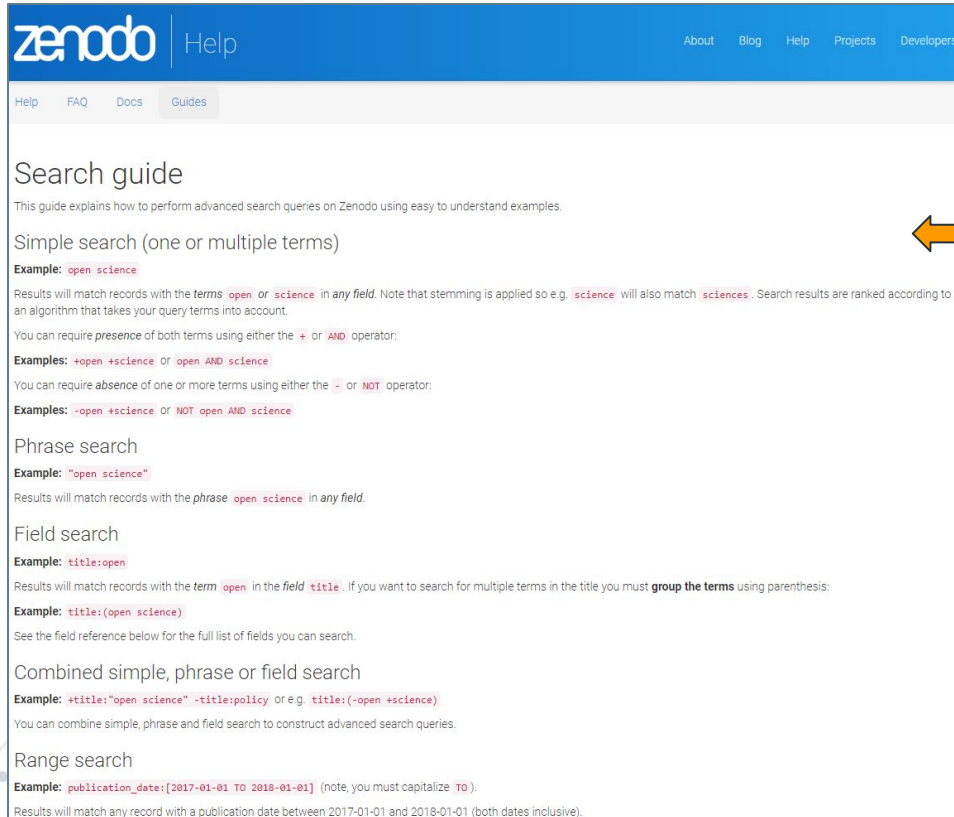
**Newsletter**

## Zenodo: key features

- ◎ **Search:** Zenodo possesses powerful search capabilities including “elasticsearch”, detailed search on the metadata and an unlimited number of optional keywords added to committed items.
- ◎ **Data Format Flexibility:** data in any format can be handled, either as individual files or as groups of files. Datasets, ROOT files, Docker images, pictures, videos etc.
- ◎ **Curated Communities:** an organization can create a managed collection of documents, and enhance its brand and findability
- ◎ **Access control:** a few access levels managed by the organization
- ◎ **DOI:** one of the factors to ensure durability of the materials (see the DOI slide below for more info)

# Zenodo: search functionality, extensively documented

<https://help.zenodo.org/guides/search/>



The screenshot shows the Zenodo help page for search. The header includes the Zenodo logo and navigation links for About, Blog, Help, Projects, and Developers. Below the header is a sub-navigation bar with links for Help, FAQ, Docs, and Guides. The main content area is titled "Search guide" and contains several sections: "Simple search (one or multiple terms)", "Phrase search", "Field search", "Combined simple, phrase or field search", and "Range search". Each section provides examples and explanations of search syntax. An orange arrow points from the right side of the slide to the "Simple search" section.

zenodo | Help

About Blog Help Projects Developers

Help FAQ Docs Guides

## Search guide

This guide explains how to perform advanced search queries on Zenodo using easy to understand examples.

### Simple search (one or multiple terms)

**Example:** `open science`

Results will match records with the **terms** `open` or `science` in **any field**. Note that stemming is applied so e.g. `science` will also match `sciences`. Search results are ranked according to an algorithm that takes your query terms into account.

You can require **presence** of both terms using either the `+` or **AND** operator:

**Examples:** `+open +science` or `open AND science`

You can require **absence** of one or more terms using either the `-` or **NOT** operator:

**Examples:** `-open +science` or `NOT open AND science`

### Phrase search

**Example:** `"open science"`

Results will match records with the **phrase** `open science` in **any field**.

### Field search

**Example:** `title:open`

Results will match records with the **term** `open` in the **field** `title`. If you want to search for multiple terms in the title you must **group the terms** using parenthesis:

**Example:** `title:(open science)`

See the field reference below for the full list of fields you can search.

### Combined simple, phrase or field search

**Example:** `+title:"open science" -title:policy` or e.g. `title:(-open +science)`

You can combine simple, phrase and field search to construct advanced search queries.

### Range search

**Example:** `publication_date:[2017-01-01 TO 2018-01-01]` (note, you must capitalize `TO`).

Results will match any record with a publication date between 2017-01-01 and 2018-01-01 (both dates inclusive).

Just a fragment of the help page  
– lots of information there

Rich, flexible functionality

# PHENIX: keyword references to Zenodo

## PHENIX Keywords

Listed on this page are *recommended* keywords used for two purposes:

- to **tag materials** placed on this site so they can be consistently referenced across this web resource
- to enhance discoverability of the PHENIX materials committed to [Zenodo](#)

Consistent use of the keywords is strongly recommended. Please note that [HEPData materials](#) follow a *different set of conventions*.

**The keywords are case-sensitive.** We adopted lowercase convention for all keywords for the following reason: Zenodo is using a complex query syntax so it's best to avoid ambiguity. Note that Zenodo keywords can actually be a combination of words and white space (i.e. phrases). Multiple such combinations are possible.

In the tables below, the keywords are grouped in categories. Each entry in the left column acts as a query link to [Zenodo](#), for that specific keyword.

### General (12 items)

Keyword	Description
<a href="#">alice</a>	ALICE - an experiment at CERN
<a href="#">atlas</a>	ATLAS - an experiment at CERN
<a href="#">bup</a>	Beam Use Proposal
<a href="#">cms</a>	CMS - an experiment at CERN
<a href="#">decadal plan</a>	Two long-term PHENIX research proposals
<a href="#">phenix</a>	PHENIX - an experiment at RHIC
<a href="#">phobos</a>	PHOBOS - an experiment at RHIC
<a href="#">rhic</a>	Relativistic Heavy Ion Collider (RHIC)
<a href="#">sphenix</a>	sPHENIX - an experiment at RHIC
<a href="#">star</a>	STAR - an experiment at RHIC
<a href="#">tutorial</a>	a category of PHENIX documents on Zenodo
<a href="#">wa98</a>	WA98 - an experiment at CERN

### Conferences (117 items)

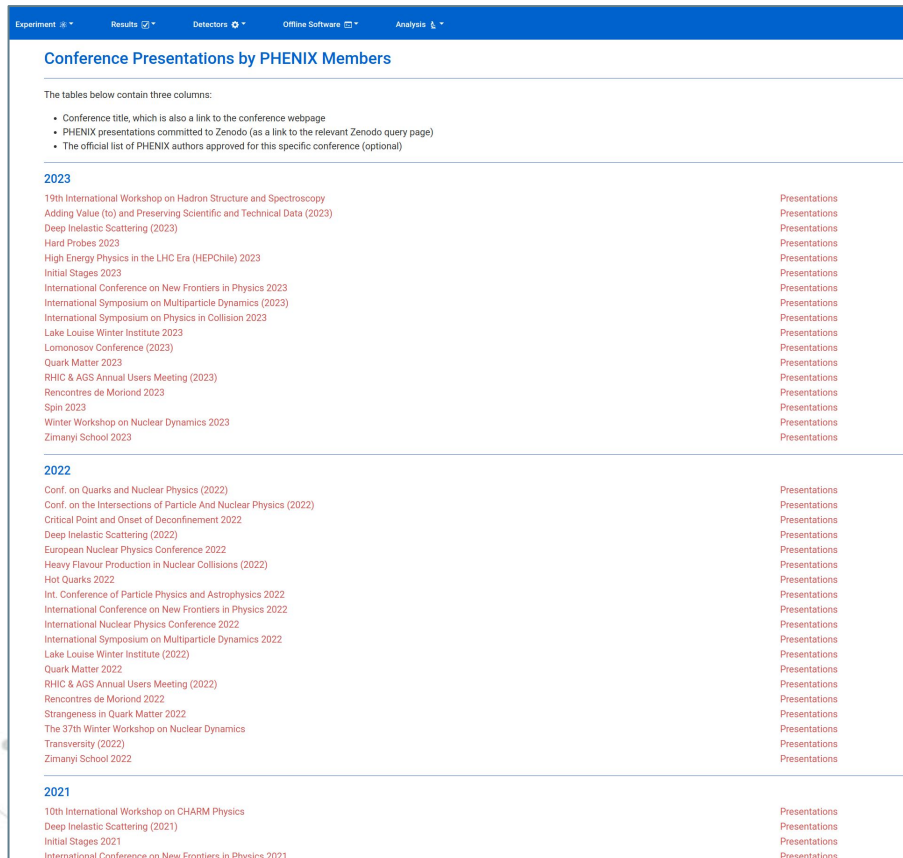
Keyword	Description
<a href="#">aum16</a>	RHIC & AGS Annual Users Meeting (2016)
<a href="#">aum17</a>	RHIC & AGS Annual Users Meeting (2017)
<a href="#">aum18</a>	RHIC & AGS Annual Users Meeting (2018)

Keywords are classed as “general”, “conferences”, “detector”, “physics”, “software” etc, for ease of navigation.

They are rendered as **functioning links to Zenodo queries**.

Keywords are documented in the easy to read and maintain YAML files, which are rendered automatically on the website.

# The conference page



Experiment Results Detectors Offline Software Analysis

## Conference Presentations by PHENIX Members

The tables below contain three columns:

- Conference title, which is also a link to the conference webpage
- PHENIX presentations committed to Zenodo (as a link to the relevant Zenodo query page)
- The official list of PHENIX authors approved for this specific conference (optional)

Year	Conference Title	Presentations
2023	19th International Workshop on Hadron Structure and Spectroscopy	Presentations
	Adding Value (to) and Preserving Scientific and Technical Data (2023)	Presentations
	Deep Inelastic Scattering (2023)	Presentations
	Hard Probes 2023	Presentations
	High Energy Physics in the LHC Era (HEPChile) 2023	Presentations
	Initial Stages 2023	Presentations
	International Conference on New Frontiers in Physics 2023	Presentations
	International Symposium on Multiparticle Dynamics (2023)	Presentations
	International Symposium on Physics in Collision 2023	Presentations
	Lake Louise Winter Institute 2023	Presentations
	Lomonosov Conference (2023)	Presentations
	Quark Matter 2023	Presentations
	RHIC & AGS Annual Users Meeting (2023)	Presentations
	Rencontres de Moriond 2023	Presentations
	Spin 2023	Presentations
Winter Workshop on Nuclear Dynamics 2023	Presentations	
Zimanyi School 2023	Presentations	
2022	Conf. on Quarks and Nuclear Physics (2022)	Presentations
	Conf. on the Intersections of Particle And Nuclear Physics (2022)	Presentations
	Critical Point and Onset of Deconfinement 2022	Presentations
	Deep Inelastic Scattering (2022)	Presentations
	European Nuclear Physics Conference 2022	Presentations
	Heavy Flavour Production in Nuclear Collisions (2022)	Presentations
	Hot Quarks 2022	Presentations
	Int. Conference of Particle Physics and Astrophysics 2022	Presentations
	International Conference on New Frontiers in Physics 2022	Presentations
	International Nuclear Physics Conference 2022	Presentations
	International Symposium on Multiparticle Dynamics 2022	Presentations
	Lake Louise Winter Institute (2022)	Presentations
	Quark Matter 2022	Presentations
	RHIC & AGS Annual Users Meeting (2022)	Presentations
	Rencontres de Moriond 2022	Presentations
Strangeness in Quark Matter 2022	Presentations	
The 37th Winter Workshop on Nuclear Dynamics	Presentations	
Transversity (2022)	Presentations	
Zimanyi School 2022	Presentations	
2021	10th International Workshop on CHARM Physics	Presentations
	Deep Inelastic Scattering (2021)	Presentations
	Initial Stages 2021	Presentations
	International Conference on New Frontiers in Physics 2021	Presentations

Using Zenodo, we preserved presentations by PHENIX members, made at **117** conferences (and counting), with modest effort.

The conference page contains **links to collections of slides** for specific conferences, as well as to the conference pages themselves. All rendering of the YAML content is **100% automatic**.

Reliance on the Zenodo instance at CERN provides a degree of certainty of durability of the materials thus preserved, and transparent access for years to come.

# An example of a PHENIX Zenodo entry (conference slides)

zenodo Search records... Communities My dashboard Log in Sign up

PHENIX PHENIX Collaboration

Published September 5, 2022 | Version v1 Presentation Open

## Study of forward J/Psi production vs event multiplicity in p+p/A collisions at PHENIX

Liu, Ming<sup>1</sup> Show affiliations

Talk presented at the Conference on the Intersections of Particle and Nuclear Physics (CIPANP) 2022

### Files

Ming\_PHENIX-JPsi-multiplicity-CIPANP-092022-v2.pdf

PHENIX

### Study of Forward J/ $\psi$ Production vs Event Multiplicity in p+p/A Collisions at PHENIX

Ming X. Liu  
Los Alamos National Laboratory  
(for the PHENIX Collaboration)  
CIPANP 2022

Files (4.4 MB)

Name	Size	Download all
Ming_PHENIX-JPsi-multiplicity-CIPANP-092022-v2.pdf md5:49c789269a4cad5de72ee96f247d2ec7	4.4 MB	Preview Download

### Citations

Show only: Literature (0) Dataset (0) Software (0) Unknown (0) Citations To This Version

No citations found

16 VIEWS 12 DOWNLOADS Show more details

### Versions

Version	Published
Version v1 10.5281/zenodo.7052131	Sep 5, 2022

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.7052130. This DOI represents all versions, and will always resolve to the latest one. Read more.

### External resources

Indexed in OpenAIRE

### Communities

PHENIX Collaboration

### Keywords and subjects

ciparp22 phenix J/ $\psi$  multiplicity p+p p+al p+au qcd gluon charmonium forward rapidity dimuon cpc fvtx

### Details

DOI 10.5281/zenodo.7052131

Resource type Presentation

Publisher Zenodo

Conference Conference on the Intersections of Particle and Nuclear Physics (CIPANP) 2022 (CIPANP) , Lake Buena Vista, FL, USA, Aug. 29 - Sep. 4, 2022

## Notable features:

- Custom keywords
- Slide viewer
- Community attribution
- Reference to the conference

# The 2023 Platform Migration and its impact

- © After a long convergence between Zenodo and Invenio, there was migration of Zenodo from its legacy platform (derived from older versions of Invenio) to the newer Invenio standard in October 2023. This impacted the metadata format and the API. For example, data items committed to Zenodo prior to the migration have a number of legacy attributes that are not directly supported in the current version.
- © However, it is important to note that this is accommodated in the API and the updated JSON schema for the metadata import and export, so for migration purposes the differences can be gracefully resolved by the application logic.
- © See examples of mitigation strategies in the next two slides.

## Example 1: handling of the discontinued “embargo” attribute

**Old API:** a single `access_right: "embargoed"` field plus `embargo_date`.

**New API:** the record has `access.files: "restricted"` and an `access.embargo` sub-object, for example:

```
"access": {  
  "record": "public",  
  "files": "restricted",  
  "embargo": {  
    "active": true,  
    "until": "2025-12-31",  
    "reason": "Publisher embargo"  
  },  
  "status": "embargoed"  
}
```

## Example 2: handling of the conference info

In the InvenioRDM/Zenodo record JSON schema, conference data is stored under `custom_fields` with the key `meeting:meeting`.

For example:

```
"custom_fields": {  
  "meeting:meeting": {  
    "acronym": "HMC",  
    "dates": "2024-11-04/2024-11-06",  
    "place": "Virtual",  
    "session": "Findability D2",  
    "title": "Helmholtz Metadata Collaboration Conference 2024",  
    "url": "https://events.hifis.net/event/1627"  
  }  
}
```

## Zenodo vs document development workflow systems

- ◎ Zenodo is a repository. It provides a robust versioning system.
- ◎ At the same time, it is explicitly not a document development environment where “issues” can be created and tracked. However, upon submission to a community a thread of **comments** can be created by the curators prior to publishing, which covers many use cases.
- ◎ Such tasks can and are typically done using the issue tracker on **GitHub**, **Overleaf**, **Google Docs** etc. There is a positive experience with all these platforms.
- ◎ Documents are typically uploaded to Zenodo when they are in the final or close to final shape, with possibility of revisions – each assigned a unique DOI.

# Using YAML for the website/Zenodo integration

```
- name: wwnd24
  title: 'The 39th Winter Workshop on Nuclear Dynamics (2024)'
  url: 'https://indico.cern.ch/event/1345629/'
  year: 2024

- name: zs23
  title: 'Zimanyi School 2023'
  url: 'http://zimanyischool.kfki.hu/23/'
  year: 2023

- name: pic23
  title: 'International Symposium on Physics in Collision 2023'
  url: 'https://indico.cern.ch/event/1190468/'
  year: 2023

- name: spin23
  title: 'Spin 2023'
  url: 'https://indico.jlab.org/event/663/'
  year: 2023

- name: qm23
  title: 'Quark Matter 2023'
  url: 'https://indico.cern.ch/event/1139644/'
  year: 2023
```

Conversion of the conference keywords to Zenodo queries is automated using simple macros.