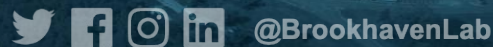




# InvenioRDM at the SDCC

Uma Ganapathy, Louis Pelosi  
V. Garonne, R. Hancock, J. Lauret, C. Lepore, M. Poat



# What is InvenioRDM?

- Turn-key research data management repository
- Community framework allowing for deep customization
- Ability to store and catalog digital data such as documents, images, datasets, etc.
- Contributions are made from many institutions, the development is driven by the very users of the framework
- Support for revisions, collaboration, updating on cataloged information
- Support for multiple authentication methods

# Research Data Management

## What's the purpose of RDM?

- Disseminate & archive
- Enable reproducibility
- Enable reuse



Half the links in papers are inaccessible after 10 years!

89% of 53 landmark cancer research papers irreproducible!



Ships logs from 18th century used for climate research

= Open Science Enabler!

# What makes it different

## Highly scalable:

- Can handle uploads up to 100 million records and several petabytes
- Accepts any format - the infrastructure behind it will support
- Versioning support

## Highly customizable:

- Style and theme can be easily edited to reflect user preference
- Custom vocabularies for resource types, data types, roles, languages, etc
- Authentication can easily be customized with simple configuration settings

**REST API** provides the ability to develop future integrations where needed/requested extending its usage

**Built with OAI-PMH** (Open Archives Initiative Protocol for Metadata Harvesting) to allow metadata harvesting of records

- Records can be uploaded as Metadata-only record or with metadata and files

# (continued)

## Search:

- Support for customizable faceted search
- Advanced query syntax allows search by term, phrase, range, etc
- Auto-completion to make it easy for users to enter metadata. Ability to link to authority records for names, affiliations, licenses and more.

**Citation** formatting allows for customized generating of citation strings using the Citation Style Language (800+ journal citation styles)

Quick **previews** for common file formats

- Including record previewing before publishing

## Access control:

- Restricted records access by user role, community access, etc
- Secret links can be created for collaboration on restricted records
- Restricted records can be set to public on a set date

## Security:

- Users can view a list of devices logged in with
- Security policy by InvenioRDM for reporting/fixing/announcing vulnerabilities when they occur

# SDCC - sPHENIX showcase

- sPHENIX had a critical need for a document repository to secure vital internal collaboration notes, external presentation reviews, scientific papers, and more. While default solutions fell short to meet the needs, both technical and non-technical, the SDCC provided a necessary solution
- The SDCC has taken the base InvenioRDM framework and extended certain features to meet the needs of researchers and publishers alike. Most of these extensions involved code additions/changes while a small part of it was through configuration.

## Login and streamlining users' access

- Information from the centralized organizational account is utilized for another level of access control for the login process and automatic addition of members to certain restricted communities.
- Addition of an authorization layer to allow only sPHENIX experiment members to login
  - This check is made at every login attempt ensuring that up to date membership is considered
- Upon login, members are added to the restricted “sPHENIX” community
  - This ensures that assigned communities are available for all authenticated users
  - Additional communities can be created by members for specialized interests

# SDCC - Extension of Collaborative comments

- InvenioRDM provides community members the ability to engage in centralized dialogue, for each record under review. This conversation is further enhanced by the SDCC implementation to provide nested replies for comments.
- Nested comments supported up to 3-level deep
- This enhances the discussion by community members, when the record is being reviewed before its publication

The screenshot displays a conversation thread with the following structure:

- Conversation** / **Record**
- Comment 1:** User [redacted] commented 14 minutes ago. Text: "The abstract need to be extended".
- Reply 1:** User [redacted] commented 13 minutes ago. Text: "The abstract need to be extended - This will be worked on after file uploads are done and close to finalizing documents".
- Reply 2:** User [redacted] commented 13 minutes ago. Text: "The abstract need to be extended - Files will be uploaded soon.".
- SubReply 1:** User [redacted] commented 11 minutes ago. Text: "The abstract need to be extended - Files will be uploaded soon - Uploaded files.".
- Reply 3:** User [redacted] commented 12 minutes ago. Text: "Please review the metadata and file uploads".

Navigation controls include "Replies" and "SubReplies" dropdowns.

# SDCC - Community Access Level

## Addition of a new access level to Community

- A 'restricted' community in InvenioRDM has a few access levels for its members.
- In addition to the existing ones, another access level 'Reviewer' has been added.
- This level would allow the members to take part in records' review conversations but not allow them to accept or decline records under review to the community.
- This allows collaboration of members in the review process while limiting the curation of records only to members with a higher access level.



# SDCC - Extension of Search on 'Requests'

## Search on 'Requests' in Community

- Many of the documents in the project's records are live and are being revised often and hence remain in 'draft' state, as 'requests'.
- While InvenioRDM has an extensive search capability for published records, the search for records as 'requests' is limited to their titles.
- Because of the larger proportion of drafts in the repository, there was a need to extend the search capability of 'requests' in a community. This includes full or partial search of all metadata fields of records under review.
- Enabling the ability to search on 'requests' is a crucial addition to assist in their easier retrieval within a community.

# SDCC - Access Control & Customization

- **Access Control**
  - sPHENIX is a repository where authentication is required to access any data in the repository and the default open access in InvenioRDM has been made more restrictive, to support this.
  - All the UI and REST API access to the application have been made to require authentication.
- **Customization**
  - Users' information have been added to utilize the autocomplete feature for contributors.
  - More data types have also been added as required by the users.

# SDCC - Deployment

## Deployment

The InvenioRDM based application has been deployed on a test and a production server on RHEL 8. A puppetized script is in development to **automate** all the steps involved in deploying this application from scratch.

- Installation of prerequisite packages
- Adding configuration for all services (postgres, redis, opensearch, rabbitmq) to run them natively on the system
- Configuration for uwsgi, nginx and authentication systems used
- Installation of InvenioRDM base code and the custom code changes for the application
- Additional custom configuration and starting services

# SDCC - DOI Extension

## DOI Extension

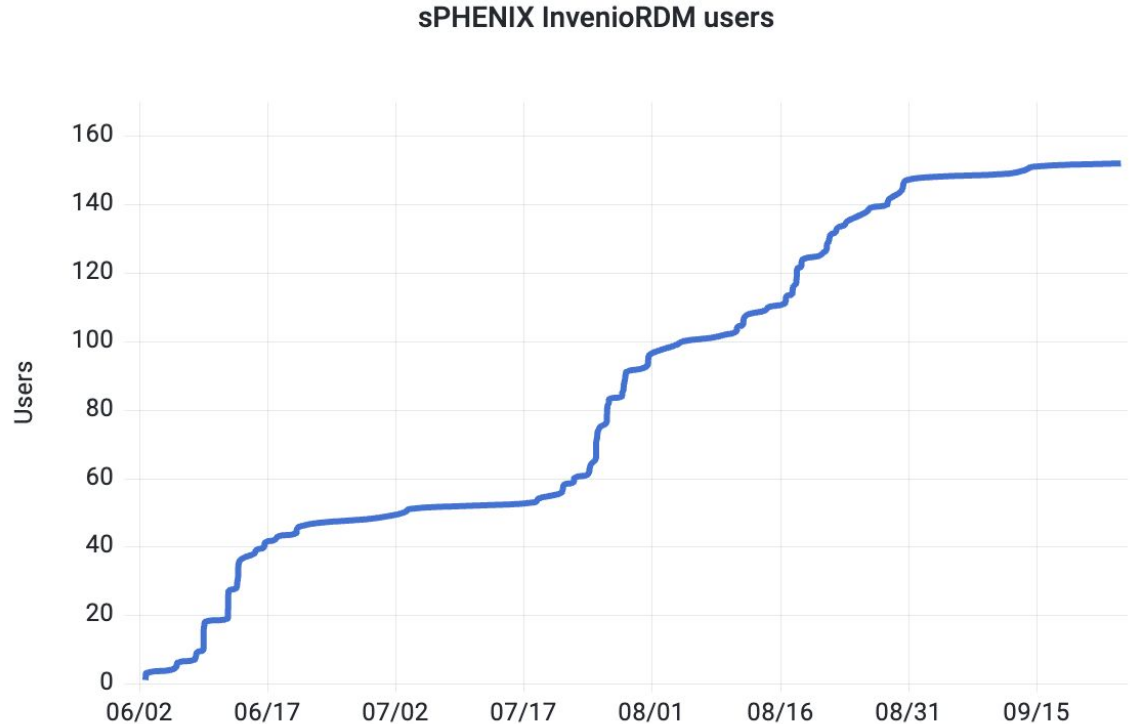
- InvenioRDM by default registers DOIs through Datacite
- All DOE documents are required to have DOIs from OSTI (Office of Scientific and Technical Information)
- OSTI DOIs are free and OSTI allows DOIs to be registered for datasets
- The extension of InvenioRDM to support OSTI DOIs has been tested locally and integration with InvenioRDM code base is being worked on

## Summary of Features added to InvenioRDM for sPHENIX Repository

1. Login using SDCC account and a layer of authorization to allow only sPHENIX group members to login to the repository
2. All sPHENIX users who login are made members of the sPHENIX Community in which all the data is expected to reside in the repository
3. The conversation associated with each record under review is made to have nested replies for comments, and the ability to modify/delete them
4. A new access level to allow members to take part in records' review conversations but not to have privileges for more managerial actions
5. InvenioRDM having an open access model, by default allows certain pages to be accessed without login. All UI and REST API access are made to require authentication for this project
6. Addition of sPHENIX usernames and other data to be populated for selection of Authors and additional values for metadata for records (Resource Types)
7. There are a few paths for creating a record and uploading data. Limiting this to a single path, to ensure all uploaded content will be added for review by sPHENIX users
8. InvenioRDM provides extensive search capability for records which are finalized and published, but not for those which are being revised. Most of the records for sPHENIX are drafts, being revised constantly. The search capability has been extended to drafts
9. Configuration of the application and scripts to automate installation of source code that have been added to InvenioRDM, to provide the above mentioned features

# sPHENIX user growth since release

- As of 09/25/2023 (4 months since release) there are 152 active users utilizing the sPHENIX InvenioRDM service at the SDCC



# Conclusions

- InvenioRDM provides an optimal solution to collect research and have work cited
- It is easily scalable with SDCC mounted storage, maintenance, and daily backups provide users with a secure pathway for progress
- There is active development for future updates provided by InvenioRDM community members
- We currently see a demand for the implementations discussed today and a crucial need for a “one stop shop” for scientific collaboration, publication, and dissemination
- While the community grows so will the platform and, in addition, the needs of the user. Actively engaging with the scientific community and meeting these demands ensures collaboration needs are met