

INVENIO RDM

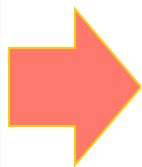
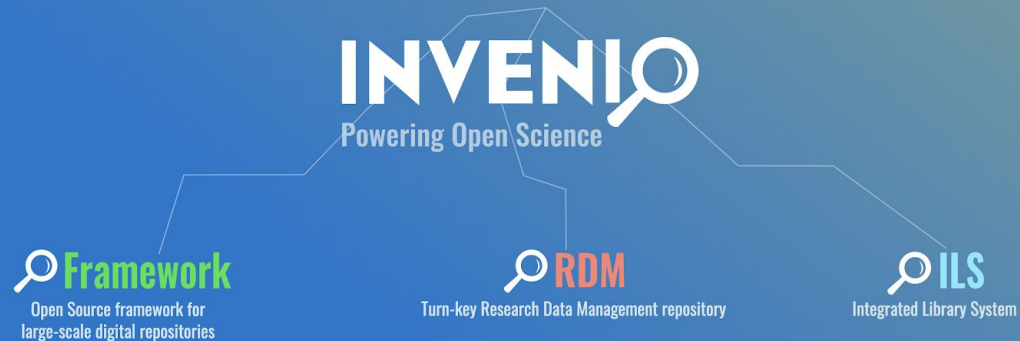
Powering open science and collaboration with Invenio

EIC Computing Forum

Northwestern University Invenio Team

November 15, 2021

Invenio software powers open science



Free Open Source Software

Invenio is Free Open Source Software supported by a committed community of multidisciplinary institutions.

[Code](#) | [Docs](#) | [Examples](#)



Friendly and Responsive Community

Although Invenio was born at CERN, its community is growing bigger every day. Talk to the team now in our chatroom or forum.

[Chatroom](#) | [Forum](#) | [Get Involved](#) | [Events](#)



The "Safe bet"

Invenio community has been around for 20 years. **Solid services** have been built on top of it to ensure long-term confidence.

[Live services](#) | [Products](#)

How did this collaboration start (and what about Zenodo?!)

What motivated the InvenioRDM project?

- Some organizations tried to reuse the existing open source Zenodo source code
- Other orgs tried to use the Invenio Framework to build a RDM repository from scratch
- Several orgs tried to make the same modifications but had no easy way of sharing their changes

All these groups came together to create a collaborative open source project and grow a sustainable community.

The Zenodo logo is displayed in white lowercase letters on a solid blue rectangular background. The font is a clean, sans-serif typeface.

Zenodo will also run on InvenioRDM by the end of the project period.

We're leveraging Invenio as a strong foundation. Here's why.

RDM platforms
are critical to help preserve and share research, enable reproducibility, and empower reuse of datasets, protocols, engagement or study materials, & a wide range of other research products.

- **Research, shared.** Securely share and preserve data records and a wide range of research types with collaborators. Allows easy dissemination to the community.
- **Discoverable.** Leverages metadata standards and the powerful Elasticsearch full-text search engine retrieves, facets, sorts, and filters your searches with ease.
- **Scalable.** Invenio is fast. Designed to manage 100+ million records and petabytes of files. All data can be archived independently of the size.
- **Communities.** Create and curate your own community (e.g., workshop, project, lab, or journal).
- A robust community: Large team of developers & active open source community. A SAAS-model for service via TIND (CERN spinoff). Invenio is widely used by [many organizations](#) & underlying technology (Python, Flask) widely supported.
- **Next-Generation:** With InvenioRDM, any organization can launch a turn-key open source next-generation repository platform with world-class features to support open and FAIR science. <http://ngr.coar-repositories.org/>
- **Get credit & be cited.** Get a DOI to make records easily and uniquely citable. Pre-formatted citation text makes it easy to cite your work and be cited. Contributor roles allow you to recognize the whole team.
- **Metrics.** Industry standard usage statistics for record pages with all tracking completely anonymized.
- **FAIR.** Advanced features to make your research Findable, Accessible, Interoperable, & Reusable.
- **Compliance-friendly.** Comply with data sharing mandates* and acknowledge your funders.
- **Easy.** Turn-key research data management platform & index can be easily deployed in the local environment by your team or by a service provider, such as TIND. Customize the look and feel to your local environment.

The InvenioRDM project has two goals:



Repository Platform

Build a turn-key research data management (RDM) repository platform based on [Invenio Framework](#) and [Zenodo](#).

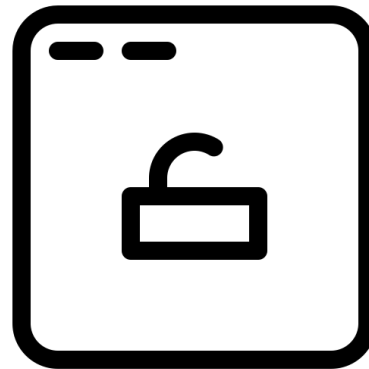


Community

Grow a community of research institutions, private companies and individuals to sustain the platform going forward.

The platform

A few highlights...

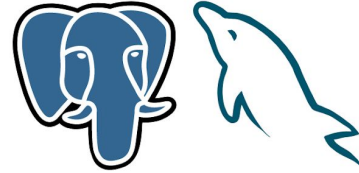


InvenioRDM stack



Elasticsearch

Elasticsearch is an extremely fast JSON-native distributed search engine supporting anything from full-text to geospatial queries.



PostgreSQL or MySQL

PostgreSQL and **MySQL** are powerful relational databases with JSON-support as well as a strong reputation for reliability, robustness, and performance.



Python/Flask

Invenio is built using **Python 3**, the **Flask** micro web framework and a suite of the best community-built Python libraries.

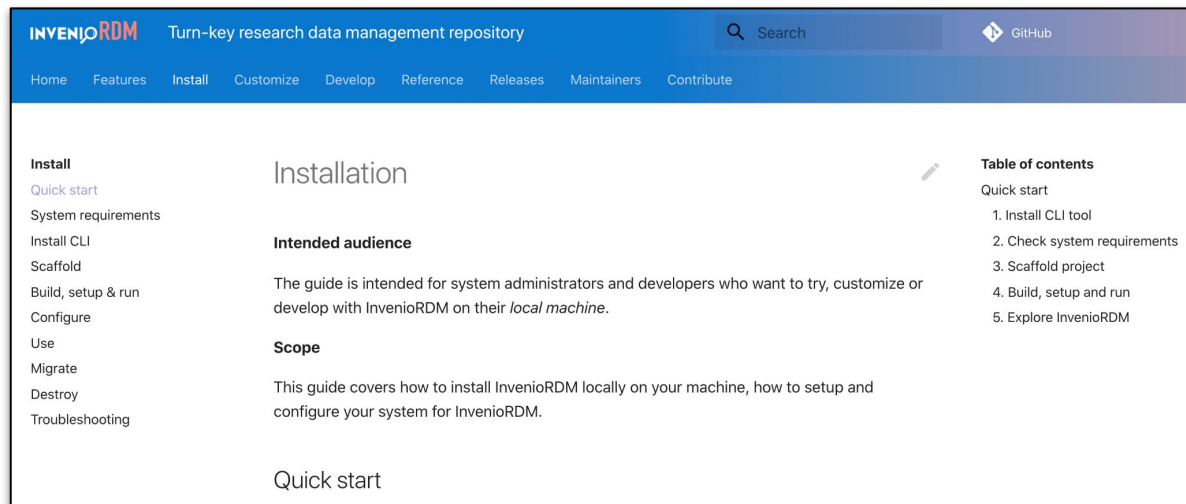


React

InvenioliS UIs are built using **React**, the well-known JavaScript library.

Invenio is JSON-native and provides RESTful APIs to make it easy to build apps on top of the framework

Standing up InvenioRDM



The screenshot shows the InvenioRDM documentation website. The header includes the InvenioRDM logo, the tagline "Turn-key research data management repository", a search bar, and a GitHub link. The navigation menu contains links for Home, Features, Install, Customize, Develop, Reference, Releases, Maintainers, and Contribute. The main content area is titled "Installation" and includes a sidebar with a table of contents. The table of contents lists: Quick start, System requirements, Install CLI, Scaffold, Build, setup & run, Configure, Use, Migrate, Destroy, and Troubleshooting. The main text under "Installation" includes sections for "Intended audience" (system administrators and developers) and "Scope" (local machine installation).

- 1- Install invenio-cli
`pip install invenio-cli`
- 2- Initialize your project
`invenio-cli init`
`cd <project name>`
- 3- Install it and run it
`invenio-cli install`
`invenio-cli services setup`
`invenio-cli run`
- 4- Visit `https://localhost`
`firefox https://127.0.0.1:5000`

System requirements

Invenio can run in Docker, on virtual machines, or on physical machines. Invenio can run on a single machine or a cluster of 100s of machines.

It all depends on exactly how much data you are handling and your performance requirements.

Small installation:

- Web/app/background servers and Redis: 1 node
- Database: 1 node
- Elasticsearch: 1 node

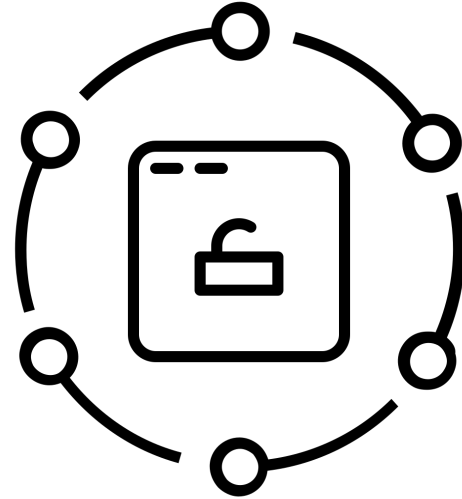
Medium installation:

- Load balancer: 1 node
- Web/app servers and background workers: 2 nodes
- Database: 1 node
- Elasticsearch: 3 nodes
- Redis/RabbitMQ: 1 node

Large installation:

- Load balancer: 2 node (with DNS load balancing)
- Web/app servers: 3+ nodes
- Background workers: 3+ nodes
- Database: 2 nodes (master/slave)
- Elasticsearch: 5 nodes (3 data, 2 clients)
- Redis: 3 nodes (HA setup)
- RabbitMQ: 2 nodes (HA setup)

The community





The turn-key research data management repository

 Aug 5, 2021: InvenioRDM v6.0 Long-Term Support Released



Status

Follow the latest project status.



Roadmap

See our next major milestones



Demo

Demo of InvenioRDM showing latest development state.



Forum

Join our project forum and collaborate.



Chat

Find all the partners in our official chatroom.



Code

Have a look at InvenioRDM code evolution.

InvenioRDM collaborators



data futures



NORTHWESTERN UNIVERSITY



Search and retrieve datasets using standards-based documentation

Robust search enhanced by:

- Standardized forms of name (LDAP + ORCID coming soon)
- Standard subject terms (MeSH, Library of Congress Subject terms)
- Standardized citation formats
- Clear levels of access
- Standard application of licenses

The screenshot displays the InvenioRDM search interface at Northwestern University. The search term 'diabetes' has yielded two results. The first result is the 'Hyperglycemia and Adverse Pregnancy Study (HAPO) dataset', which is available for 'Open Access'. The second result is the 'Methylomics of Prenatal Gestational Diabetes Mellitus (GDM) dataset', also available for 'Open Access'. A search filter panel is overlaid on the right, showing 'Subjects' with two categories: 'Medical' (MeSH terms) and 'Topical' (Library of Congress (FAST) terms). Both categories have a '+ Add another' button.

InvenioRDM @ Northwestern University
Catalog your Research

Search

diabetes

found 2 results.

Hyperglycemia and Adverse Pregnancy Study (HAPO) dataset Open Access

Metzger, Boyd; Lowe, Lynn P.; Dyer, Alan R.

HAPO aim to assess the association between maternal glucose levels and the adverse outcome of mothers diagnosed with gestational diabetes mellitus (GDM). 25,000 pregnant women from 15 clinics across 9 countries were enrolled with 21,064 able to complete and participated in the study. Women's course of care throughout pregnancy, glucose tolerance, A1C levels and their outcomes were analyzed for any associations or independent associations to create updated guidelines for diagnosing Gestational Diabetes Mellitus (GDM) based off the data collected and adopted by WHO.

Methylomics of Prenatal Gestational Diabetes Mellitus (GDM) dataset Open Access

Hou, Lifang; Hu, Gang; Bacarelli, Andrea; Liu, Lei; Zhang, Wei

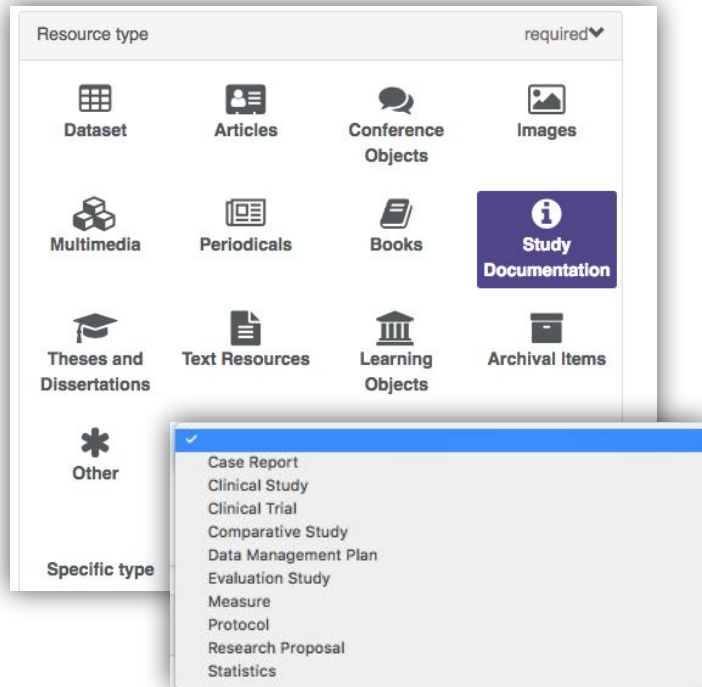
Building off of The Tianjin GDM Observational Study in China, this project brings the addition of genetics and lifestyle intervention to make a longitudinal study. Blood samples from both the mother and children are taken to be analyzed for DNA methylation and phenotypes. This analysis investigates the effects GDM and the risks for obesity and cardiometabolic issues that may arise in children as a result of in utero exposure. This project aims to explore the influence of lifestyle choices and intervention with children affected.

Subjects optional

Medical Select Medical Subject Heading (MeSH) terms X
Medical Subject Heading (MeSH) terms.
[+ Add another MeSH term](#)

Topical Select Library of Congress (FAST) topical terms X
Library of Congress (FAST) topical terms.
[+ Add another FAST term](#)

Data management for reproducibility and Open Access: study-focused resource types



InvenioRDM helps you store, manage and, if needed, share your study's outputs:

- Study-based resource types to manage a large range of assets
- **Reproducibility** is enhanced: store research proposals, datasets, code
- Be **compliant** with data sharing mandates
- **Cite** and **attribute** the work of all contributors to research
- **Reuse** deposited data or measures from other studies

Communities & Collections

Community: Define your research group or other collaborative unit

Collection: Create multiple Collections under the umbrella of the Community. Within Collections, deposit and describe your:

Phenotype Definitions

Definitions
Characterizations
Evaluations
Metadata
Dissemination Strategy

Clinical Studies

Research Proposals
Protocols
Data Management Plans
Methods Descriptions
Measures
Case Reports
Datasets and Analyses

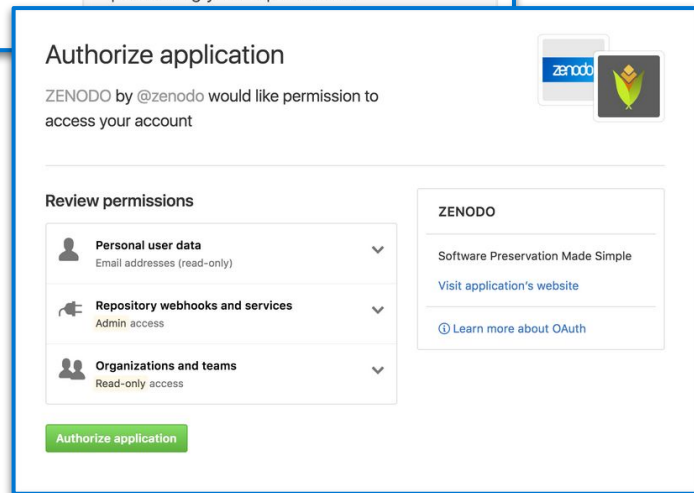
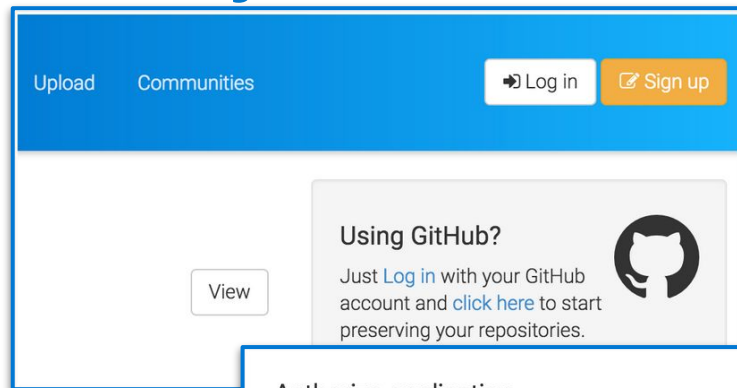
Collections bring together related groupings of documentation to communicate process, enable sharing of results, and support publication, compliance, and reproducibility

The screenshot displays a user interface for research data management. At the top, there are three main sections: "Research. Shared.", "Safe", and "Citeable. Discoverable.", each with a corresponding icon and text. Below this, the "Popular communities" section features a "view all" link and a card for the "Biodiversity Literature Repository" with its logo. To the right, two collection cards are shown: "Phenotype Definitions" with a grid of eyes and "XYZ Clinical Study" with a microscope icon. Below these are sections for "Popular records" (with a "This week" filter), "New records", and "Blog posts" (with a link to "Zenodo and General Data Protection Regulation"). A yellow arrow points from the "Phenotype Definitions" collection to the "Phenotype Definitions" section header in the adjacent text block. The footer includes a "Funded by" section with three logos and a navigation bar with links for "Status", "Privacy policy", "Terms of use", and "Support".

Software publication and citeability

InvenioRDM will support software publication or deposit in a number of ways, including:

- Generation of citations for all resource types, including software
- Planned support for custom fields which may be necessary to integrate software metadata
- 3rd party integrations allowing a direct link to GitHub, as can currently be seen in Zenodo
 - Login to GitHub from Zenodo
 - Authorize GitHub to give Zenodo needed permissions
 - Pick the repository you want to archive
 - Create a new release



Software publication example

August 23, 2016

Software Open Access

ChIPseqRUs: A pipeline for ChIP-seq preprocessing

Loh, Yong-Hwee; Shao, Ning-Yi; Shen, Li

A preprocessing pipeline for ChIP-seq, including alignment, quality control, and visualization.

Preview

chip-seq_preprocess-v1.0.zip

shenlab-sinai-chip-seq_preprocess-42aae7e

○ .gitignore	311 Bytes
○ LICENSE	18.0 kB
○ README.md	4.7 kB
○ all_flowchart.png	94.5 kB
○ bin	
▪ SourceMe_minverva_profile	230 Bytes
▪ fastq2bam_by_bowtie.sh	1.6 kB
▪ fastq2bam_by_bowtie2.sh	1.4 kB
▪ genNormedNgsplotConfig.py	1.0 kB
▪ genTDF	144 Bytes
▪ ngsplot_all.sh	1.2 kB
▪ rmdup.bam.sh	490 Bytes
▪ rmdup_PE.bam.sh	487 Bytes
▪ runPhantomPeak.sh	267 Bytes
▪ run_spp_nodups.R	37.9 kB
○ project	

Files (131.6 kB)

Name

Size

157

views

7

downloads

See more details...

Available in

GitHub

Indexed in

OpenAIRE

Publication date:

August 23, 2016

DOI:

DOI: [10.5281/zenodo.60674](https://doi.org/10.5281/zenodo.60674)

Keyword(s):

chip-seq quality control pipeline workflow
visualization alignment preprocessing parallelization

Published in:

GitHub repository: (2016).

Software publication example

September 3, 2020

Lesson **Open Access**

Building Websites with Jekyll and Github/Gitlab: Pre-Alpha test release

Toby Hodges; Aleksandra Nenadic; Julian Karl Bauer; Anne Foullioux; Sarah Stevens; Renato Alves

A (very) pre-alpha release of a lesson teaching the skills required to design, author, and publish web pages with Jekyll and GitHub. The material is nowhere near complete: this release was made to test out the Zenodo/GitHub integration.

Preview

[building-websites-with-jekyll-and-github-or-gitlab-v0.1-pre.alpha.zip](#)

- carpentries-incubator-building-websites-with-jekyll-and-github-or-gitlab-c781181
 - .editorconfig 364 Bytes
 - .github
 - FUNDING.yml 109 Bytes
 - ISSUE_TEMPLATE.md 944 Bytes
 - PULL_REQUEST_TEMPLATE.md 847 Bytes
 - .gitignore 169 Bytes
 - .gitlab-ci.yml 376 Bytes
 - .travis.yml 1.0 kB
 - AUTHORS 33 Bytes
 - CITATION 22 Bytes
 - CODE_OF_CONDUCT.md 330 Bytes
 - CONTRIBUTING.md 6.8 kB
 - Gemfile 244 Bytes
 - LICENSE.md 3.3 kB
 - Makefile 4.8 kB
 - README.md 451 Bytes

Files (67.7 kB)

38

views

0

downloads

[See more details...](#)

Available in

GitHub

Indexed in

OpenAIRE

Publication date:

September 3, 2020

DOI:

DOI [10.5281/zenodo.4013386](https://doi.org/10.5281/zenodo.4013386)

Keyword(s):

education jekyll github gitlab carpentries-incubator

Related identifiers:

Supplement to

<https://github.com/carpentries-incubator/building-websites-with-jekyll-and-github-or-gitlab/tree/v0.1>

<https://doi.org/10.5281/zenodo.4013386>

With thanks...

Teams

- The Invenio team @ CERN & RDM collaborators ([here](#))
- Galter Health Sciences Library & Learning Center
- Northwestern University Clinical and Translational Sciences Institute
- The NU Institute for Innovations in Developmental Sciences
- Confederation of OA Repositories (COAR)

Support

Work presented here is supported in part by:

CERN Knowledge Transfer Fund

NUCATS: UL1TR001422 (NCATS)

★ all of the InvenioRDM project partners



Sara Gonzales



Guillaume Viger



Matt Carson



Kristi Holmes



Northwestern University
NUCATS
Clinical and Translational Sciences Institute