InvenioRDM
A Collaborative Next-Generation Research Data Management and Repository Solution

Thomas Morrell @temorrell
Research Data Specialist, Library
California Institute of Technology

Kristi Holmes @kristiholmes
Director, Galter Health Sciences Library & Learning Center
Associate Professor of Preventive Medicine-Health and Biomedical Informatics
Northwestern University Feinberg School of Medicine

@inveniosoftware
Invenio software powers open science

**Framework**
Open Source framework for large-scale digital repositories

**RDM**
Turn-key Research Data Management repository

**ILS**
Integrated Library System

**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
Institutional Repository and Data Index & catalog
https://inveniosoftware.org/products/rdm/

The turn-key research data management repository

Launching in the summer 2020

Institutional Repository and Data Index & catalog
https://inveniosoftware.org/products/rdm/

Roadmap
We intend to be ready by summer 2020.

Talk
Join our project forum and collaborate.

Chat
Find all the partners in our official chatroom.

Code
Have a look at InvenioRDM code evolution.

Events
InvenioRDM project events for partners

Features
Sneak peak at the future InvenioRDM.
• We’re going install Invenio RDM during this presentation
• This is an Alpha release (limited features, temporary data model, no UI design)
• Follow along at github.com/inveniosoftware/invenio-app-rdm/wiki
• Let’s set up the repository, build docker images, and configure database
How does Zenodo fit into all of this?

- Zenodo is a great model of a modern research data repository
- Even though it’s open source, Zenodo was not designed for non-CERN deployment
- Zenodo is a repository service hosted by CERN, InvenioRDM is a repository application

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.

Zenodo will also run on InvenioRDM by the end of the project period.
InvenioRDM Features (Research Data Management)

✔ Research, shared. Securely share and preserve data records and research materials with collaborators. Allows easy dissemination to the community.

✔ 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.

✔ Discoverable. Leverages metadata standards and the powerful Elasticsearch full-text search engine retrieves, facets, sorts, and filters your searches with ease.

✔ Metrics. Industry standard usage statistics for record pages with all tracking completely anonymized.

✔ FAIR. Advanced features to make your research Findable, Accessible, Interoperable, & Reusable.

✔ Communities. Create and curate your own community (e.g., workshop, project, lab, or journal).

✔ Comply. Comply with data sharing mandates* and acknowledge your grants.

✔ 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.


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.
Why this, why now? Examine current repositories

- Dated technology, nearly impossible to advance
- Minimal interoperability
- Requirements for modern research not supported
- No end-to-end support for FAIR
- No end-to-end support for NGR
- Silos, anyone?
- Rarely coordinated vision and direction
- They’re really hard to implement and maintain
- Closed (truly or effectively) hamper access, innovation, translation, impact, accountability, …
Why this, why now? Examine InvenioRDM benefits

**Top of the class UX**
Finally a repository that your users will love! We have thought hard about user experience for everyone: end-users, curators, system administrators, and developers. Expect more of your repository!

**Institutional integration**
InvenioRDM integrates with your institution. It supports your Single Sign-On (SSO) solution, your storage systems, your public/private clouds, your permissions models as well as your custom fields.

**Repository Profiles**
InvenioRDM comes with pre-configured repository profiles for institutional repositories (IRs), research data management (RDM) systems as well as domain-specific repositories for health and biomedical sciences.

**Scalability**
InvenioRDM scales with you! Handle 1 or 100 million records, 1 byte or several petabytes. It runs on bare-metal, VMs and container platforms such as Kubernetes and OpenShift.

**Next Generation Repository**
InvenioRDM implements the Next Generation Repository (NGR) vision and integrates with other Open Science infrastructures such as ORCID, DataCite & OpenAIRE.

**Resilient**
Architected from ground up to be resilient to changing business requirements. Stable APIs, easy maintenance and upgrade, as well as an extensible platform to ensure a reliable core on which you can build new innovative features and integrations.
The InvenioRDM project is an open source collaboration with two main 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.
Building the community
The InvenioRDM collaboration

The turn-key research data management repository

Launching in the summer 2020

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 project is scheduled to deliver the first release of InvenioRDM by summer 2020

FAQ
How do we support open, collaborative work on geographically dispersed teams?

InvenioRDM resources: open source via GitHub
- Repository: [github.com/inveniosoftware/invenio-app-rdm](https://github.com/inveniosoftware/invenio-app-rdm)

Invenio RFC (Request For Comments):
- Coordinate the design process
- Produce consensus among Invenio stakeholders
- Document design decisions

RFCs aid communication between geographically dispersed teams and document Invenio development so that we can avoid knowledge loss when people leave and ease knowledge transfer when people join.

[github.com/inveniosoftware/rfcs](https://github.com/inveniosoftware/rfcs)

Project Boards
- Coordinate development during the monthly sprint
- Document responsibility and priorities
- Identify issues and context in the monthly sprint: to-do, in-progress, blocked, in review, and done

Work in progress for InvenioRDM this month:
How do we support open, collaborative work on geographically dispersed teams?
How do we support open, collaborative work on geographically dispersed teams?

**Discourse Forum**

- [invenio-talk.web.cern.ch](https://invenio-talk.web.cern.ch)

**Gitter (incl. daily development updates):**

- [gitter.im/inveniosoftware/InvenioRDM](https://gitter.im/inveniosoftware/InvenioRDM)
How do we support open, collaborative work on geographically dispersed teams?

Invenio RDM Roadmap:
inveniosoftware.org/products/rdm/roadmap/
Ongoing support for the community

Events:

Webinars coming soon
We will soon be providing webinars on Invenio products.

Training bootcamp
We regularly hold bootcamps to show the nuts and bolts of Invenio. This year’s materials are available at [indico.cern.ch/event/773969](indico.cern.ch/event/773969)

We also have a nice set of tutorials:
[github.com/inveniosoftware/training](github.com/inveniosoftware/training)

User Group Workshop
We run an Invenio User Group Workshop every year. See the full list at [indico.cern.ch/category/6240](indico.cern.ch/category/6240)
Collaborative work, globally: Next Generation Repository (NGR) behaviors & technologies help guide our journey

Behavior user stories:
“As a machine or human user, I need to easily and uniformly identify the licensing and re-use conditions of a scholarly resource, so that I know what I am allowed to do with it.”
InvenioRDM as a foundation for local engagement


Search public content now or log in to try a deposit:
Test Login: gla3975
Password: InvenioRDM@NU_2019
Live Demo

• We’ll populate some demo records
• Then we’ll start up the server
Institutional Repository for Data and Software
Zenodo clone, launched in 2017
Hosted service from TIND
Easy to describe and upload files
Library takes care of preserving and maintaining access to files

data.caltech.edu
Institutional Repository as Infrastructure

tccon.ornl.gov

Private data files
- Ascension Island
  - asc20120722_20120813.txt
  - asc20120317_20120818.txt
  - asc20120911_20121219.txt
  - asc20130408_20130416.txt
  - asc20130121_20131015.txt
  - asc20130101_20130330.txt
  - asc20150311_20150429.txt
  - asc20150410_20150430.txt
  - asc20150501_20150526.txt
  - asc20151006_20151328.txt

Departmental Server at Caltech

tccondata.org

Public data files
- Index of /2014Public/ascension01
  - Name: README.txt
  - Size: 11.8KB
  - Date Modified: 10/2014, 5:06:00 PM
  - Name: asc20120722_20120813.public.txt
  - Size: 10.2MB
  - Date Modified: 5/2017, 4:28:00 PM

Migration

CaltechDATA

https://doi.org/10.14291/tccon.ggr2014.parkfalls01.R1
API-based innovation

Departmental Server at Caltech

California Institute of Technology
Research Data Repository

Data Files
Metadata

Update metadata
Dates
Version Information
Usage Information
Update files
Netcdf data
README

caltechdata_api

Monthly automatic updates, new data version or new site

https://github.com/caltechlibrary/caltechdata_api
https://github.com/inveniosoftware/datacite
API-based innovation

- Binder provides web-based access to Jupyter notebooks, RStudio, and other computing environments
- Pulls content from GitHub, Zenodo, and other repositories
- Researchers can test at mybinder.org
- Then get a badge in CaltechDATA: https://doi.org/10.22002/D1.1259
Live Demo

- A full data repository is running on my laptop
- Let’s get records using the API
- Let’s make a record using the API
- Then let’s search for is using the API
- We can look at the web interface, and customize it
With sincere thanks…

Teams
- Lars Holm-Nielsen & Jose Benito Gonzalez Lopez
- The Invenio team, & RDM collaborators (here)
- Caltech Library
- Galter Health Sciences Library & Learning Center
- TIND
- Northwestern University Clinical and Translational Sciences Institute (NUCATS)
- CTSA Program Center for Data to Health (CD2H) team
- The NU Institute for Innovations in Developmental Sciences (DevSci)
- Confederation of OA Repositories (COAR) & Kathleen Shearer

Support
*Work presented here is supported in part by:*
- CERN Knowledge Transfer Fund
- CD2H: U24TR002306 (NCATS)

⭐ all of the InvenioRDM project partners!

Come pick up your InvenioRDM stickers!