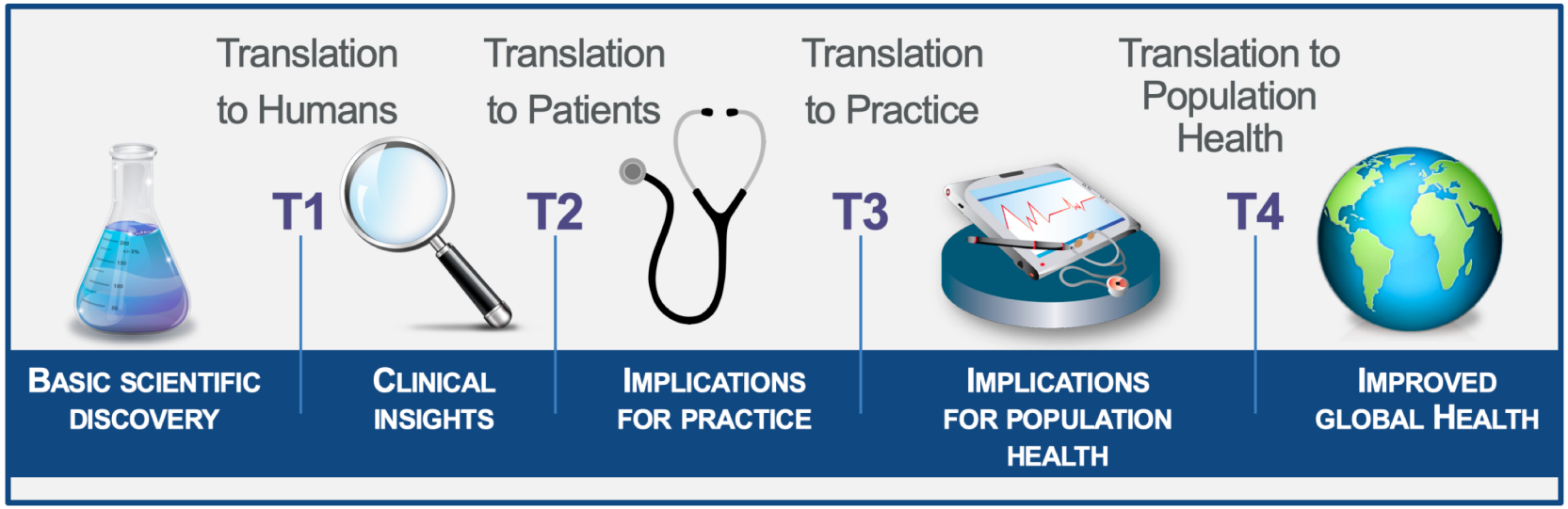
Northwestern Medicine

Using Invenio RDM to power biomedical research

Kristi Holmes
kristi.holmes@northwestern.edu
@kristiholmes
Northwestern University Feinberg School of Medicine
Galter Health Sciences Library & Learning Center
Northwestern University Clinical and Translational Sciences Institute
CTSA Program National Center for Data to Health (CD2H)

A Focus on Translational Science



The CTSA Program from the National Institutes of Health

- A national network of **>50 medical research institutions in the US** that work together to improve the translational research process to get more treatments to more patients more quickly.
- CTSA Program support enables research teams including scientists, patient advocacy organizations and community members to tackle system-wide scientific and operational problems in clinical and translational research that no one team can overcome.

CTSA Clinical & Translational Science Awards Program

The National Center For Data to Health (CD2H)

Informatics & data science coordinating center for the CTSA Program

Accelerating Informatics Innovation to Advance Translational Research

Make Data Easier to Share and Re-use

Make Tools More Accessible and Interoperable

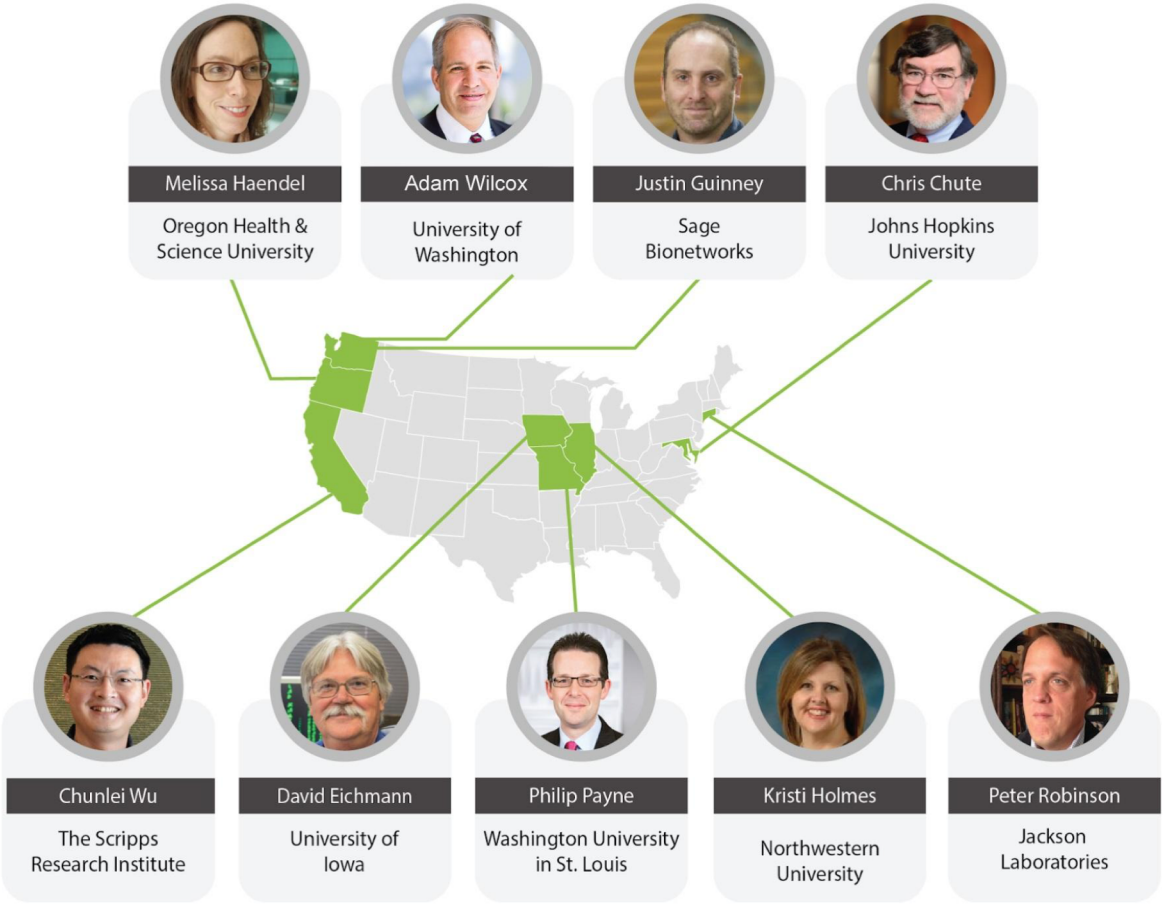
Leverage Expertise and Foster a More Collaborative CTSA Culture

Better translation of research and improved patient care

CTSA Clinical & Translational Science Awards Program

NATIONAL CENTER FOR DATA TO HEALTH

Who we are and who we serve



all >50 CTSA sites

Workforce Development | Collaboration/ Engagement | Integration Across the Lifespan | Methods/ Processes | Informatics

CLIC Center for Leading Innovation & Collaboration

…& the larger informatics community

What’s important: patient care.

https://www.cancer.northwestern.edu/types-of-cancer/brain/index.html

What's important: healthy communities and empowering access to research & partnership.

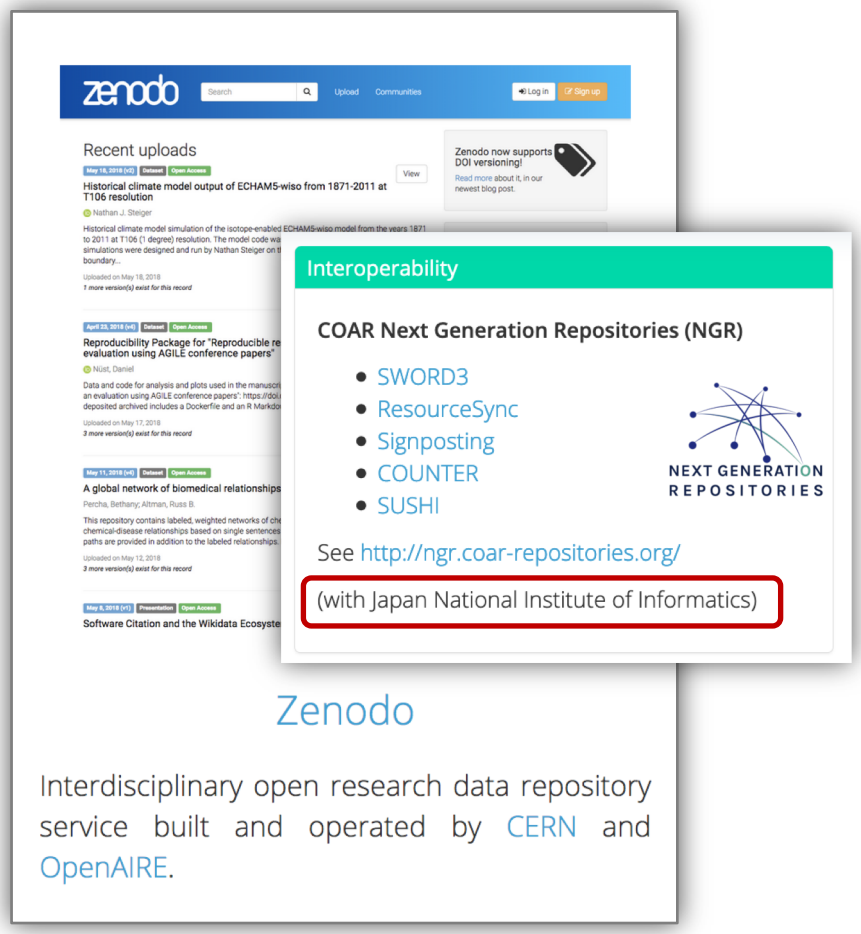
Requires

- *tri-institutional partnership and a focus on cancer health equity.*
- *collaborations with the community on cancer health equity issues.*

To deliver access to research, improve patient care, and engage our communities, it takes both technology & culture.

Repositories play a big role in making this possible.

We're leveraging Invenio v3 as a strong foundation. Here's why.



Safe: Invenio has been created with security and long-term preservation in mind.

Scalable: Invenio is fast. Designed to manage 100+ million records and petabytes of files. All data can be archived independently of the size.

RESTful: Invenio was born for the web, is JSON-native and provides RESTful APIs out of the box that will allow anyone to build on top of it.

Open: Invenio is 100% open source licensed under MIT license. Open standards for open science.

A robust community: Large team of developers & active open source community. A SAAS-model for service via TIND (CERN spinoff). 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.

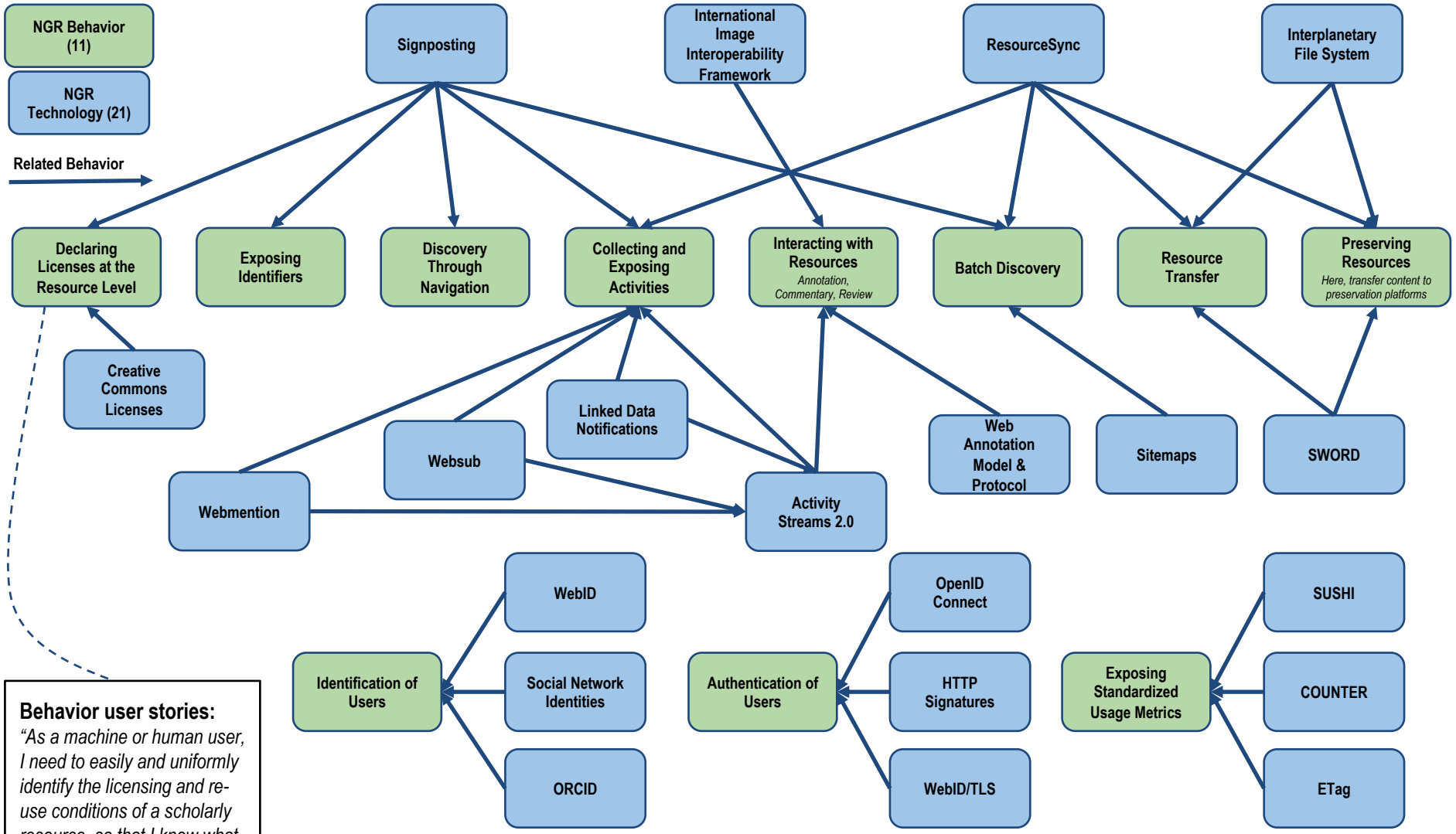
http://ngr.coar-repositories.org/

Collect, record, preserve, & disseminate a wide range of research objects from across the translational workforce (e.g., datasets, protocols, consent materials, education or engagement materials, technical reports, case studies, supplemental materials, survey instruments, lay summaries, …)

NATIONAL CENTER FOR DATA TO HEALTH

Work so far…

Mapping COAR NGR Behaviors & Technologies helps guide our own journey: where do we want to go & how to get there

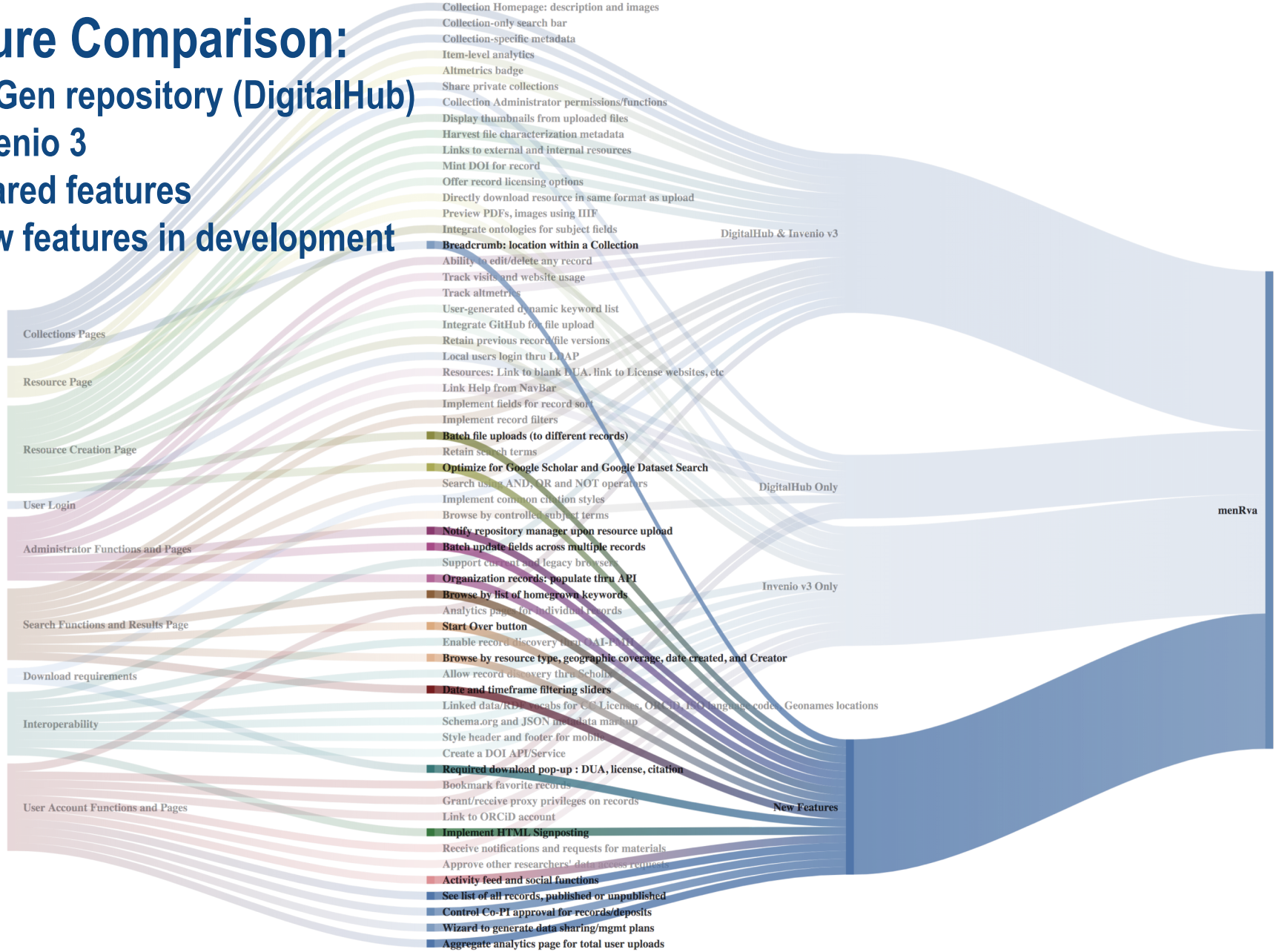


Behavior user stories:
"As a machine or human user, I need to easily and uniformly identify the licensing and re-use conditions of a scholarly resource, so that I know what I am allowed to do with it."

https://github.com/data2health/menRva

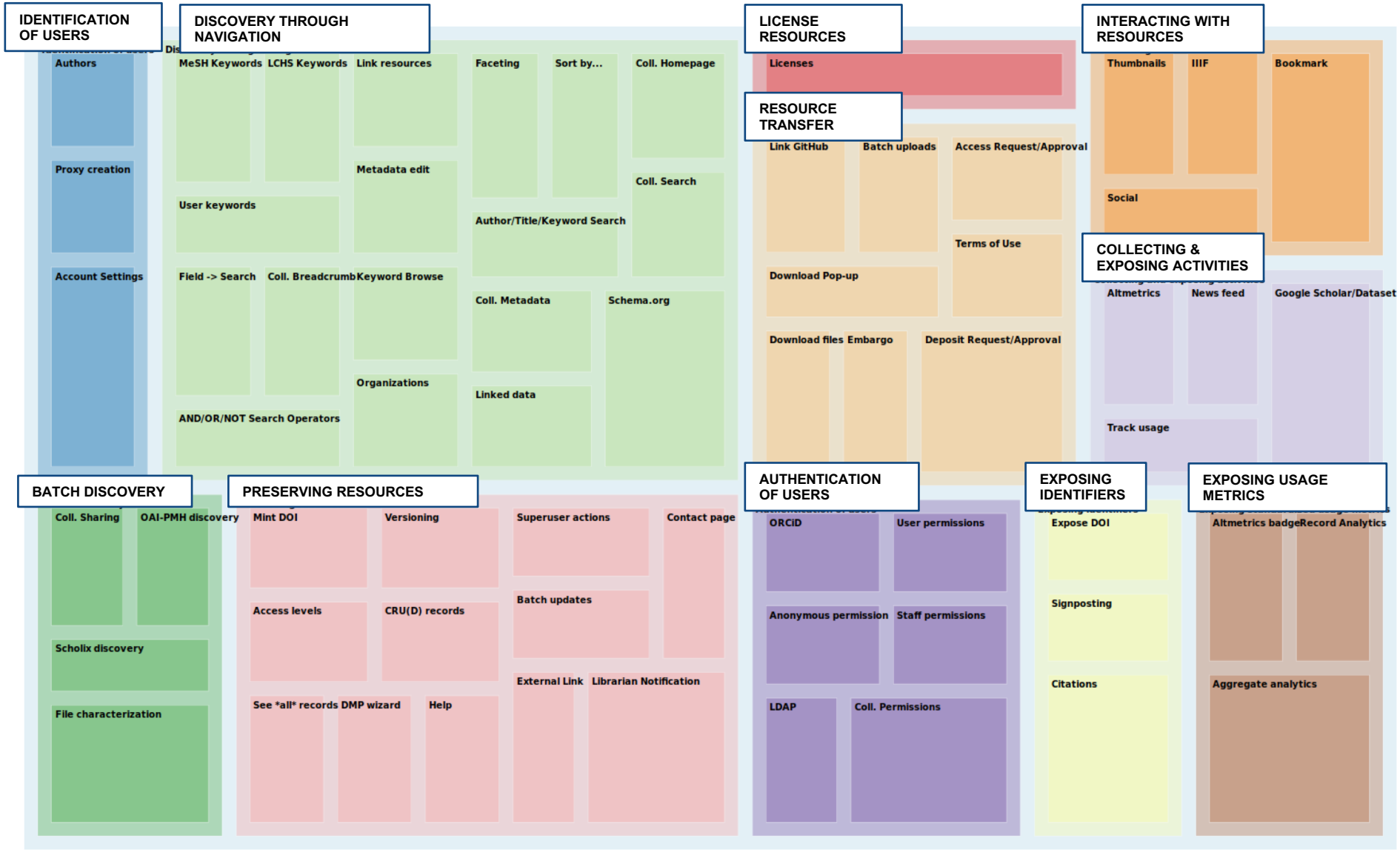
Feature Comparison:

- 1st Gen repository (DigitalHub)
- Invenio 3
- Shared features
- New features in development



Gonzales, S., Viger, G., Carson, M.B., & Holmes, K. (2019). *Beyond Accessibility: Taking the Institutional Repository to the Next Generation.* Galter Health Sciences Library & Learning Center. doi:10.18131/g3-b7wj-vh65

Our own next-generation requirements mapped against the COAR's Behaviors and Technical Recommendations for NGRs



Gonzales, S., Viger, G., Carson, M.B., & Holmes, K. (2019). *Beyond Accessibility: Taking the Institutional Repository to the Next Generation.* Galter Health Sciences Library & Learning Center. doi:10.18131/g3-b7wj-vh65

Joining Invenio RDM (1-year project)

INVENIO

Get started | Features | Examples | Documentation | Community | Roadmap | Blog

Invenio RDM: a turn-key open source research data management platform

Lars Holm Nielsen Apr 29, 2019 Invenio

CERN has partnered with 10 multidisciplinary institutions and companies to build a turn-key open source research data management platform called Invenio RDM, and grow a diverse community to sustain the platform.

The Invenio RDM project is funded by the CERN Knowledge Transfer Fund, as well as all the participating partners, including:

- Brookhaven National Laboratory (US)
- Caltech Library (US)
- Data Futures (UK)
- Helmholtz Zentrum Dresden-Rossendorf (DE)
- Northwestern University (US)
- OpenAIRE (GR)
- TIND (NO)
- Tubitak (TK)
- University of Hamburg (DE)
- University of Münster (DE)

The project has an ambitious one year schedule in which it will deliver:

- Invenio RDM - A research data management platform based on Zenodo and Invenio v3 Framework.
- A community of public and private institutions to sustain Invenio RDM.
- Minimum two existing repositories migrated to Invenio RDM, with Zenodo being one of them.

For more information about the Invenio RDM project, please contact:

Lars Holm Nielsen
Invenio Product Manager
CERN IT Department
info@inveniosoftware.org

https://invenio-software.org/blog/2019-04-29-rdm/

CD2H Update: Invenio RDM

INVENIO

Get started | Features | Examples | Documentation | Community | Roadmap | Blog

InvenioRDM: a turn-key open source research data management platform

Lars Holm Nielsen Apr 29, 2019 Invenio

CERN has partnered with 10 multidisciplinary institutions and companies to build a turn-key open source research data management platform called InvenioRDM, and grow a diverse community to sustain the platform.

The InvenioRDM project is funded by the CERN Knowledge Transfer Fund, as well as all the participating partners, including:

- Brookhaven National Laboratory (US)
- Caltech Library (US)
- CTSA Program National Center for Data to Health (US)
- Data Futures (UK)
- Helmholtz Zentrum Dresden-Rossendorf (DE)
- Northwestern University (US)
- OpenAIRE (GR)
- TIND (NO)
- Tubitak (TK)
- University of Hamburg (DE)
- University of Münster (DE)

INVENIO

https://invenio-software.org/
https://invenio-software.org/blog/2019-04-29-rdm/

What is a RDM platform?

Research Data Management (RDM) platforms play an important role in today's research ecosystem to ***disseminate & preserve***, ***enable reproducibility***, and ***empower reuse***. RDM platforms allow researchers to share and preserve scientific results and support sharing of a wide variety of resources.

What does this project include?

This project builds on the CD2H work with Invenio v3 and advances it through a collaboration with CERN and several other partner organizations.

Invenio RDM includes Zenodo features, such as DOI minting capabilities, versioning support, and COUNTER compliant usage statistics, among others.

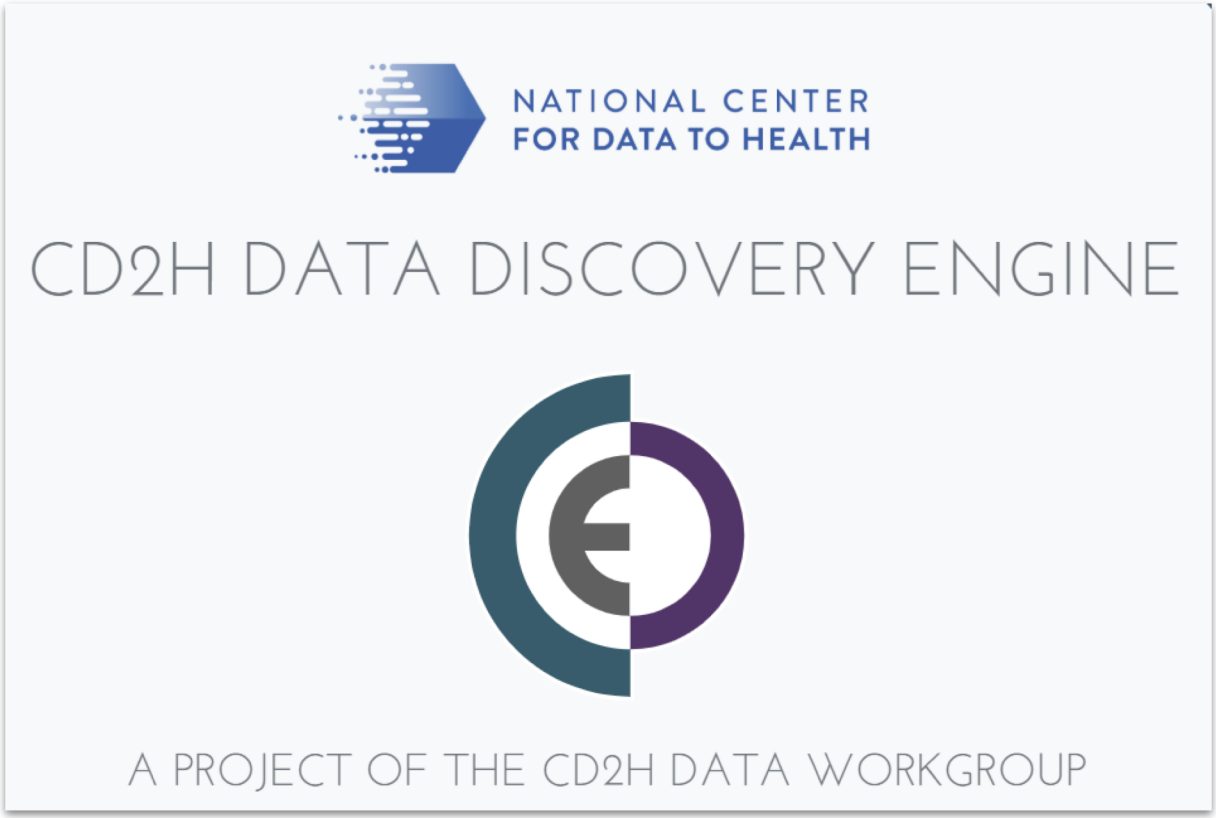
Transforming Zenodo into a general purpose RDM-platform will require work in three key areas:

1. **Core repository**
2. **Packaging and distribution**
3. **Customization and extendibility**

Next steps for our team and looking down the road...

CD2H Project Highlight: **Data Discovery Engine**

Make data more discoverable and reusable



http://discovery.biothings.io

Scripps Research

Northwestern Medicine
Feinberg School of Medicine

Moving toward a **Data Discovery Engine** for biomedicine

I have a dataset,
and I want to share it properly

So that others can find it, use it for their research, and attribute my work properly.

Now tell me what to do

I'm starting a new project,
And want to plan ahead,

So that the data coming out of my project will be FAIR from the beginning.

Now tell me what to do

Metadata Widgets
Authoring dataset metadata connect to multiple data-portals

Schema Hub
Developers' tool for hosting and extending schemas

InvenioRDM

Invenio-based data portals

centralized & standalone

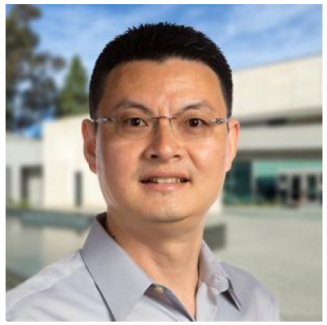
Other Tools & Services

Other Tools & Services

Other Tools & Services

DISCOVERY

Prototype at:
http://discovery.biothings.io

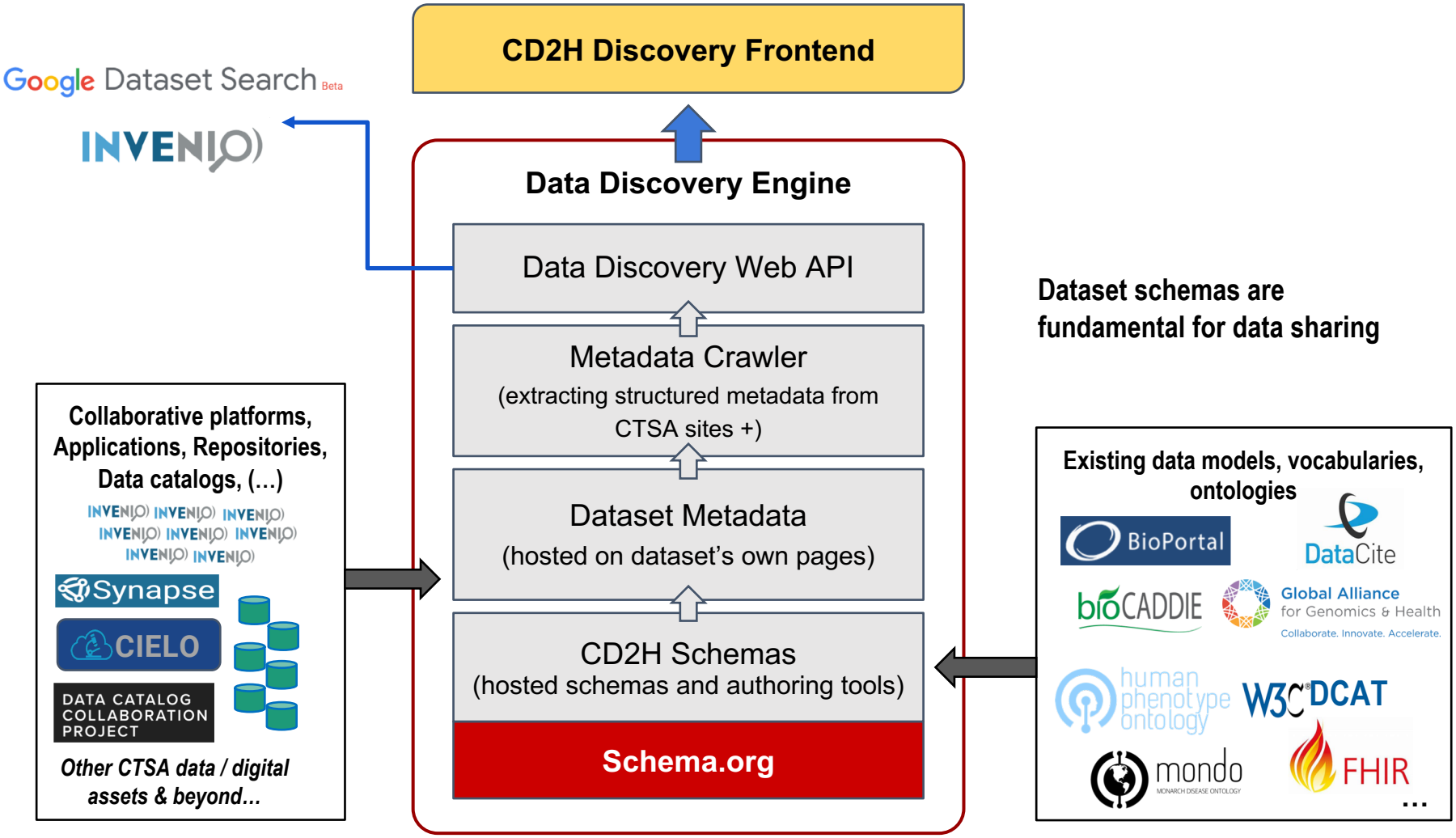


Work with Chunlei Wu (Scripps)
https://wulab.io/

Invenio-RDM *nodes and other trusted repositories can enable a distributed interoperable architecture. Empowering a range of new and existing repositories & data catalogs will result in a robust, collaborative community.*

NATIONAL CENTER FOR DATA TO HEALTH

CD2H Project Highlight: **Data Discovery Engine**

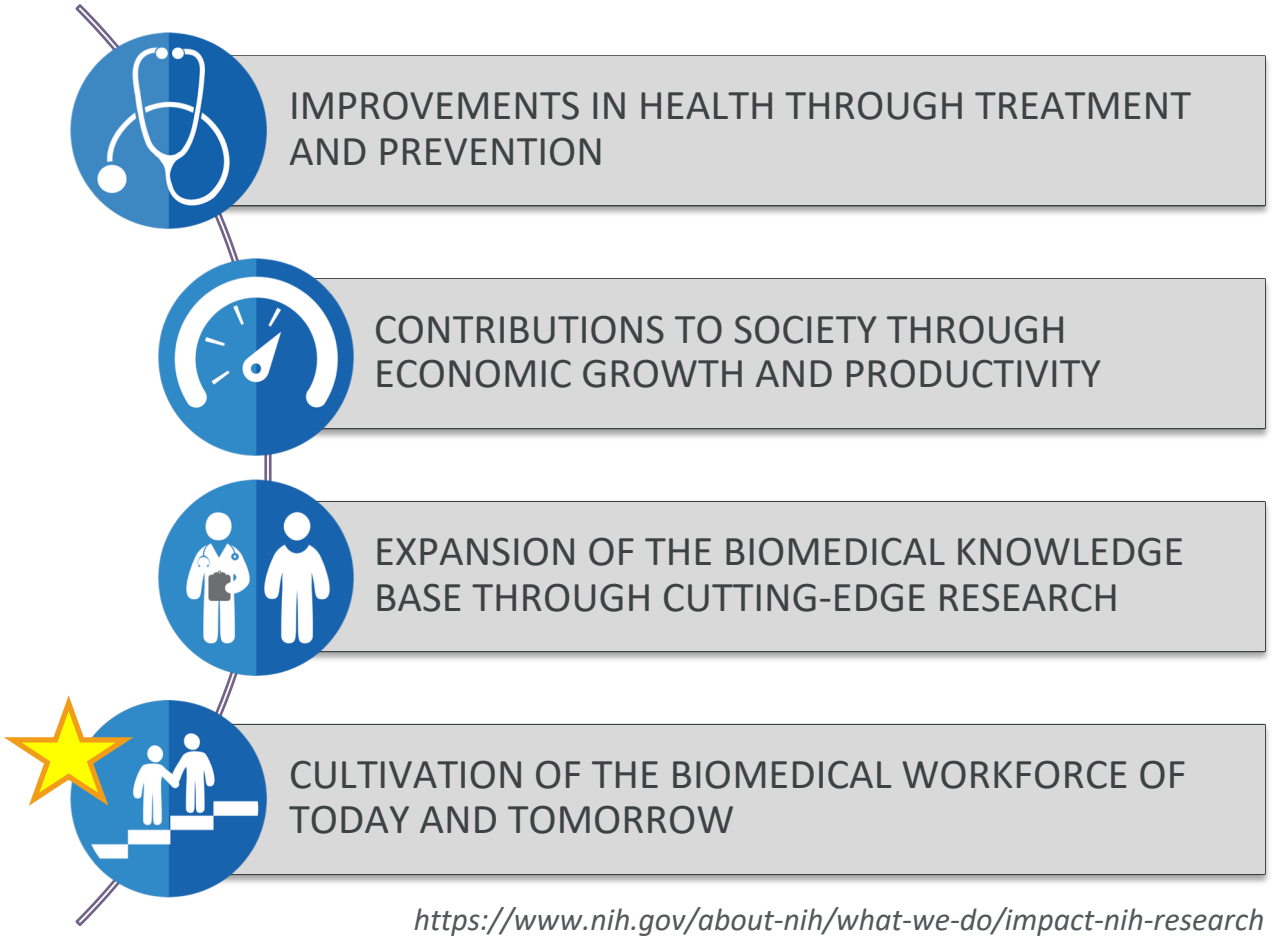


Prototype at: http://discovery.biothings.io

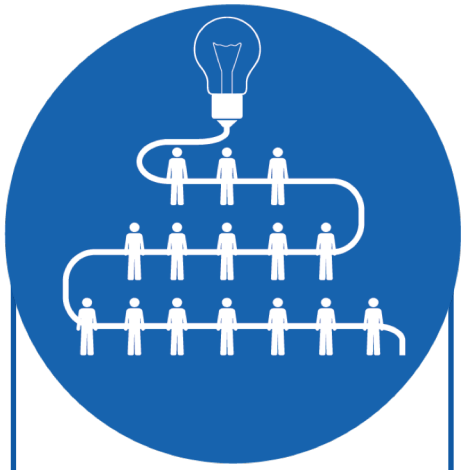
One more note on technology + culture.

What is impact?

More than papers and grants – we are driving toward improved health and wellbeing

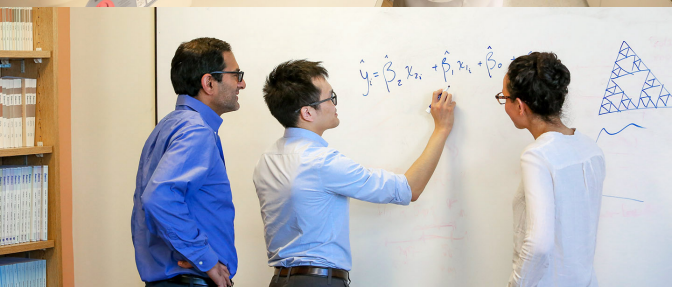


https://www.nih.gov/about-nih/what-we-do/impact-nih-research



For effective translation of knowledge and discoveries into the improved health of our communities, it is essential to incorporate evaluation strategies that enable investigators and teams to measure, monitor, and communicate the impact of their work

Biomedical research evolves: we must consider the entire research workforce and all that they do



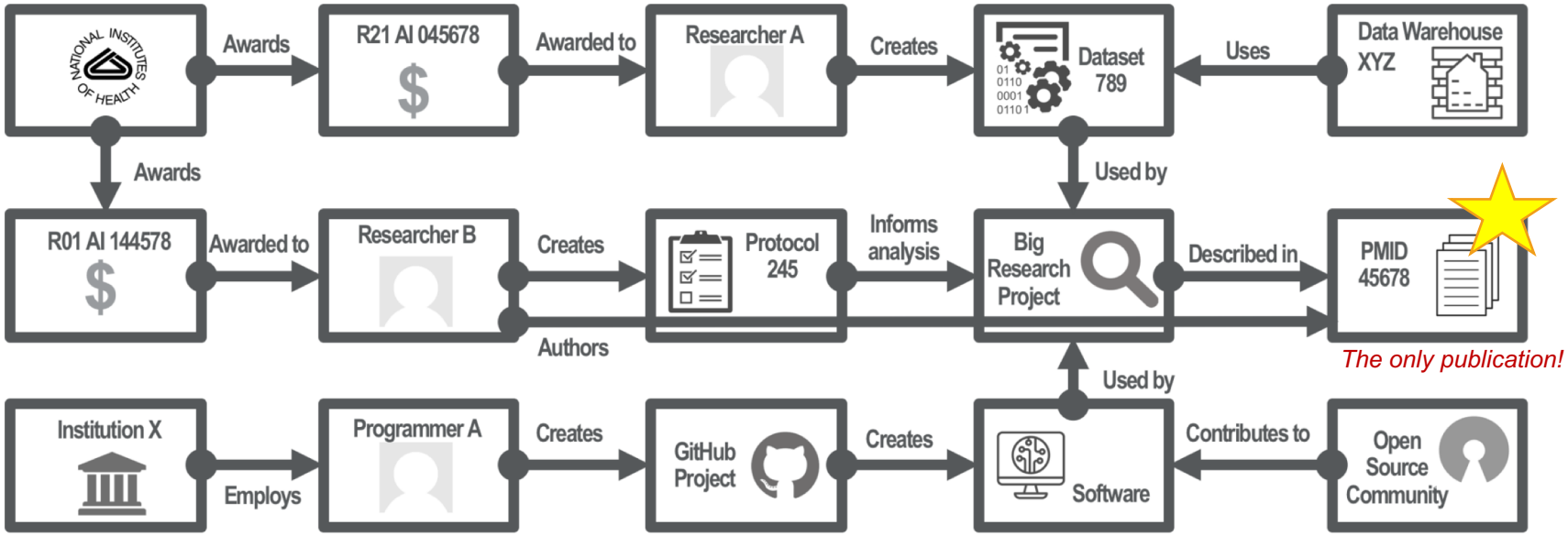
Diverse roles, diverse outputs, diverse impacts

Each a critical component of the research process

- New experimental methods or protocols
- Datasets, data models, or databases
- software tools or algorithms
- New diagnostic criteria
- New standards of care
- Biological materials or animal models
- Consent documents
- Clinical/practice guidelines
- Quality measure guidelines
- Maps and visualizations
- Clinical trials
- Measurement instruments
- Continuing education materials
- Quality measure guidelines
- Cost-effective intervention
- Consensus development conferences
- American Medical Association Current Procedural Terminology (CPT) codes
- Change in delivery of healthcare services
- Gray literature

https://becker.wustl.edu/impact-assessment
http://nucats.northwestern.edu/
https://libguides.galter.northwestern.edu/MIC

Architecting Attribution



Adapted from original by @figgyjam

A wide range of contributions beyond traditional papers are required to drive research. We're building on CRedIT and community input to make it possible to describe, give credit for, highlight the impact of non-traditional contributions to research

https://github.com/data2health/architecting_attribution

Interested in learning more or joining us? sign-up at the bottom of the page linked here

CTSA Clinical & Translational Science Awards Program

NATIONAL CENTER FOR DATA TO HEALTH

With thanks…

Guillaume Viger

Sara Gonzales

Lisa O'Keefe

Matt Carson

Teams

- Galter Health Sciences Library & Learning Center
- Center for Data to Health (CD2H)
- Lars & the Invenio team
- Chunlei Wu, Scripps
- Northwestern University Clinical and Translational Sciences Institute (NUCATS)
- Collaborators: ChicagoCHEC, FIRST DailyLife, Health for All, OpenVIVO

NIH Support

- U24TR002306 (NCATS)
- UL1TR001422 (NCATS)
- U54CA202995, U54CA202997, U54CA203000 (NCI)
- P30AR072579 (NIAMS)
- G08LM012688 (NLM)