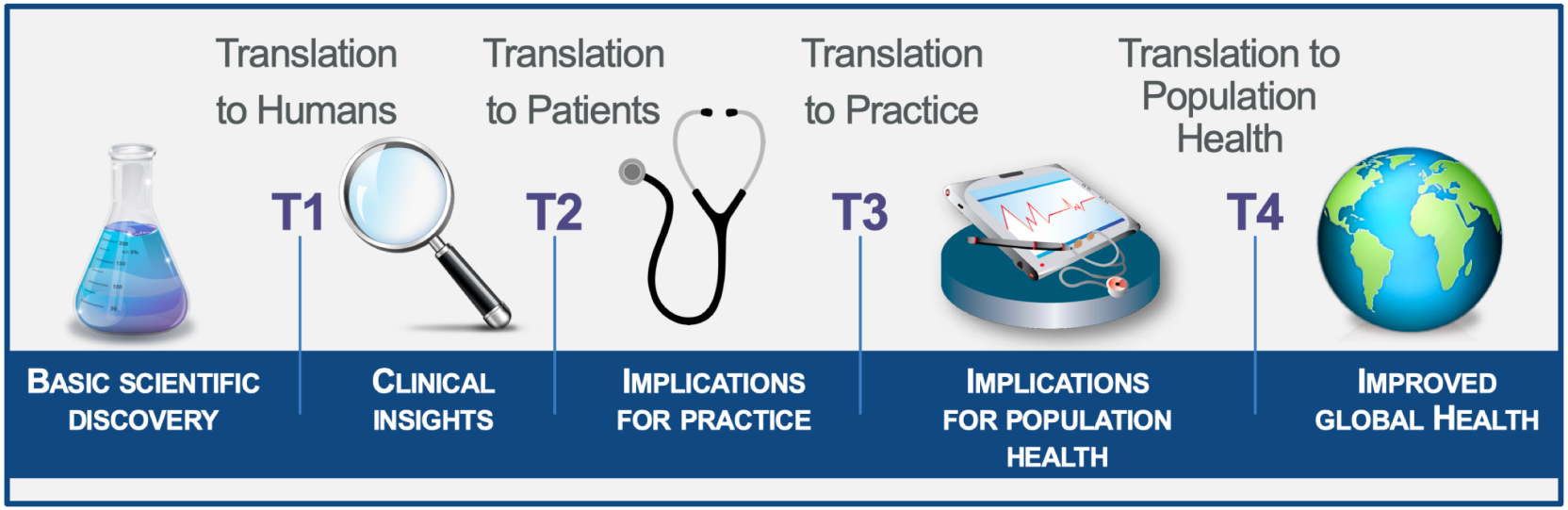
Northwestern Medicine®
Feinberg School of Medicine

CTSA Program National Center for Data to Health (CD2H)
Galter Health Sciences Library & Learning Center
NUCATS

InvenioRDM Project Kick-Off
24 June 2019

kristi.holmes@northwestern.edu
@kristiholmes

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

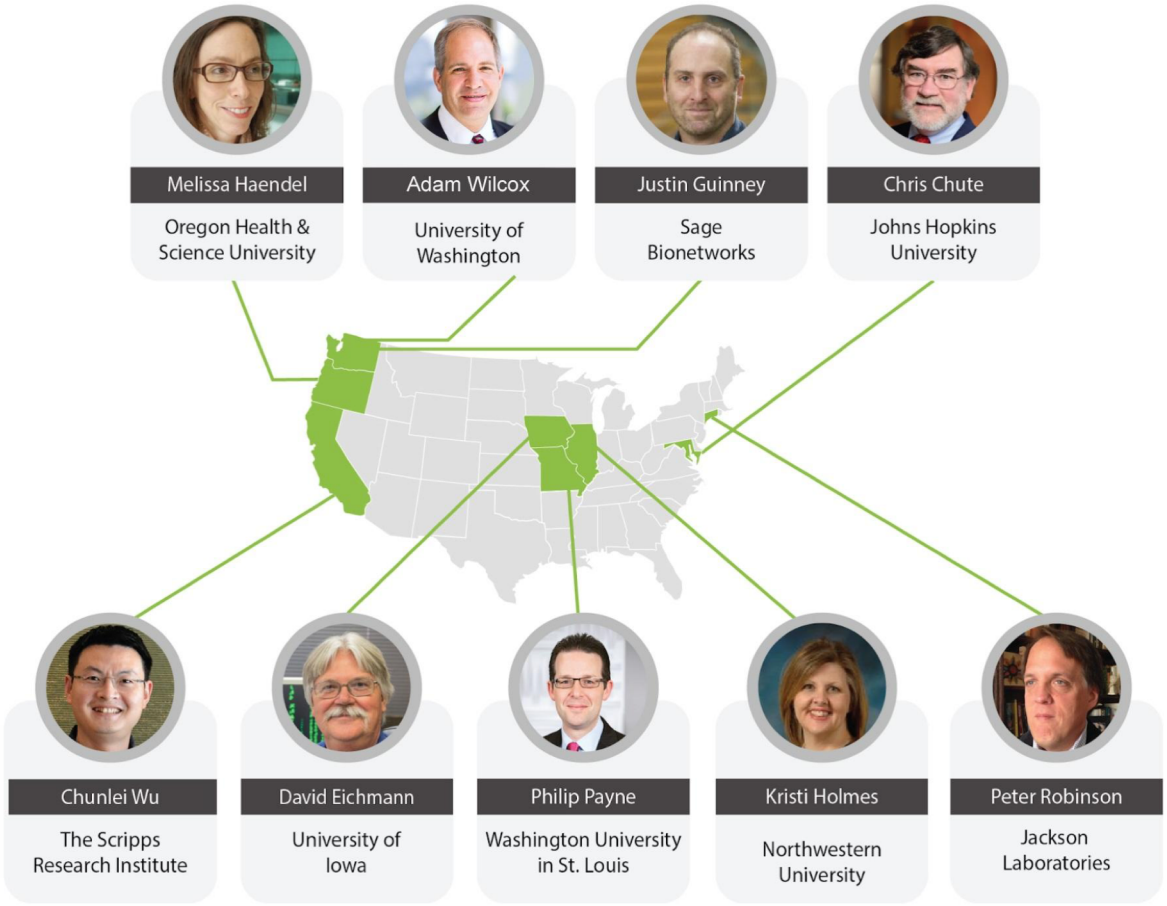
Leverage Expertise and Foster a More Collaborative CTSA Culture

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



Workforce Development

Collaboration/ Engagement

Integration Across the Lifespan

Methods/ Processes

Informatics

CLIC Center for Leading Innovation & Collaboration

…& the larger informatics community

Who are we?

Guillaume Viger

Sara Gonzales

Lisa O'Keefe

Matt Carson

Kristi Holmes

Teams

- **Galter Health Sciences Library & Learning Center**
- **Center for Data to Health (CD2H)**
- Northwestern University Clinical and Translational Sciences Institute (NUCATS)
- Collaborators: Institute for Innovations in Developmental Sciences (DevSci), ChicagoCHEC, FIRST DailyLife, Health for All, OpenVIVO

This work is supported by the National Institutes of Health's National Center for Advancing Translational Sciences, Grant Number U24TR002306.

Northwestern University Feinberg School of Medicine

In Chicago? Stop by for a visit!

Matt Carson matthew.carson@northwestern.edu
Kristi Holmes kristi.holmes@northwestern.edu

What can we contribute?

- We can contribute aligned work output from one developer (Guillaume - potentially two).
- Sara, our data librarian with domain expertise, has been doing significant user consultations, persona establishment, user story write-ups and will be conducting user testing with Guillaume.
- We also have team members (Kristi, Lisa, Matt) with strong experience in community engagement, dissemination, and governance models for collaborative communities. We have multiple potential test sites/partners of our own.
- DigitalHub migration (Fedora/Samvera)

Other resources:

- Repository and Index Software Comparison. Available at https://github.com/data2health/repository-and-index-software
- 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

Questions…

What are areas where we can contribute and how so?

- Commitment levels? Organization - through sprints, tight / loose alignments?
- Ideally flexible alignments for us and reporting through automated tests and few presentations (main developer is already spending up to 20% of his time reporting)

If you are already running a Zenodo clone (what works, what can be improved):

- We are not running a Zenodo clone. We started with a cookie cutter instance and have been truly using Invenio as a framework from the start.
- Our work has ranged from configuration of Invenio's features, to theming, to forking modules, to developing internal instance modules and finally to developing an external module.

Most difficult?

Some of the more friction-full tasks have been:

- Finding out how some things are achieved and why Invenio does it one way or the other; until we find it, we sometimes go against the grain. (Additional documentation about reasoning always helps)
- Working with the front-end (javascript) part of the code (we know there is work on this :))

- Defining boundaries and integration points between modules
  -- Use events, imported functions, entrypoints or configuration? (might be clearer with experience)
  -- How to create modules that extend the core record data model, while still letting the instance developer have the last say.
  -- Decoupling certain functionality: e.g., some indexing triggered at the POST processing could be triggered lower down

- Local ElasticSearch index used and destroyed in test (pet peeve :))

- Devops work
- There is a lot to digest for a single person

Best?

- People and community are truly friendly and knowledgeable
- Invenio is versatile and can be adapted to one's needs
- Performance. It has even been noticed by users of legacy system

- Technology choices
- Openness. Having Invenio and Zenodo open-source allows one to see how features of Invenio were used.
- Good vision for the project / understanding of the landscape
- NGR Collaborations

- Enabling straightforward access to repository technologies

What outcome(s) do we expect & wish for?

- A working InvenioRDM that will be usable for a consortium-shared cloud instance or on-premise installations starting with our own library. We are already working on such an instance and the Invenio-RDM is a natural generalization of our work.
- A fully realized Next-Generation Repository (roadmap).
- A framework to support interoperability and future development (e.g. data discovery engine with CD2H)
- A collaborative and friendly community that develops and grows around the InvenioRDM project. Materials to support and empower users, including documentation, user stories, workflows, and interfaces with key initiatives (i.e. discovery, FAIR, metrics, good data practice, etc.).
- A strong and sustainable open source community, empowered with useful governance to support participation; mechanisms to give credit for contributions

Thank you!

This work is supported by the National Institutes of Health's National Center for Advancing Translational Sciences, Grant Number U24TR002306.