Making it count
A computational approach to attribution

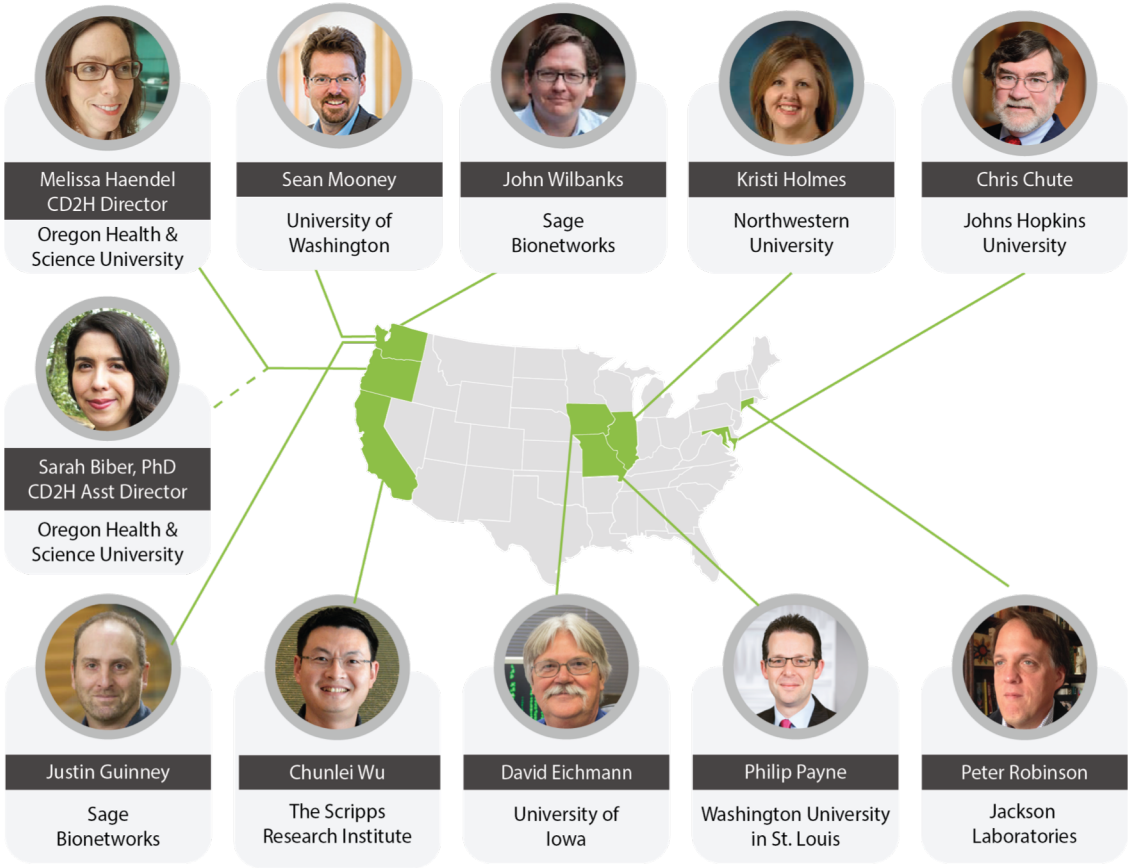
A project to make more meaningful connections between people, their roles, their work and impacts.

29 October 2018
Kristi Holmes, PhD
#RO2018. @kristiholmes

CTSA Clinical & Translational Science Awards Program

NATIONAL CENTER FOR DATA TO HEALTH

Who we are and who we serve



The community we serve

9 CD2H Sites

iDTF

CLIC Center for Leading Innovation & Collaboration

NIH National Center for Advancing Translational Sciences

The larger informatics community

CD2H: National Center for Data to Health

Data & Informatics 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

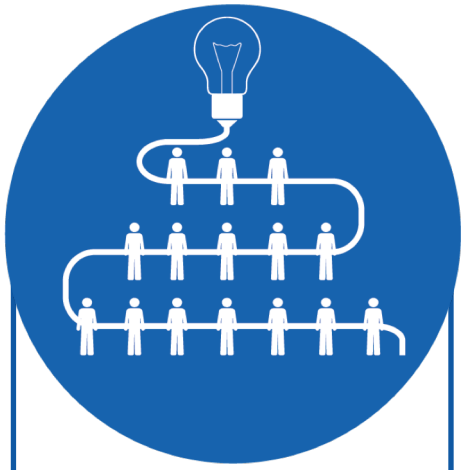
Better translation of research and improved patient care

What *IS* impact?

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

- IMPROVEMENTS IN HEALTH THROUGH TREATMENT AND PREVENTION
- CONTRIBUTIONS TO SOCIETY THROUGH ECONOMIC GROWTH AND PRODUCTIVITY
- EXPANSION OF THE BIOMEDICAL KNOWLEDGE BASE THROUGH CUTTING-EDGE RESEARCH
- CULTIVATION OF THE BIOMEDICAL WORKFORCE OF TODAY AND TOMORROW

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

CTSA Clinical & Translational Science Awards Program

NATIONAL CENTER FOR DATA TO HEALTH

Contributors and expertise needed for a genetic diagnosis

Clinical/care

Thomas Markello
Dong Chen
Justin Y. Kwan
Iren Horkayne-Szakaly
Alan Morrison
Olga Simakova
Irina Maric
Jay Lozier
Andrew R. Cullinane
Tatjana Kilo
Lynn Meister
Kourosh Pakzad
Sanjay Chainani
Roxanne Fischer
Camilo Toro
James G. White
David Adams
Cornelius Boerkoel
William A. Gahl
Cynthia J. Tifft
Meral Gunay-Aygun

Pathology

Hans Goeble
Karen Balbach
Nadine Pfeifer
Sandra Werner
Christian Linden

Ontologist

Melissa Haendel
Peter Robinson
Chris Mungall
Sebastian Kohler
Cindy Smith
Nicole Vasilevsky
Sandra Dolken

CS/informatics

Elizabeth Lee
Amanda Links
Will Bone
Murat Sincan
Damian Smedley
Jules Jacobson
Nicole Washington
Elise Flynn
Sebastian Kohler
Orion Buske
Marta Girdea
Michael Brudno
Jeremy Band

Curator

Melissa Haendel
David Adams
David Draper
Bailey Gallinger
Joie Davis
Nicole Vasilevsky
Heather Trang
Rena Godfrey
Gretchen Golas
Catherine Groden
Michele Nehrebecky
Ariane Soldatos
Elise Valkanas,
Colleen Wahl
Lynne Wolfe

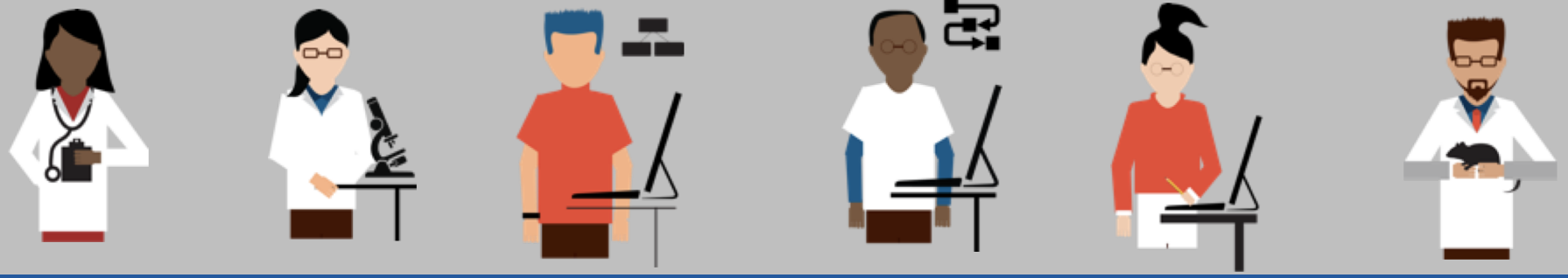
Basic research

Johannes Grosse
Attila Braun
David Varga-Szabo
Niklas Beyersdorf
Boris Schneider
Lutz Zeitlmann
Petra Hanke
Patricia Schropp
Silke Mühlstedt
Carolin Zorn
Michael Huber
Carolin Schmittwolf
Wolfgang Jagla
Philipp Yu
Thomas Kerkau
Harald Schulze
Michael Nehls
Bernhard Nieswandt

@ontowonka

Contributors and expertise needed for a genetic diagnosis

Clinical/care | Pathology | Ontologist | CS/informatics | Curator | Basic research



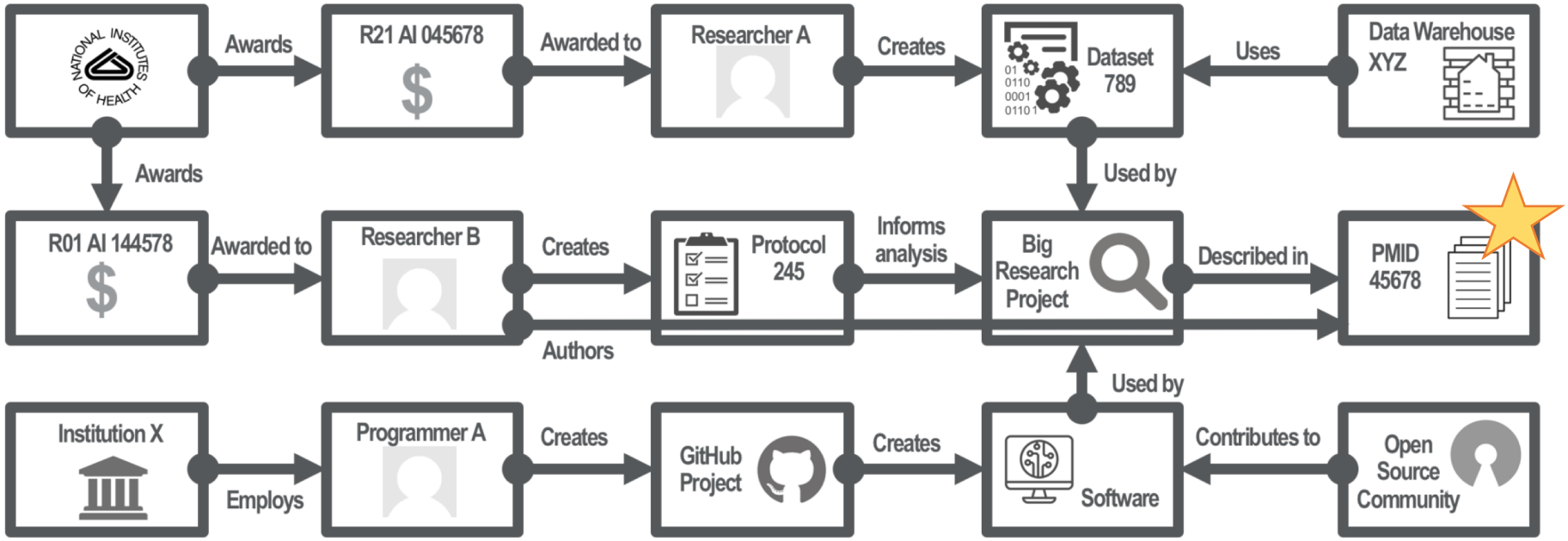
How can we credit all of these contributors?

Moreover, how can we find the resources and people to form the scientific teams, collaborations, reviewers, we need?

A community journey

Contributor Roles & Research Outputs

Better attribution: extending credit beyond the publication



Adapted from Julie McMurry

What work is being done, who is doing it, and what outputs are being created?

1. Understand deeply the requirements for a computable attribution system from a large diversity of stakeholders;
2. Build model(s) to meet these requirements (CRO, ROO);
3. Evaluate the models in real pilot systems with real data.

By using contribution roles & research outputs to develop infrastructure to understand the scholarly ecosystem, we can better understand, leverage, and credit a diverse translational workforce

The Informatics of Attribution

1. Understand deeply the requirements for a computable attribution system from a large diversity of stakeholders;
2. Build model(s) to meet these requirements; and
3. Evaluate the models in real pilot systems with real data.

Development of data models to address these needs demands a rigorous requirements-driven approach

Key modeling challenges for development of integrative community standards

1. Accommodation of diverse and complex data types
2. Support needs of different applications and systems
3. Interoperability with broader data landscape

Key tools necessary to drive change

1. Technology
2. Persistent identifiers
3. Data models
4. Connections – of all kinds!

NATIONAL CENTER FOR DATA TO HEALTH

@kristiholmes & @ontowonka

OpenVIVO

Implementation of a community-driven concept of credit

1. Provide a VIVO experience for everyone, a demonstration of VIVO, a platform for experimentation, and an ownership experience for the VIVO team
2. Use persistent identifiers for all entities – people (ORCiD), works (DOI and PMID), organizations (GRID), journals (ISSN), concepts (FAST)
3. Automatic, real-time ingest of metadata from identifiers via public APIs
4. Publication of data
5. Consumption and reuse of data
6. Attribution of works by scholars to indicate roles in works

--Mike Conlon, VIVO Project Director

My own profile was completed entirely with publicly available data via ORCID and DOIs and it took about 15 minutes to complete from start to finish.

http://openvivo.org/

OpenVIVO

Implementation of a community-driven concept of credit

1. Provide a VIVO experience for everyone, a demonstration of VIVO, a platform for experimentation, and an ownership experience for the VIVO team
2. Use persistent identifiers for all entities – people (ORCiD), works (DOI and PMID), organizations (GRID), journals (ISSN), concepts (FAST)
3. Automatic, real-time ingest of metadata from identifiers via public APIs
4. Publication of data
5. Consumption and reuse of data
6. Attribution of works by scholars to indicate roles in works

--Mike Conlon, VIVO Project Director

My o *DOIs and i*

http://openvivo.org/

frontiers in Research Metrics and Analytics

TECHNOLOGY REPORT ARTICLE
Front. Res. Metr. Anal., 01 March 2018 | https://doi.org/10.3389/frma.2017.00012

OpenVIVO: Transparency in Scholarship

Violeta Ilik[1]*, Michael Conlon[2], Graham Triggs[3], Marijane White[4], Muhammad Javed[5], Matthew Brush[4], Karen Gutzman[6], Shahim Essaid[4], Paul Friedman[6], Simon Porter[7], Martin Szomszor[7], Melissa Anne Haendel[4], David Eichmann[8] and Kristi L. Holmes[6]

[1]Stony Brook University, Stony Brook, NY, United States
[2]University of Florida, Gainesville, FL, United States
[4]land, OR, United States
n University, Chicago, IL, United States
[7]Digital Science, London, United Kingdom
[8]University of Iowa, Iowa, IA, United States

Front. Res. Metr. Anal. **2,** (01 March 2018) doi:10.3389/frma.2017.00012

OpenVIVO is a free and open-hosted semantic web platform that anyone can join and that gathers and shares open data about scholarship in the world. OpenVIVO, based on the VIVO open-source platform, provides transparent access to data about the scholarly work of its participants. OpenVIVO demonstrates the use of persistent identifiers, the automatic real-time ingest of scholarly ecosystem metadata, the use of VIVO-ISF and related ontologies, the attribution of work, and the publication and reuse of data—all critical components of presenting, preserving, and tracking scholarship. The system was created by a cross-institutional team over the course of 3 months. The team created and used RDF models for research organizations in the world based on Digital Science GRID data, for academic journals based on data from CrossRef and the US National Library of Medicine, and created a new model for attribution of scholarly work. All models, data, and software are available in open repositories.

Transparency in Scholarship

Scholarship requires knowledge of previous work. The growth of scholarship worldwide and the proliferation of scholarly output types—from papers and monographs to preprints, conference papers, datasets, posters, and presentation slides—have fundamentally changed the scholarly ecosystem from an environment dependent on libraries to one that is dependent on the electronic resources made available by libraries to support discovery and knowledge transfer. This shift clearly drives a need for the representation of scholarly works using standard metadata formats to facilitate indexing and discovery.

For scholars to have knowledge of previous work, the work must be indexed and discoverable *via* electronic systems. Metadata regarding the

PEOPLE ALSO LOOKED AT

Eugene Garfield's Influences upon the Future of Evaluative Informetrics
Henk F. Moed

Openness and Impact of Leading Scientific Countries
Caroline S. Wagner, Travis Whetsell, Jeroen Baas and Koen Jonkers

Active Role of the Necrotic Zone in Desensitization of Hypoxic Macrophages and Regulation of CSC-Fate: A hypothesis
Maryam Mehrabi, Fatemeh Amini and Shima Mehrabi

Predictive Effects of Novelty Measured by Temporal Embeddings on the Growth of Scientific Literature
Jiangen He and Chaomei Chen

Interdependent Followers Prefer Avoidant Leaders: Followers' Cultural Orientation Moderates Leaders' Avoidance Relationships

frontiers in Research Metri and Analytics

JOURNAL ABOUT ARTICLES RESEARCH TOPICS FOR AUTHORS EDITORI

‹ Articles

THIS ARTICLE IS P
Evaluating Researc

TECHNOLOGY REPORT ARTICLE

Front. Res. Metr. Anal., 01 March 2018 | https://doi.org/10.3389/frma.2017.00012

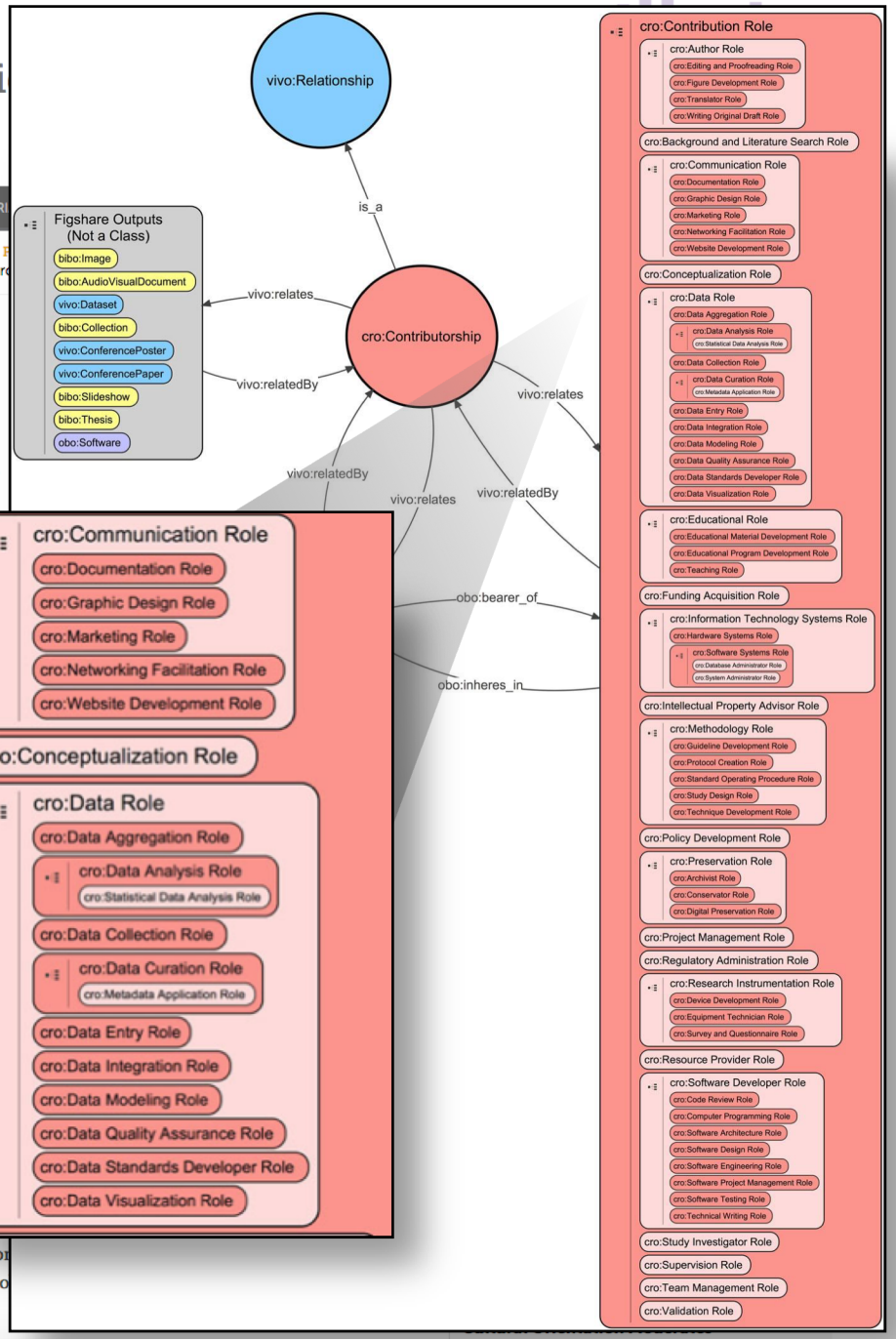
OpenVIVO: Transparency in Scholarship

Violeta Ilik[1]*, Michael Conlon[2], Graham Triggs[3], Marijane White[4], M Gutzman[6], Shahim Essaid[4], Paul Friedman[6], Simon Porter[7], Martin Szo Eichmann[8] and Kristi L. Holmes[6]

, United States

Front. Res. Metr. Anal. **2,** (01 March 2018) doi:10.3389/frma.2017.00012

Join us! Submit tickets! ☺
https://github.com/data2health/contributor-role-ontology

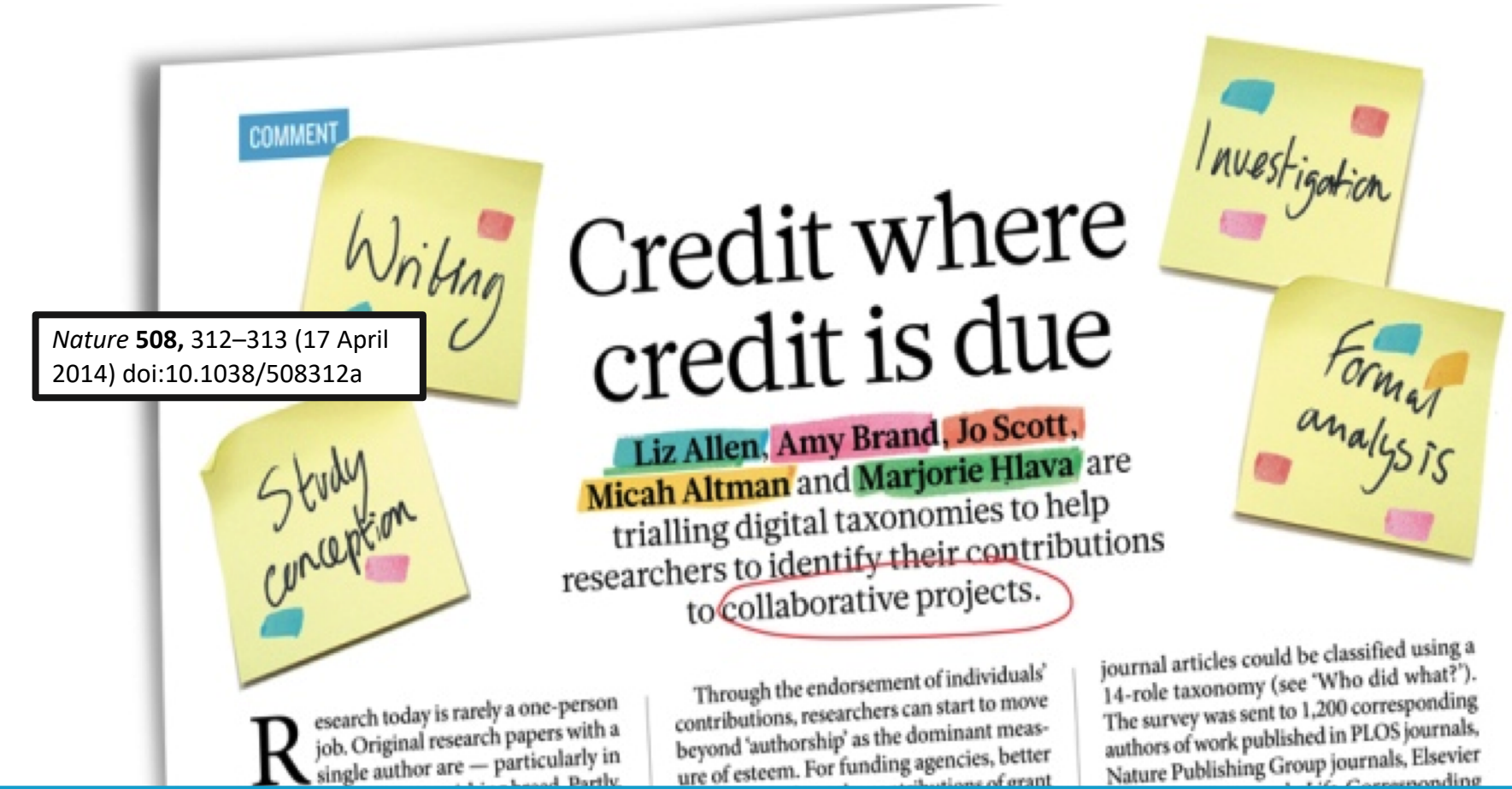


latform that anyone can joi
IVO open-source platform,
strates the use of persistent
scholarly ecosystem metadata, the use of VIVO-ISF and related ontologies, the a
data—all critical components of presenting, preserving, and tracking scholarshi
team over the course of 3 months. The team created and used RDF models for re
Science GRID data, for academic journals based on data from CrossRef and the
new model for attribution of scholarly work. All models, data, and software are a

Transparency in Scholarship

Scholarship requires knowledge of previous work. The growth of scholarship worldwide a
papers and monographs to preprints, conference papers, datasets, posters, and presentat
ecosystem from an environment dependent on libraries to one that is dependent on the electro
discovery and knowledge transfer. This shift clearly drives a need for the representation of scho
facilitate indexing and discovery.

For scholars to have knowledge of previous work, the work must be indexed and discoverable *via* electronic systems. Metadata regarding the



Nature **508,** 312–313 (17 April 2014) doi:10.1038/508312a

CRT

CRediT

CRediT ontology in OWL:
https://github.com/data2health/credit-ontology

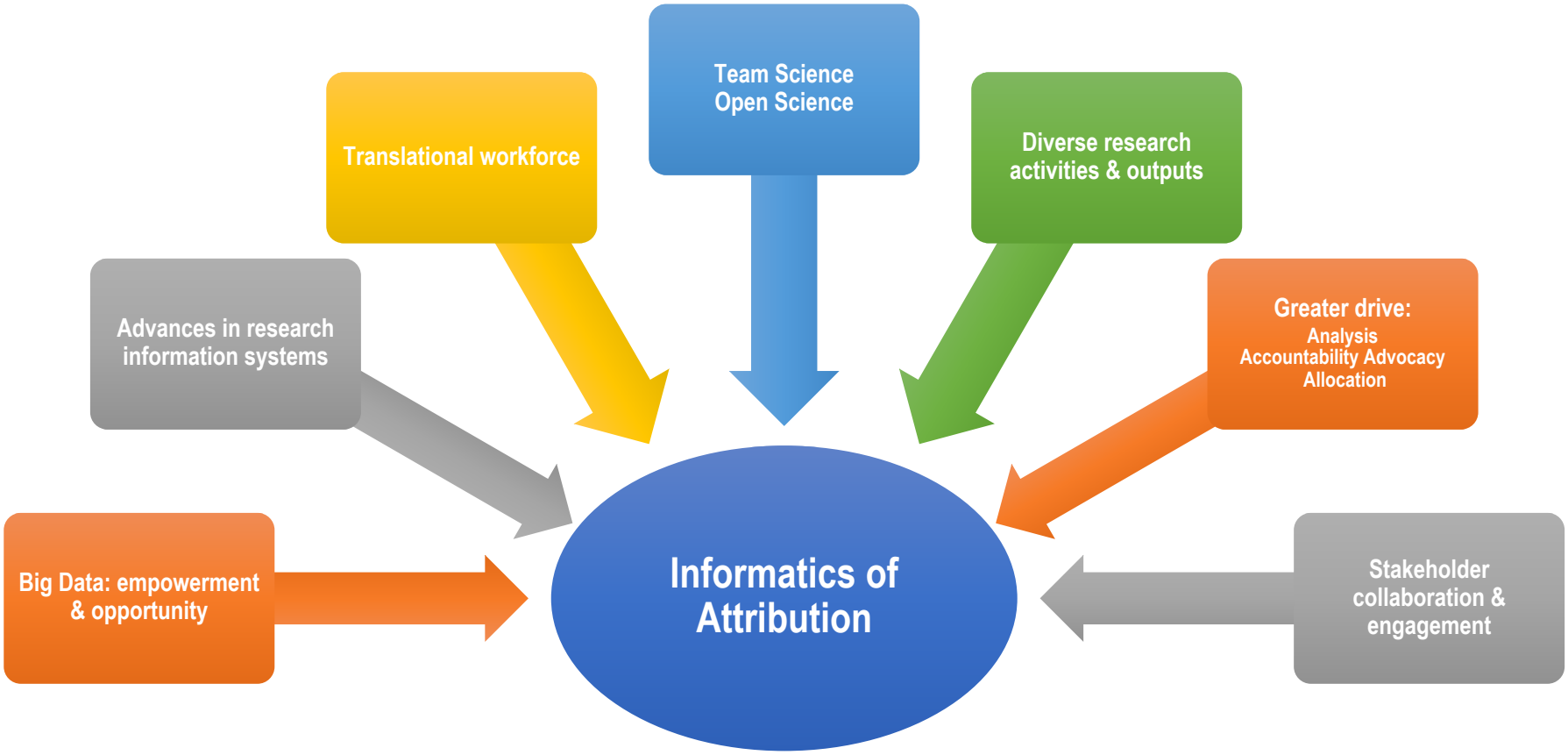
CRediT is high-level taxonomy, including 14 roles, that can be used to represent the roles typically played by contributors to scientific scholarly output. The roles describe each contributor's specific contribution to the scholarly output.

https://casrai.org/credit/

It takes technology + culture.

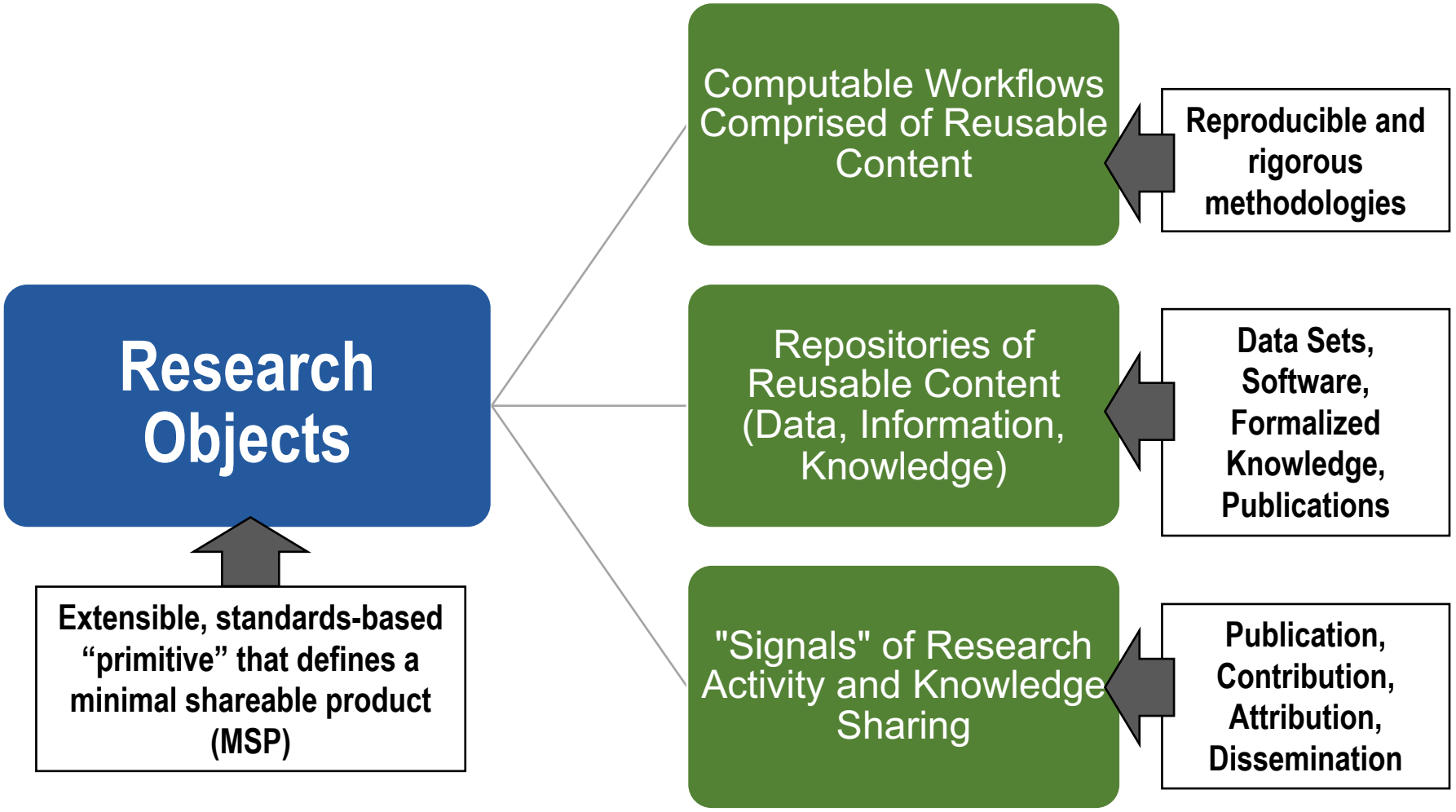
Why now & how do we get there?

http://bit.ly/AttributionSignUp



Research Objects: A Common Unit of Sharing Across Use Cases

@prpayne5



Thank you

Teams

- CD2H
  - ***NU team:*** *Karen Gutzman, Patty Smith, Sara Gonzales*
  - ***OHSU team****: Marijane White, Nicole Vasilevsky, Melissa Haendel*
- Northwestern University Clinical and Translational Sciences Institute
- OpenVIVO collaborators, Force11 Attribution WG, NISO, Cathy Sarli & Becker Library
- Galter Library, NUCATS, ChicagoCHEC, FIRST DailyLife, Health for All

NIH Support

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

Contact us!

kristi.holmes@northwestern.edu
@kristiholmes

https://ctsa.ncats.nih.gov/cd2h/
@data2health

Links to selected resources and projects

- National Center for Advancing Translational Sciences: https://ncats.nih.gov/
- Clinical and Translational Science Award (CTSA) Program: https://ctsacentral.org/
- Northwestern University Clinical and Translational Sciences Institute: https://nucats.northwestern.edu/
- OpenVIVO: http://openvivo.org/ and https://wiki.duraspace.org/display/VIVO/OpenVIVO+Task+Force
- CD2H: https://ctsa.ncats.nih.gov/cd2h/ and https://github.com/data2health
- FORCE11 Attribution Working Group: https://www.force11.org/group/attributionwg
- National Information Standards Organization. (2016). Outputs of the NISO Alternative Assessment Metrics Project. [Recommended Practice RP-25-2016]. Retrieved from https://goo.gl/n7JV2z. National Information Standards Organization. Outputs of the NISO Alternative Assessment Metrics Project. 2016. Report No.: Recommended Practice RP-25-2016.
- Sarli CC, Dubinsky EK, Holmes KL. Beyond citation analysis: a model for assessment of research impact. J Med Libr Assoc. 2010; 98(1):17-23. PMID: 20098647; PMCID: PMC2801963.
- Becker Medical Library Model for Assessment of Research Impact. St. Louis, MO: Bernard Becker Medical Library; 2014. Available from: https://becker.wustl.edu/impact-assessment
- Gutzman KE, Konkiel S, White M, et al. Attribution of Work in the Scholarly Ecosystem 2016. Available from: https://figshare.com/articles/Attribution_of_Work_in_the_Scholarly_Ecosystem/3175198 .
- White M, Haendel M, Brush M. Contribution Ontology: A repository for representation of a person's role in research processes and outputs [Data model]. 2016 Available from: https://github.com/openrif/contribution-ontology
- European Commission Directorate-General for Research and Innovation, Open Science Working Group on Rewards/Recognition. Evaluation of Research Careers fully acknowledging Open Science Practices: Rewards, incentives and/or recognition for researchers practicing Open Science [Report].. Brussels, Belgium: European Commission; 2017. Available from: https://ec.europa.eu/research/openscience/pdf/os_rewards_wgreport.pdf - view=fit&pagemode=none