

Architecting Attribution

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



24 January 2019
Kristi Holmes, PhD
PIDapalooza 2019 @kristiholmes

Who we are and who we serve



The community we serve



9 CD2H Sites

iDTF



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



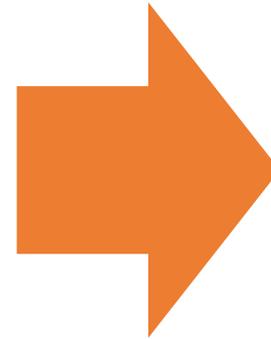
Make Data Easier to Share
and Re-use



Make Tools More Accessible
and Interoperable



Leverage Expertise and
Foster a More Collaborative
CTSA Culture



Better translation
of research and
improved patient
care



Tools



Ontologies



Algorithms



Data



Standards



Basic

Genomics

Proteomics

Metabolic Pathways

Molecular Modeling

Molecular Simulation

Cellular Models

Molecular Assays

Chasm of semantic despair



Biospecimens

Lab Data

Trials Data

Disease & Syndrome

Medical Imaging

EHR Structures

Patient Record Data

Clinical



What is attribution???



WIKIPEDIA
The Free Encyclopedia

[Main page](#)
[Contents](#)
[Featured content](#)
[Current events](#)
[Random article](#)
[Donate to Wikipedia](#)
[Wikipedia store](#)

[Interaction](#)

Article [Talk](#)

[Read](#) [Edit](#) [View](#)

Attribution

From Wikipedia, the free encyclopedia

Attribution may refer to:

- [Attribution \(copyright\)](#), concept in copyright law requiring an author to be credited
- [Attribution \(marketing\)](#), concept in marketing of assigning a value to a marketing activity based on desired outcome
- [Attribution \(psychology\)](#), concept in psychology whereby people attribute traits and causes to things they observe
- [Performance attribution](#), technique in quantitative finance for explaining the active performance of a portfolio
- [Journalism sourcing](#) (or attribution), journalistic practice of attributing information to its source

Drosophila Muller F Elements Maintain a Distinct Set of Genomic Properties Over 40 Million Years of Evolution



[« Previous](#) | [Next Article »](#)
[Table of Contents](#)

Wilson Leung^a, Christopher D. Shaffer^a, Laura K. Reed^b, Sheryl T. Smith^c,
William Barshop^a, Wil
Jeannette Wong^a, Dav
Joshua F. Machone^d,
Srebrenka Robic^e, Erin K. Luippold^e, Shannon R. McCartha^e, Tezin A. Walji^e,
Chelsea A. Walker^e, Kenneth Saville^f, Marita K. Abrams^f, Andrew R. Armstrong^f,
William Armstrong^f, Robert J. Bailey^f, Chelsea R. Barberi^f, Lauren R. Beck^f,
Amanda L. Blaker^f, Christopher E. Blunden^f, Jordan P. Brand^f, Ethan J. Brock^f,
Dana W. Brooks^f, Marie Brown^f, Sarah C. Butzler^f, Eric M. Clark^f, Nicole B. Clark^f,
Ashley A. Collins^f, Rebecca J. Cotteleer^f, Peterson R. Cullimore^f,
Seth G. Dawson^f, Carter T. Docking^f, Sasha L. Dorsett^f, Grace A. Dougherty^f,
Kaitlyn A. Downey^f, Andrew P. Drake^f, Erica K. Earl^f, Trevor G. Floyd^f,
Joshua D. Forsyth^f, Jonathan D. Foust^f, Spencer L. Franchi^f, James F. Geary^f,
Cynthia K. Hanson^f, Taylor S. Harding^f, Cameron B. Harris^f,
Jonathan M. Heckman^f, Heather L. Holderness^f, Nicole A. Howey^f,
Dontae A. Jacobs^f, Elizabeth S. Jewell^f, Maria Kaisler^f, Elizabeth A. Karaska^f,

Over 1000 authors

This Article

Investigation

Early Online March 4, 2015
doi: 10.1534/g3.114.015966
G3 May 1, 2015 vol. 5 no. 5
719-740

Supporting Information
Blog Post: Undergrads power
genomics research

All Versions of this Article:
g3.114.015966v1
5/5/719 most recent

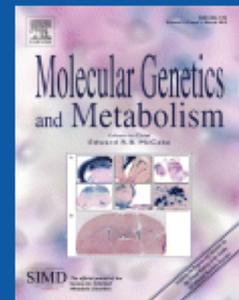
Services

- Email this article to a colleague
- Alert me when this article is cited
- Alert me if a correction is posted
- Alert me when eletters are published
- Article Usage Statistics
- Similar articles in this journal
- Similar articles in PubMed
- Download to citation manager

Contributors and expertise needed for a genetic diagnosis



Credit extends beyond the publication



- ◆ Johannes creates [stim1 mouse](#)
- ◆ Melissa curates patient data for [UDP_2542](#)
- ◆ Will performs analysis of [UDP_2542](#) data that includes [stim1 mouse](#) to generate a dataset of [prioritized variants](#)
- ◆ Tom writes publication [pmid:25577287](#) about the STIM1 diagnosis
- ◆ Tom explicitly credits Will as an author but not Melissa.

Melissa Heandel
OSU/OHSU
CD2H PEA group

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

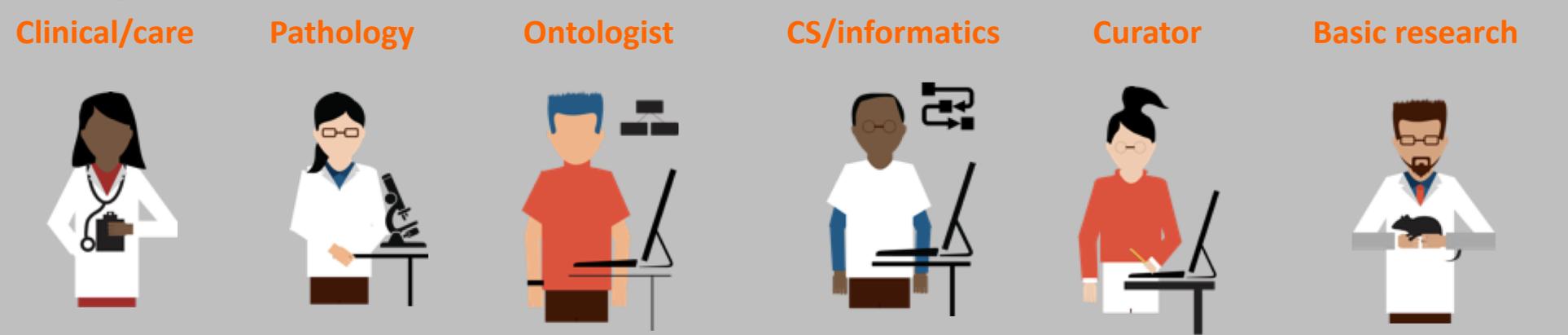
Basic research



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

@ontowonka

Contributors and expertise needed for a genetic diagnosis



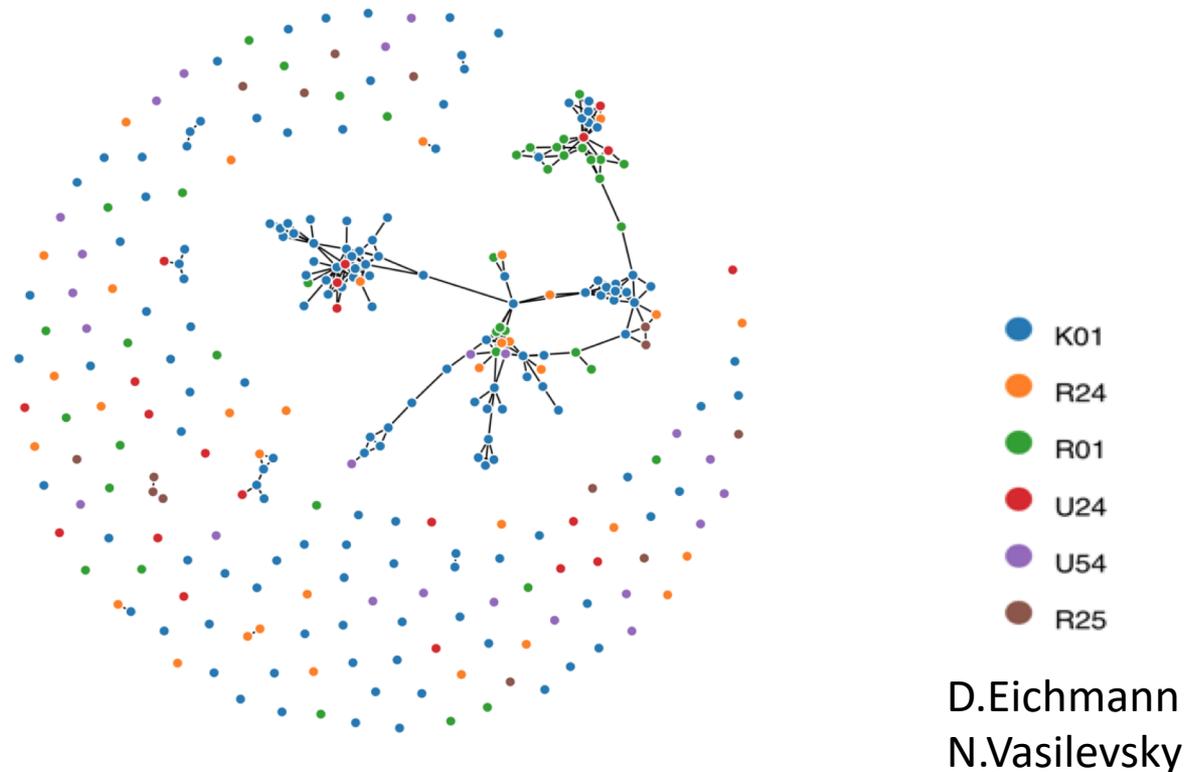
How can we credit all of these contributors?

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

Many contributions don't lead to authorship



NIH BD2K co-authorship

> 20% key personnel are not profiled using publications

PubMed Central Open Access

- 2M++ articles
- 1M+ with acknowledgements
- 5M+ sentences
- 21M+ parses
- 36M+ fragments

- MEDLINE yesterday: 30M articles

The Simple Cases

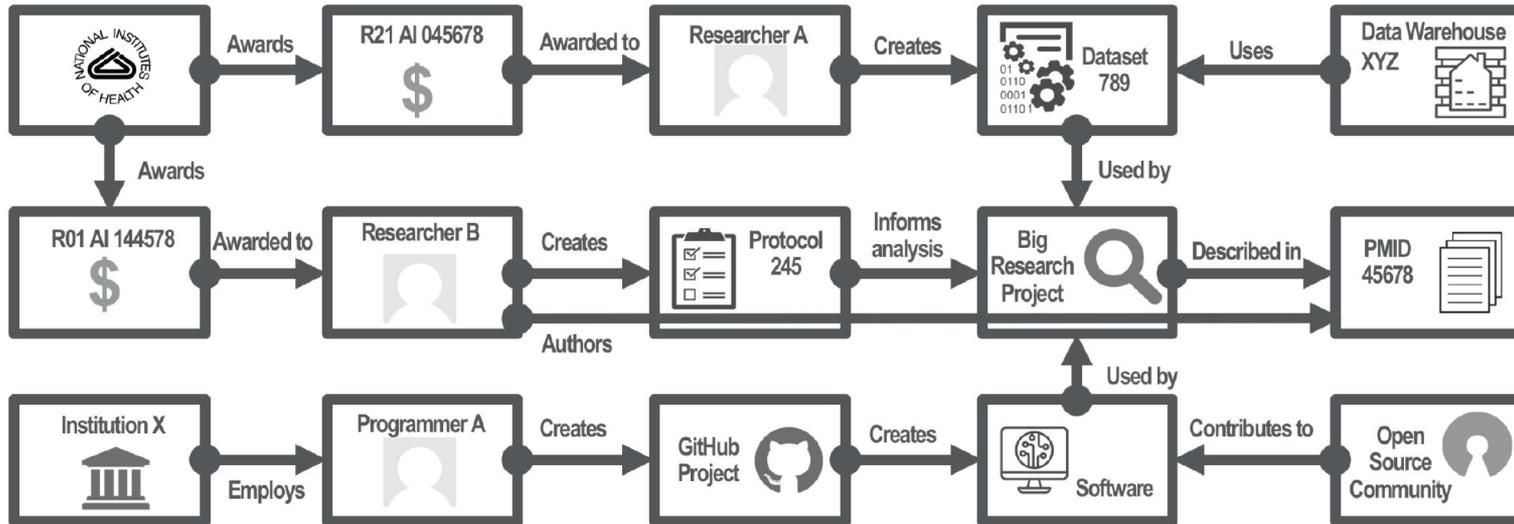
- PMID: 3008610
- SeqNum: 2
- SentNum: 6
- Sentence: EK analysed the data.
- POS: [EK/NNP, analysed/VBD, the/DT, data/NNS, ./.]
- Parse: [S
 [NP EK/NNP]
 [VP analysed/VBD
 [NP the/DT data/NNS]
] ./.]

And the Not So Simple...

- PMID: 4159542
- Sentence: We thank Sheila Harvey, Clinical Trials Unit Manager at ICNARC, and Ruth Canter, Trials Administrator at ICNARC, for their assistance in chasing completed surveys; Dr Kevin Gunning for early advice and project development; Drs Neill K. J. Adhikari and Gordon D. Rubenfeld for feedback and discussion of analysis plan; Dr Chris AKY Chong for his valuable comments on the initial draft of this manuscript; and our Responders: Addenbrooke's Hospital (Dr Kevin Gunning), Airedale General Hospital (Dr John Scriven), Alexandra Hospital (Dr Tracey Leach), Arrowe Park Hospital (Dr Lawrence Wilson), Barnet Hospital (Dr AH Wolff), ...
- 8,245 character long sentence

What needs to be considered?

Let's look at a common workflow for a technical project:



Community contributions are invisible to existing credit mechanisms

COMMENT

Writing

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

Study conception

Credit where credit is due

Liz Allen, Amy Brand, Jo Scott, Micah Altman and Marjorie Hlava are trialling digital taxonomies to help researchers to identify their contributions to collaborative projects.

Investigation

Formal analysis

Research today is rarely a one-person job. Original research papers with a single author are — particularly in

Through the endorsement of individuals' contributions, researchers can start to move beyond 'authorship' as the dominant measure of esteem. For funding agencies, better

journal articles could be classified using a 14-role taxonomy (see 'Who did what?'). The survey was sent to 1,200 corresponding authors of work published in PLOS journals, Elsevier Nature Publishing Group journals, Elsevier

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/>

Contribution Role Ontology

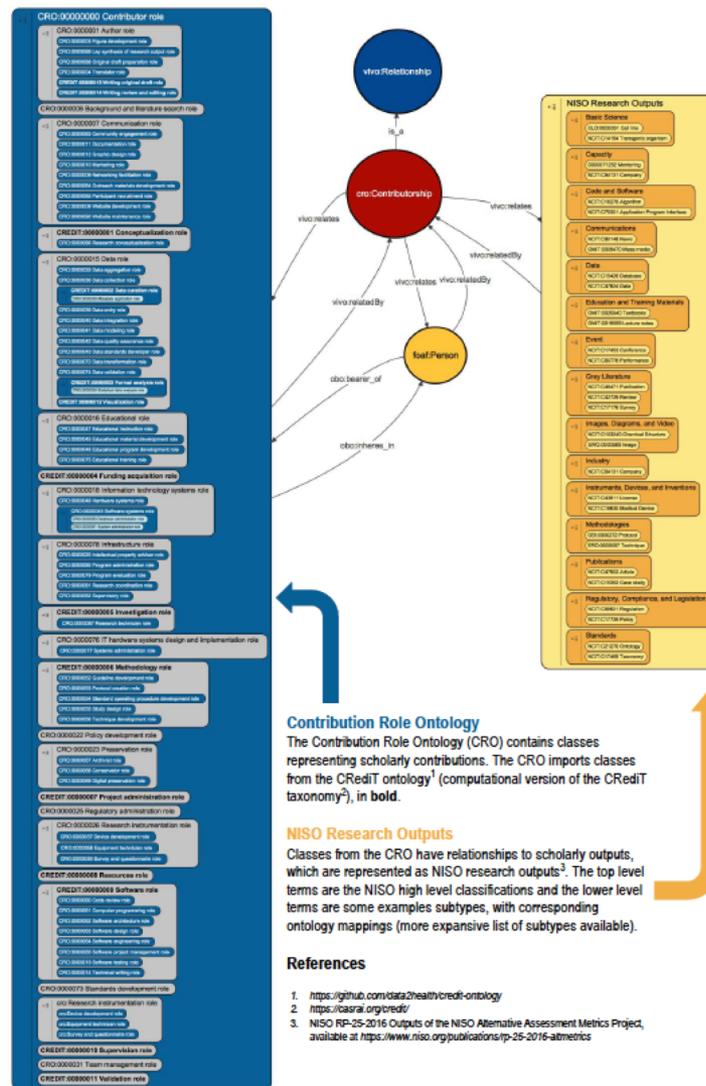
The Contribution Role Ontology (CRO) contains classes representing scholarly contributions. The CRO imports classes from the CRediT ontology¹ (computational version of the CRediT taxonomy²), in **bold**.

NISO Research Outputs

Classes from the CRO have relationships to scholarly outputs, which are represented as NISO research outputs³. The top level terms are the NISO high level classifications and the lower level terms are some examples subtypes, with corresponding ontology mappings (more expansive list of subtypes available).

References

1. <https://github.com/data2health/credit-ontology>
2. <https://casrai.org/credit/>
3. NISO RP-25-2016 Outputs of the NISO Alternative Assessment Metrics Project, available at <https://www.niso.org/publications/rp-25-2016-altmetrics>



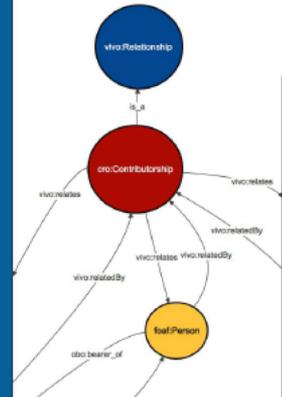
Details at
<http://bit.ly/PID19attribution>

CRO:0000015 Data role

- CRO:0000033 Data aggregation role
- CRO:0000036 Data collection role
- CREDIT:00000002 Data curation role**
 - CRO:0000038 Metadata application role
- CRO:0000039 Data entry role
- CRO:0000040 Data integration role
- CRO:0000041 Data modeling role
- CRO:0000042 Data quality assurance role
- CRO:0000043 Data standards developer role
- CRO:0000072 Data transformation role
- CRO:0000074 Data validation role
- CREDIT:00000003 Formal analysis role**
 - CRO:0000035 Statistical data analysis role
- CREDIT:00000012 Visualization role**

CRO:0000000 Contributor role

- CRO:0000001 Author role
- CRO:0000002 Data aggregation role
- CRO:0000003 Data collection role
- CRO:0000004 Data curation role
- CRO:0000005 Data entry role
- CRO:0000006 Data integration role
- CRO:0000007 Data modeling role
- CRO:0000008 Data quality assurance role
- CRO:0000009 Data standards developer role
- CRO:0000010 Data transformation role
- CRO:0000011 Data validation role
- CRO:0000012 Data visualization role
- CRO:0000013 Data aggregation role
- CRO:0000014 Data collection role
- CRO:0000015 Data curation role
- CRO:0000016 Data entry role
- CRO:0000017 Data integration role
- CRO:0000018 Data modeling role
- CRO:0000019 Data quality assurance role
- CRO:0000020 Data standards developer role
- CRO:0000021 Data transformation role
- CRO:0000022 Data validation role
- CRO:0000023 Data visualization role
- CRO:0000024 Data aggregation role
- CRO:0000025 Data collection role
- CRO:0000026 Data curation role
- CRO:0000027 Data entry role
- CRO:0000028 Data integration role
- CRO:0000029 Data modeling role
- CRO:0000030 Data quality assurance role
- CRO:0000031 Data standards developer role
- CRO:0000032 Data transformation role
- CRO:0000033 Data validation role
- CRO:0000034 Data visualization role
- CRO:0000035 Data aggregation role
- CRO:0000036 Data collection role
- CRO:0000037 Data curation role
- CRO:0000038 Data entry role
- CRO:0000039 Data integration role
- CRO:0000040 Data modeling role
- CRO:0000041 Data quality assurance role
- CRO:0000042 Data standards developer role
- CRO:0000043 Data transformation role
- CRO:0000044 Data validation role
- CRO:0000045 Data visualization role
- CRO:0000046 Data aggregation role
- CRO:0000047 Data collection role
- CRO:0000048 Data curation role
- CRO:0000049 Data entry role
- CRO:0000050 Data integration role
- CRO:0000051 Data modeling role
- CRO:0000052 Data quality assurance role
- CRO:0000053 Data standards developer role
- CRO:0000054 Data transformation role
- CRO:0000055 Data validation role
- CRO:0000056 Data visualization role
- CRO:0000057 Data aggregation role
- CRO:0000058 Data collection role
- CRO:0000059 Data curation role
- CRO:0000060 Data entry role
- CRO:0000061 Data integration role
- CRO:0000062 Data modeling role
- CRO:0000063 Data quality assurance role
- CRO:0000064 Data standards developer role
- CRO:0000065 Data transformation role
- CRO:0000066 Data validation role
- CRO:0000067 Data visualization role
- CRO:0000068 Data aggregation role
- CRO:0000069 Data collection role
- CRO:0000070 Data curation role
- CRO:0000071 Data entry role
- CRO:0000072 Data integration role
- CRO:0000073 Data modeling role
- CRO:0000074 Data quality assurance role
- CRO:0000075 Data standards developer role
- CRO:0000076 Data transformation role
- CRO:0000077 Data validation role
- CRO:0000078 Data visualization role
- CRO:0000079 Data aggregation role
- CRO:0000080 Data collection role
- CRO:0000081 Data curation role
- CRO:0000082 Data entry role
- CRO:0000083 Data integration role
- CRO:0000084 Data modeling role
- CRO:0000085 Data quality assurance role
- CRO:0000086 Data standards developer role
- CRO:0000087 Data transformation role
- CRO:0000088 Data validation role
- CRO:0000089 Data visualization role
- CRO:0000090 Data aggregation role
- CRO:0000091 Data collection role
- CRO:0000092 Data curation role
- CRO:0000093 Data entry role
- CRO:0000094 Data integration role
- CRO:0000095 Data modeling role
- CRO:0000096 Data quality assurance role
- CRO:0000097 Data standards developer role
- CRO:0000098 Data transformation role
- CRO:0000099 Data validation role
- CRO:0000100 Data visualization role



NISO Research Outputs

- Basic Science
 - CLO:0000031 Cell line
 - NCIT:C14184 Transgenic organism
- Capacity
 - D000071252 Mentoring
 - NCIT:C54131 Company
- Code and Software
 - NCIT:C16275 Algorithm
 - NCIT:C75301 Application Program Interface
- Communications
 - NCIT:C80148 News
 - OMIT:0009470 Mass media
- Data
 - NCIT:C15426 Database
 - NCIT:C47824 Data
- Education and Training Materials
 - NCIT:C15426 Database
 - NCIT:C47824 Data
- Event
 - NCIT:C15426 Database
 - NCIT:C47824 Data
- Image, Diagram, and Video
 - NCIT:C15426 Database
 - NCIT:C47824 Data
- Instrument, Device, and Instrumentation
 - NCIT:C15426 Database
 - NCIT:C47824 Data
- Methodology
 - NCIT:C15426 Database
 - NCIT:C47824 Data
- Publication
 - NCIT:C15426 Database
 - NCIT:C47824 Data
- Regulatory, Compliance, and Legislation
 - NCIT:C15426 Database
 - NCIT:C47824 Data
- Standard
 - NCIT:C15426 Database
 - NCIT:C47824 Data

Contribution Role Ontology

NISO Research Outputs

- Basic Science
 - CLO:0000031 Cell line
 - NCIT:C14184 Transgenic organism
- Capacity
 - D000071252 Mentoring
 - NCIT:C54131 Company
- Code and Software
 - NCIT:C16275 Algorithm
 - NCIT:C75301 Application Program Interface
- Communications
 - NCIT:C80148 News
 - OMIT:0009470 Mass media
- Data
 - NCIT:C15426 Database
 - NCIT:C47824 Data
- Education and Training Materials
 - NCIT:C15426 Database
 - NCIT:C47824 Data

Contribution Role Ontology (CRO) contains classes and relationships. The CRO imports classes from the NISO Research Outputs computational version of the CREDIT

relationships to scholarly outputs, and NISO research outputs³. The top level classifications and the lower level subtypes, with corresponding descriptive list of subtypes available).

Contribution Role Ontology

NISO Alternative Assessment Metrics Project, publications/tp-25-2016-atm-etrics

Join us! Submit tickets! 😊

<https://github.com/data2health/contributor-role-ontology>

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

Join us!

Workshop: An attribution workshop will be held in the spring or summer of 2019 in Chicago, IL, USA

More info: <http://bit.ly/AttributionSignUp>

It takes technology + culture.

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
- Slide credits: Haendel (contributor roles in genetic diagnosis) and Eichmann (acknowledgements project)

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