

Reusable data for biomedicine: A data licensing odyssey

Image: i.pinimg.com

Melissa Haendel

Seth Carbon, Julie McMurry, Robin Champieux,

Letisha Wyatt, Lilly Winfree

RDA September 20th, 2017



@ontowonka

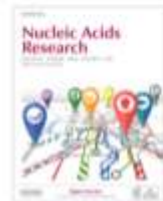
#reusabledata

THERE >1500 PUBLIC BIOMEDICAL DATABASES IN NUCLEIC ACIDS RESEARCH DATABASE COLLECTION

Nucleic Acids Research

Issues Advance Articles Publish Purchase About

Advanced Search



Volume 45, Issue D1
January 2017

Article Contents

Abstract

NEW AND UPDATED DATABASES

UPDATED NAR ONLINE
MOLECULAR BIOLOGY
DATABASE COLLECTION

LOOKING BACK: WHAT HAS

The 24th annual *Nucleic Acids Research* database issue: a look back and upcoming changes

Michael Y. Galperin; Xosé M. Fernández-Suárez; Daniel J. Rigden

Nucleic Acids Res (2017) 45 (D1): D1-D11. DOI: <https://doi.org/10.1093/nar/gkw1188>

Published: 21 December 2016 Article history

A correction has been published: [Nucleic Acids Res gkx021](#).

Views PDF Cite Share Tools

Abstract

This year's Database Issue of *Nucleic Acids Research* contains 152 papers



View Metrics

Email alerts

- New issue alert
- New advance articles alert
- Article activity alert

Receive exclusive offers and updates
from Oxford Academic

<https://doi.org/10.1093/nar/gkw1188>



@ontowonka

OPENNESS IS AN NAR REQUIREMENT, BUT ...

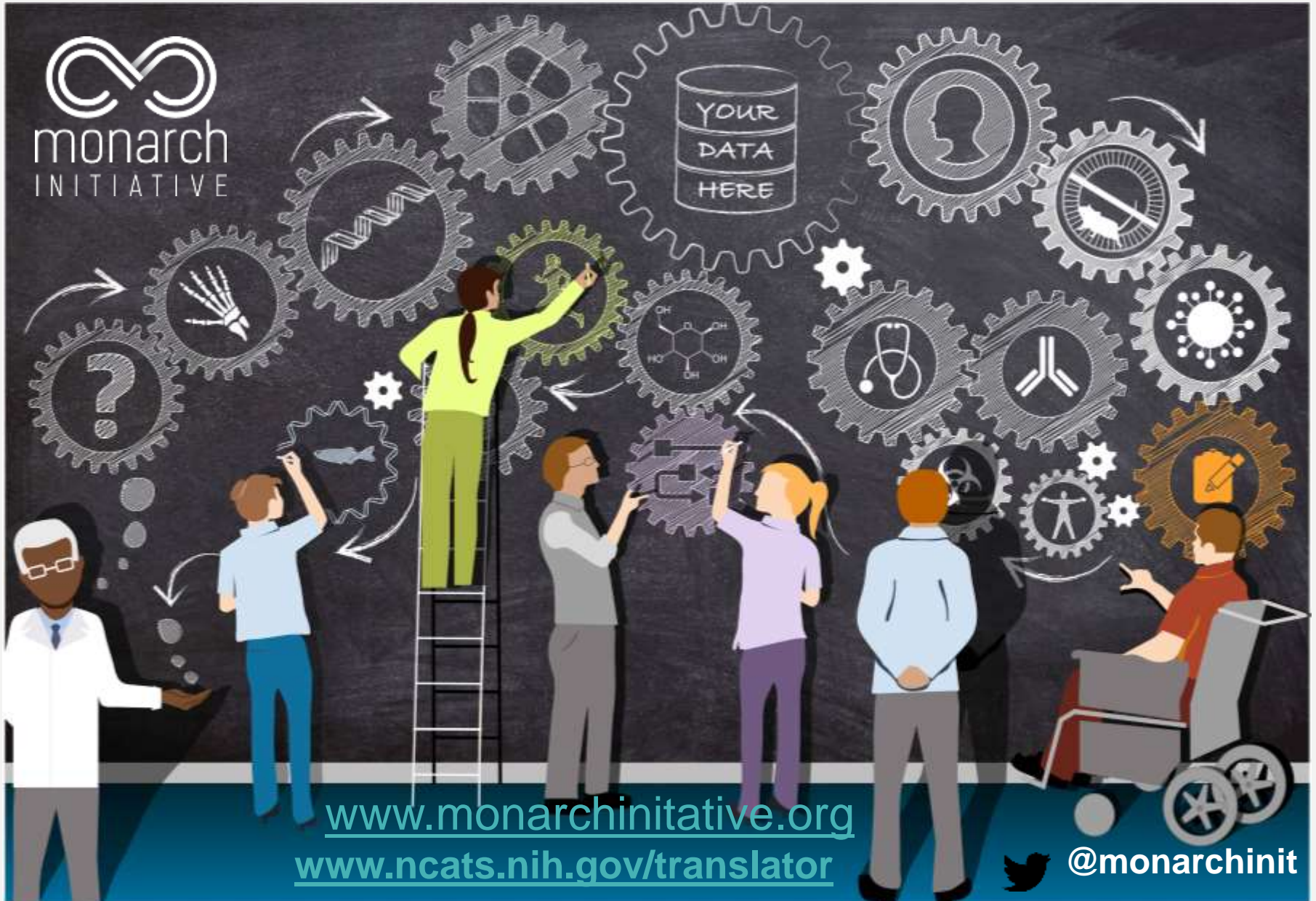


SCHRÖDINGER'S DATABASE IS

OPEN

HOW MANY OF THESE DATA ARE TRULY REUSABLE?

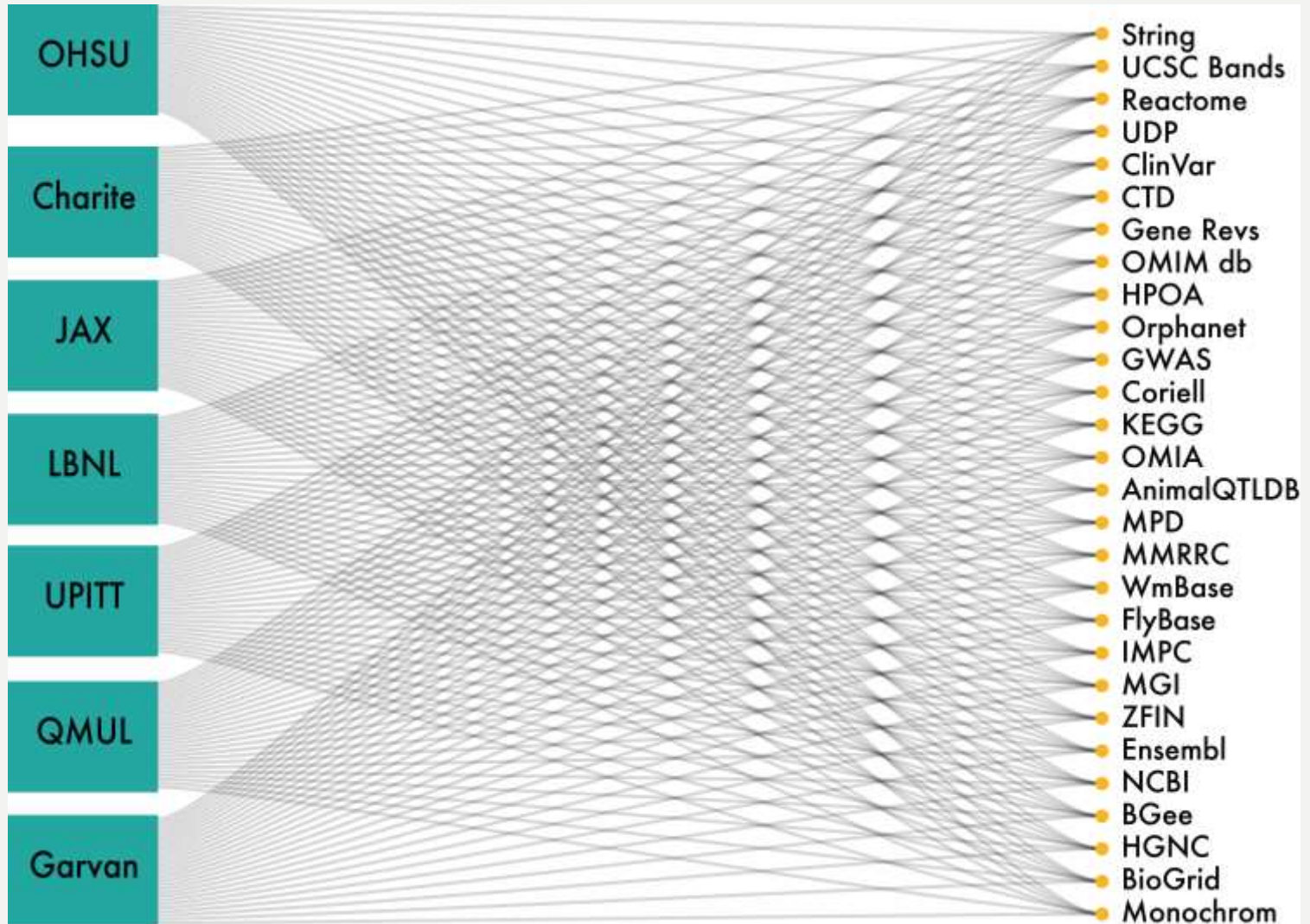
MONARCH & THE NCATS BIOMEDICAL DATA TRANSLATOR



www.monarchinitiative.org
www.ncats.nih.gov/translator

 @monarchinit

MONARCH'S LICENSING BURDEN



Request for Community partnership in data resource licensing planning

04.05.2017, 11:15 by [Melissa Haendel](#), Chris Mungall, [Andrew Su](#), Peter Robinson, Chris Chute, Russ B Altman, Phillip R.O. Payne, Mark Lawler, Tudor I. Oprea, John Willbanks, Subha Srinivasan, Lawrence Hunter, Ida Sim, Sean McDonald, Sean Mooney, Damian Smedley, Emma Ganley, Amye Kenall, Timothy Clark, Carole Goble, [Michel Dumontier](#), [Kristi Holmes](#), Mark Diekans, Adrienne Zell, Casey Overby, Gustavo Glusman, Leigh Carmody, Guoqian Jiang, Monica Munos-Torres, Maureen Hoatlin, Jeremy Goecks, Victor Jongeneel, Joshua Bittker, Jean-Philippe Gourdine, Matthew H. Brush, Richard L. Zhu, Lara Mangravite, Brett Tyler, Mark D Wilkinson, [Michael R. Crusoe](#), Raja Mazumder, Nicholas P. Tatonetti, Peter D'Eustachio, [Nicole Vasilevsky](#), [Julie McMurry](#), [Robin Champieux](#)

We write an open letter to the NIH Data Research Council to initiate a dialog regarding NIH decisions on data use agreements and licenses. We are members of NIH-funded research groups that collect and/or integrate biomedical data from diverse sources for the purpose of advancing diagnosis, prognosis, treatment selection, and mechanistic discovery.

617
views

114
downloads

0
citations



CATEGORIES

- [Bioinformatics](#)
- [Computational Biology](#)
- [Computational Biology](#)
- [Evolutionary Biology](#)
- [Genetics](#)
- [Genomics](#)
- [Medicine](#)
- [Systems Biology](#)

bit.ly/open-letter-licensing
Additional signatories welcome



@ontowonka

REUSABLEDATA.ORG

 (Re)usable Data Project [Learn more](#) ▾ [Evaluation](#) ▾ [About](#) [Contact](#)

Site search



The (Re)usable Data Project

Inspired by the efforts of scientists around the world and the game-changing efforts of projects like the Creative Commons, the Wikipedia Foundation, and the Free Software movement, we hope to engage the larger community in an open and fruitful discussion on issues concerning the use and reuse of scientific data, including the balance of openness and how to make ends meet in an increasingly competitive environment.

If you would like to join our efforts to highlight the use and reuse of data in the sciences, please feel free to contact us on our [tracker](#) or create a pull request against our [repository](#).



[Learn more »](#)

[See our data »](#)

Curate, evaluate, and provide guidance on legal and effective data reuse and

redistribution? Join us



bit.ly/reusabledata-forum

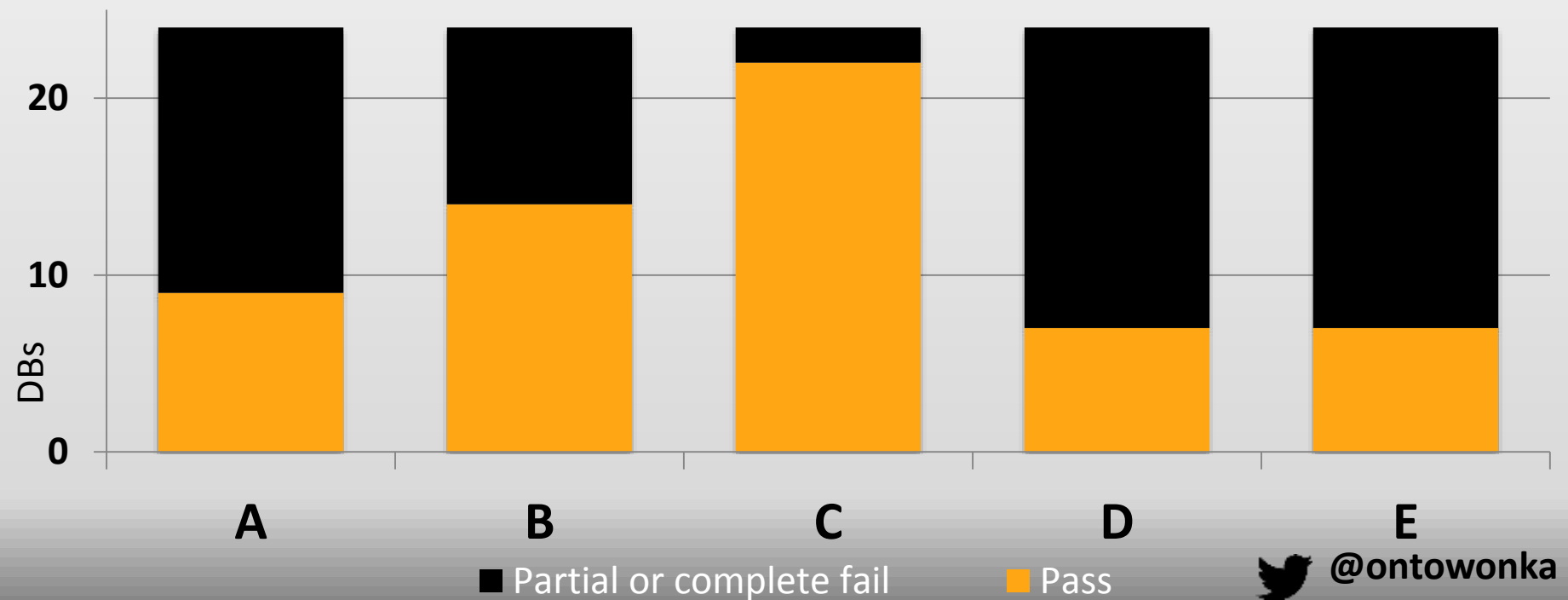


github.com/reusabledata



[@ontowonka](https://twitter.com/ontowonka)

reusabledata.org/criteria



CRITERION A: CLARITY

38% RECEIVE FULL STAR



9/24



Non Standard license
(10/24)

Multiple licenses (3/24)

Missing license (2/24)



@ontowonka

**CRITERION B:
COMPREHENSIVE & FRICTIONLESS
58% RECEIVE FULL STAR**



14/24



Reuse terms not clear 5/24
Doesn't apply to all data 4/24
Can't obtain singly licensed slice 2/24
Auto-fail due to missing/multiple license 3/24

**CRITERION C:
DATA IS ACCESSIBLE
92% RECEIVE FULL STAR**



22/24



No “reasonable good-faith location” or single action 2/24

**CRITERION D:
FEW RESTRICTIONS ON TYPES OF
REUSE: 29% RECEIVE FULL STAR**



7/24



Restrictive but allows academic use
2/24

Restrictive, no academic provisions
12/24

**CRITERION E:
FEW RESTRICTIONS ON TYPES OF
USER: 32% RECEIVE FULL STAR**



7/24



Restrictive but allows academic use 2/24
Restrictive, no academic provisions 10/24
Auto-fail due to missing/multiple license
3/24

BALANCE OF, QUALITY, SUSTAINABILITY, AND (LEGAL) REUSABILITY



OPEN DATA IS FAIR-TLC

bit.ly/fair-tlc



Findable



Accessible



Interoperable



Reusable



Traceable

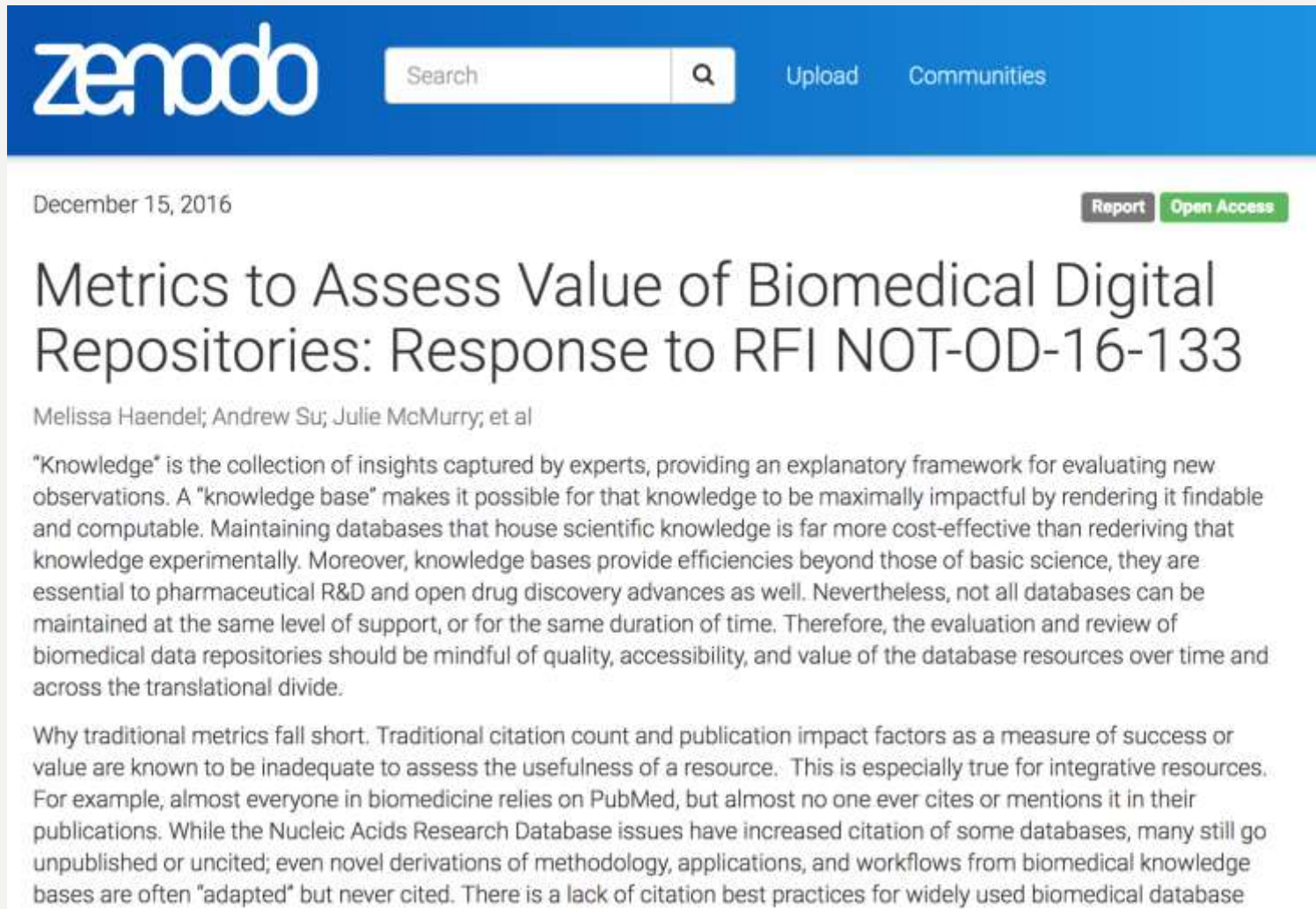


Licensed



Connected

A RUBRIC FOR EVALUATION



The screenshot shows the Zenodo website interface. At the top, there is a blue header with the Zenodo logo on the left, a search bar in the center, and 'Upload' and 'Communities' links on the right. Below the header, the date 'December 15, 2016' is displayed on the left, and 'Report' and 'Open Access' buttons are on the right. The main title of the document is 'Metrics to Assess Value of Biomedical Digital Repositories: Response to RFI NOT-OD-16-133'. Below the title, the authors are listed as 'Melissa Haendel; Andrew Su; Julie McMurry; et al'. The abstract text begins with: "Knowledge" is the collection of insights captured by experts, providing an explanatory framework for evaluating new observations. A "knowledge base" makes it possible for that knowledge to be maximally impactful by rendering it findable and computable. Maintaining databases that house scientific knowledge is far more cost-effective than rederiving that knowledge experimentally. Moreover, knowledge bases provide efficiencies beyond those of basic science, they are essential to pharmaceutical R&D and open drug discovery advances as well. Nevertheless, not all databases can be maintained at the same level of support, or for the same duration of time. Therefore, the evaluation and review of biomedical data repositories should be mindful of quality, accessibility, and value of the database resources over time and across the translational divide. The text continues with a paragraph titled 'Why traditional metrics fall short. Traditional citation count and publication impact factors as a measure of success or value are known to be inadequate to assess the usefulness of a resource. This is especially true for integrative resources. For example, almost everyone in biomedicine relies on PubMed, but almost no one ever cites or mentions it in their publications. While the Nucleic Acids Research Database issues have increased citation of some databases, many still go unpublished or uncited; even novel derivations of methodology, applications, and workflows from biomedical knowledge bases are often "adapted" but never cited. There is a lack of citation best practices for widely used biomedical database

bit.ly/fair-tlc

 @ontowonka

Nominations due September 30, 2017

[MISSION](#) [APPLY](#) [PRIZE & SUPPORTERS](#) [AWARD RECIPIENTS](#) [COI RULES](#)

THE PARASITE AWARDS

Celebrating rigorous secondary data analysis

[Tweet](#)

Researchparasite.org



THANKS TO:

SETH CARBON

JULIE MCMURRY

ROBIN CHAMPIEUX

LETISHA WYATT

LILLY WINFREE

ANDREW SU

CASEY GREENE

JOHN WILBANKS

SEAN MCDONALD

CHRIS AUSTIN

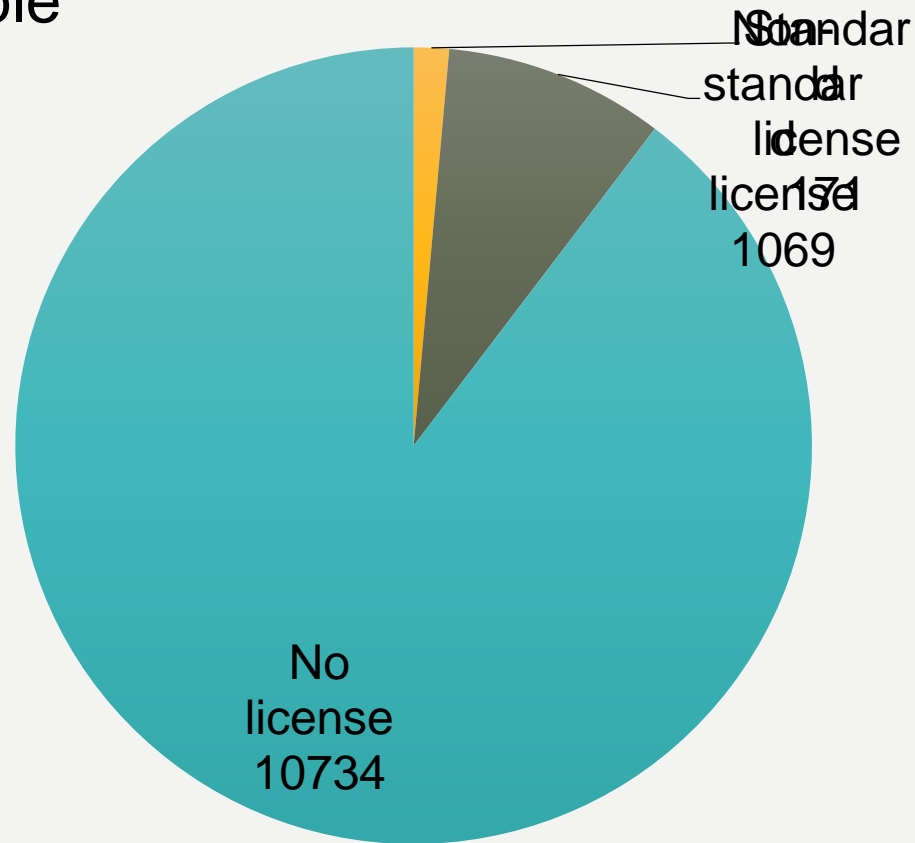
NOEL SOUTHALL

CHRISTINE COLVIS



FAIR-TLC: LICENSURE

Not all data resources are free to use, derive, and redistribute, even if publicly funded and publicly available



<http://peterdesmet.com/posts/analyzing-gbif-data-licenses.html>

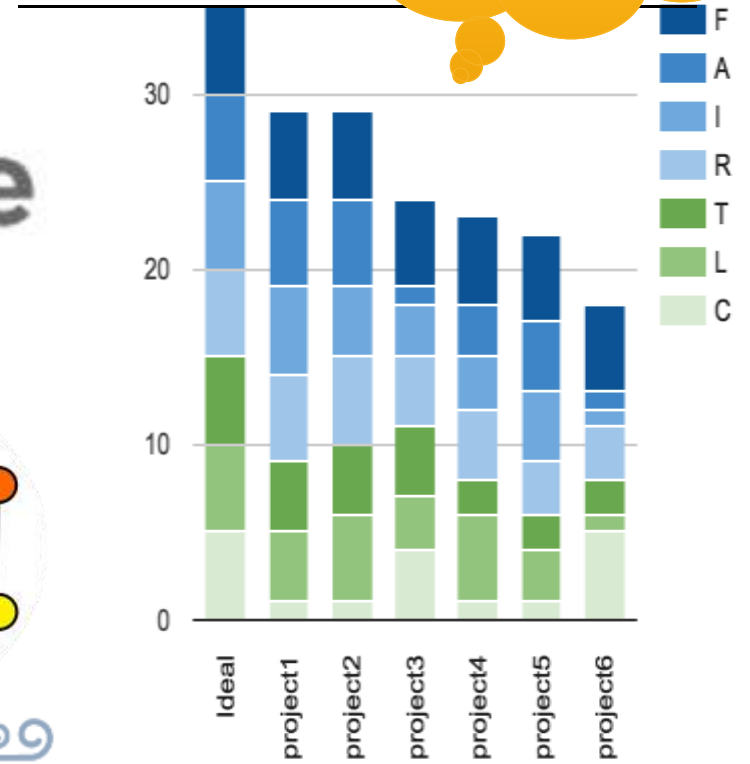
@ontowonka

FAIR-TLC EVALUATION OF THE OPEN SCIENCE CANDIDATES

Room for improvement



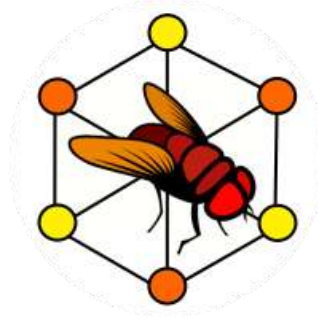
The Open Science Prize



nextstrain



Open imaging



MyGene²

OpenTrialsFDA
An OpenTrials App

openaq

bit.ly/open-science-prize

OVERVIEW OF 22 DATABASES



22 sources

2.7 out of 5 stars

