FAIR Phenotyping with APHRODITE

Juan M. Banda[1] , Andrew Williams[2], Mehr Kashyap[3], Martin G. Seneviratne[3], Aaron Potvien[4], Jon Duke[4], Nigam H. Shah[3]

[1]Department of Computer Science, Georgia State University, Atlanta GA 30303

[2]Tufts Medical Center, Boston MA 02111

[3]Center for Biomedical Informatics Research, Stanford University, Stanford CA 94305

[4]Georgia Tech Research Institute, Atlanta GA 30308

Georgia State University

The Need

- The common failure to reproduce published results has created an atmosphere of crisis even in disciplines where precise measurement and tight experimental control are the norm
- There is even more reason for vigilance in disciplines that must manage lower degrees of measurement accuracy and experimental control
- One response to this crisis has been the emergence of open science principles that publicly expose the process of defining hypotheses, data selection and development, study design and analytic choices

The solution

SCIENTIFIC DATA

Comment | OPEN | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

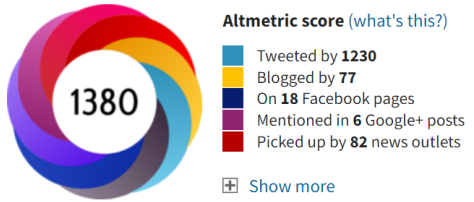
Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons ✉ - Show fewer authors

Scientific Data **3**, Article number: 160018 (2016) | Download Citation ⭳

Last updated: Mon, 9 Sep 2019 14:18:07 GMT

Total citations

Online attention



This Altmetric score means that the article is:

- in the 99th percentile (ranked 76th) of the 264,573 tracked articles of a similar age in all journals
- in the 1st percentile (ranked 1st) of the 1 tracked articles of a similar age in *Scientific Data*

Georgia State University

Rapid adoption of principles:

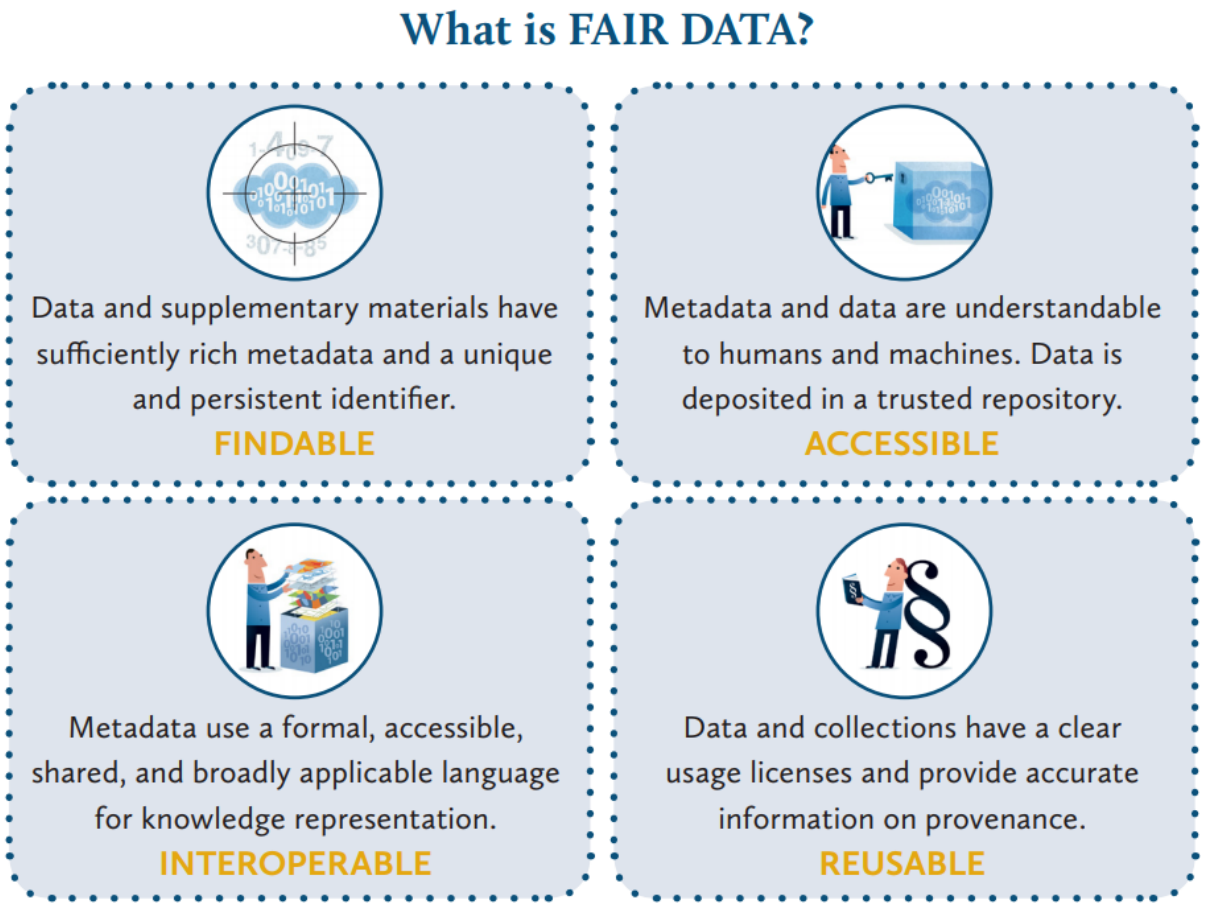
Progress towards the European Open Science Cloud

News item | 01-12-2017 | 18:00

Germany and the Netherlands establish office for the GO FAIR Initiative / France joins

What does it mean to be FAIR?

- **FAIR**
- Findable,
- Accessible,
- Interoperable,
- Reusable.



https://libereurope.eu/wp-content/uploads/2017/12/LIBER-FAIR-Data.pdf

What are we proposing?

Anatomy of an APHRODITE FAIR phenotype definition

A phenotype definition will be **Findable**

- To address the need to have a persistent global unique resource identifier (URI) for each phenotype definition version, we have utilized GitHub unique commit hash value to identify each individual phenotype definition version

- The OHDSI Gold Standard Phenotype Library workgroup has defined and created an additional abstraction layer over the phenotype definitions available as a R Shiny App

A phenotype definition will be **Accessible**

- The phenotype definition, generation script, and metadata will be retrievable by their identifier using any regular web browser or the application layer of the phenotype library

- By using a publicly and freely available resource such as GitHub, we offer better accessibility than placing the definitions on an institutional server

A phenotype definition will be **Interoperable**

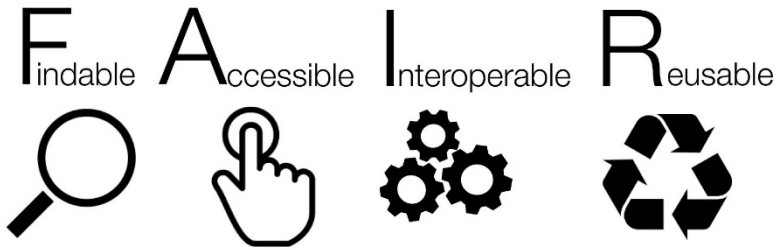
- We will leverage the OMOP CDM and associated vocabularies to solve the major obstacle to interoperability across sites. Our phenotype definitions' metadata will use JSON for knowledge representation and ease of machine readability

- When developing phenotyping definitions based on prior publications, or when a publication is generated from a definition generated from our pipeline, we will include all proper URI's to the publications in question

A phenotype definition will be **Re-usable**

- Currently APHRODITE definitions are easily shareable and re-usable for other sites. We have added meta-data elements related to software, CDM, and vocabulary versions, as well as a plurality of accurate and relevant attributes to guarantee re-usability
- All the publicly available phenotypes will be released under relevant open source licenses, details of which will be attached to the definition's meta-data
- Site and researcher information will be recorded as well as relevant publications in allowing fully traceable provenance for each definition

Questions?



Improving the FAIRness of digital resources will increase their quality and their potential for reuse

@micheldumontier::RDA:2018-01-31

Want to help? reach out: @drjmbanda or jbanda@gsu.edu