

RDA

13th

Plenary Meeting

Philadelphia, Pennsylvania USA

2-4 April 2019

Organized by RDA - United States

University

Greetings and summary from 13th Plenary Meeting in Philadelphia

Ville Tenhunen

University of Helsinki

Center for Information
Technology



"With Data Comes Responsibility"

- 448 registrants from > 30 countries
- 60 breakout sessions
- 288000 Twitter views

"Responsible data science is our new frontier!"

- Biased data reflects the world: "When data is about people, bias can lead to discrimination"
- Data science is algorithmic, therefore it cannot be biased! And yet...
 - All traditional evils of discrimination...
 - Transparency helps
 - Technology alone won't do: also need regulation and civic engagement



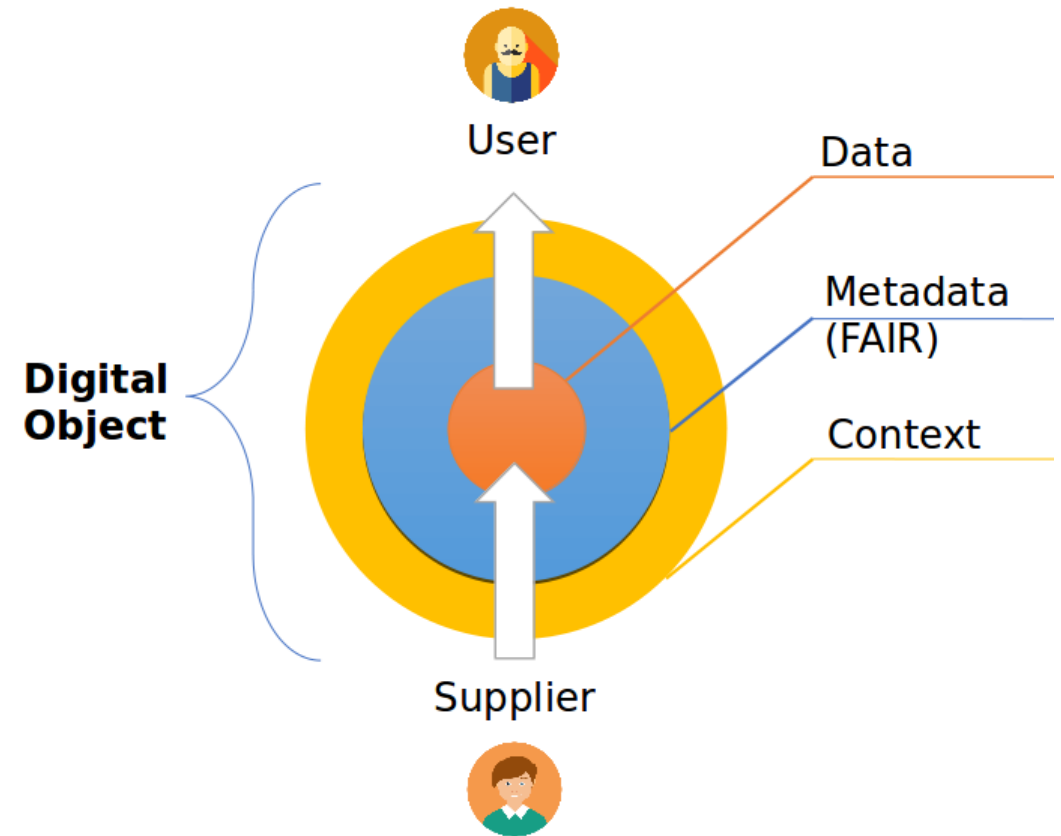
"Anonymisation is myth"

- Hanieh Razzaghi et.al.:The Lessidentify Application For Privacy Risk Reduction and FAIR Clinical Data Sharing
 - PEDSnet::Lessidentify
- Evanette K Burrows et.al.: Use of Electronic Health Record(EHR) data for research purpose: challenges and solutions
 - "Apply some de-identification rules programmatically"
 - "Techniques using rule-based natural language processing will be evaluated later"



Digital objects and automation

- Dimitris Koureas:
 - "Trust lost when datasets disconnect from: context in which they were created, or communities who created them."
- George O. Strawn:
 - "Interoperable data will become a core infrastructure, like the Internet has already become"
 - "We must automate activities that currently require much manual intervention"



Role of AI in research data management (along the data life cycle)

- creation/collection
 - metadata generation – from manual to AI
 - data quality, cleaning
 - steering the data production, e.g. in large simulations
 - automatic translation of the metadata in other languages (than english)
 - classifying data for better understanding, add context, data annotation
 - data discovery – hidden data in articles, text files
- processing
- analyzing
- archiving
- publishing
 - nlp for dissemination and explanation of data
 - translating data into natural language
- reuse
 - enhanced search
- And some more:
 - spoken language as data – automatic translation, heritage
 - data discovery – hidden data in articles, text files
 - analysis of research papers

https://docs.google.com/document/d/15posmhUFakMyTYzWDxSf6m_6hOiOAFxz8KrZ55Epz2E/edit#heading=h.yrp6lpe78sq3



FAIR

is everywhere



Thank you!

ville.tenhunen@helsinki.fi

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

