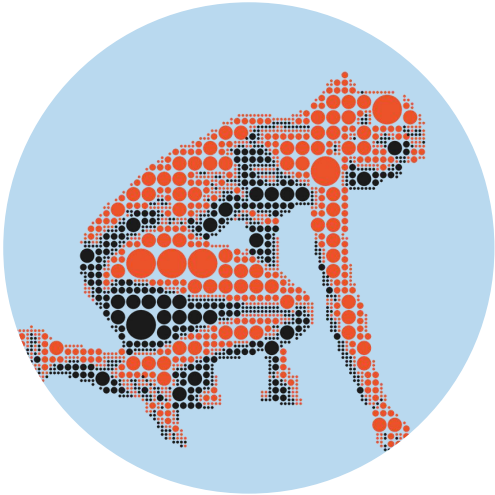


# FAIR Data Sprint

Data Science Week 2020, February 4th

Erik van den Bergh



# Programme today

## Morning

- 10.00 Today's activities and introduction to FAIR data
- 10.30 Cleaning data (example data)
- 11.00 Coffee
- 11.15 Further data refining with OpenRefine
- 12.30 Lunch

## Afternoon

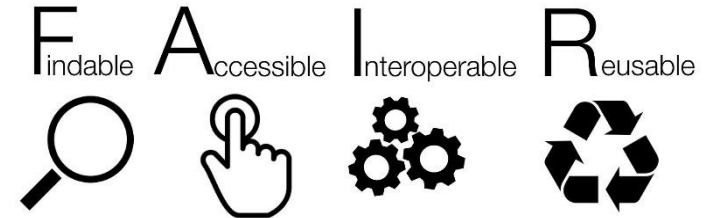
- 13.30 Introduction to metadata
- 14.00 Metadata exercises
- 15.00 Energizer with SCB Bongerd
- 15.15 FAIR data in practice presentation by Rob Lokers
- 16.00 Drinks in central hall
- 16.30 Uploading data tot repositories
- 18.30 Pizza
- Uploading data to repositories
- 20.30 conclusions & end

# FAIR data introduced by Barend Mons

- <https://www.youtube.com/watch?v=-le69A3yanw>

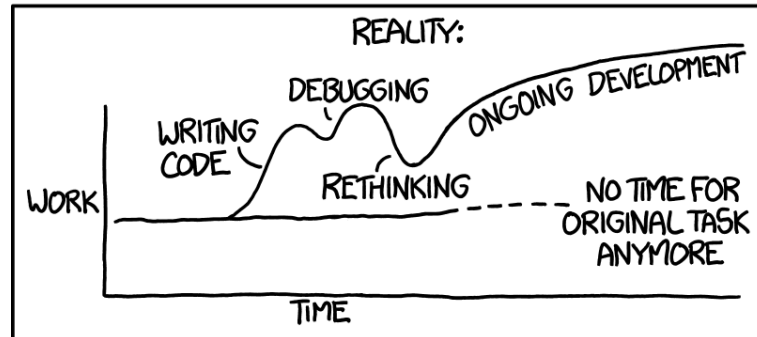
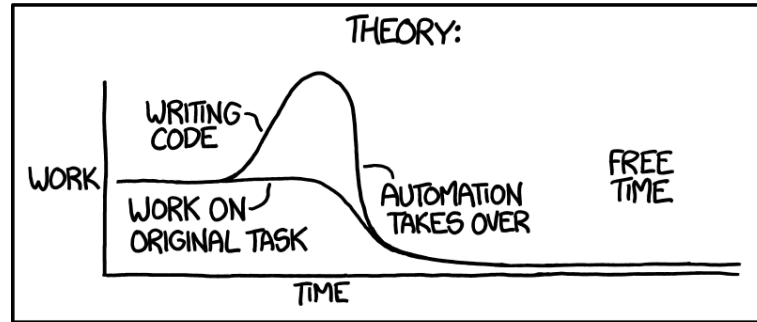
# FAIR principles

- FAIR principles are the internationally recognized standards that data should have to remain relevant
- It allows maximum reusability, now and in the future
- It allows interaction by *agents*, so computers can help analyse

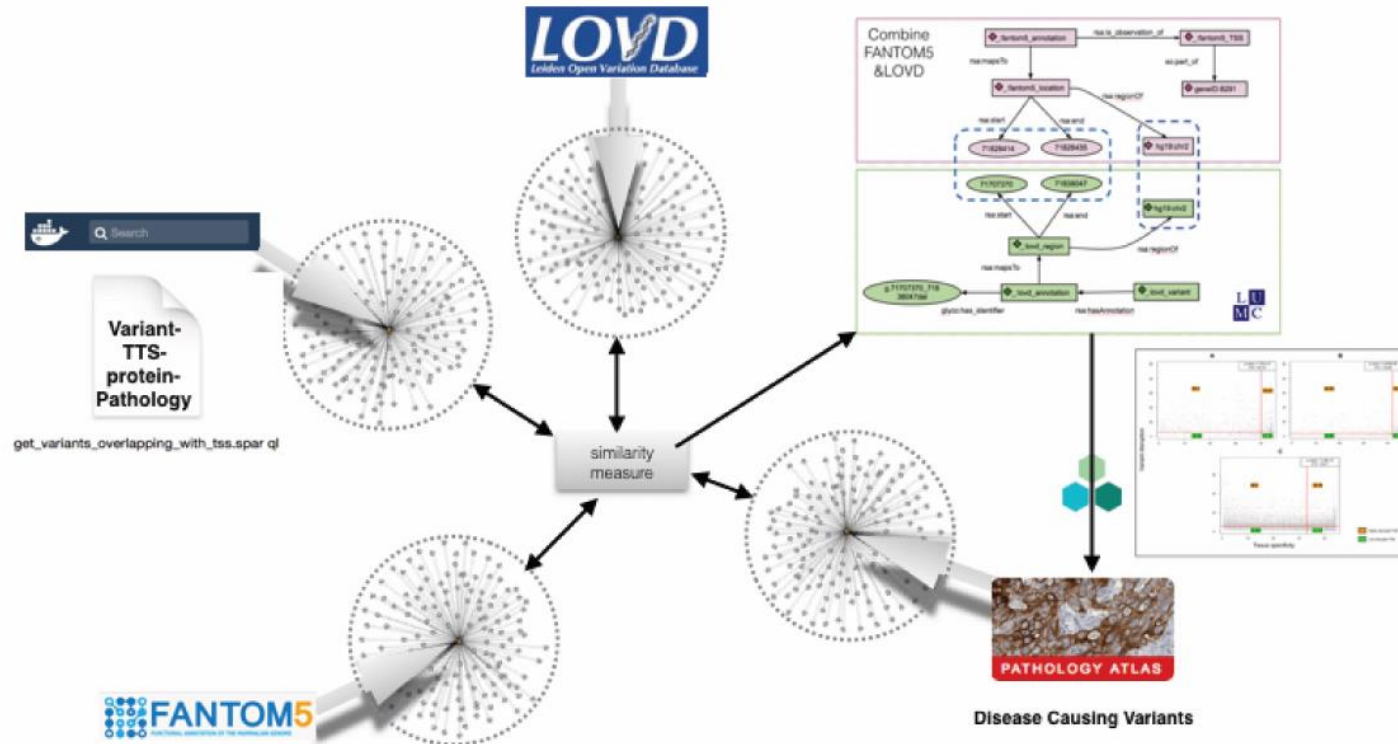


# Automated science?

"I SPEND A LOT OF TIME ON THIS TASK.  
I SHOULD WRITE A PROGRAM AUTOMATING IT!"



# FAIR science automation



# FAIR is not Open!

- Accessible = accessible under well defined conditions
  - Open can be defined by the proper license
- FAIR is about machine readability, not about open access

WordPad file on my computer  
with CC license

**100% Open**  
**0% FAIR**



Annotated genome file in data  
system with patient data under  
access conditions

**0% Open**  
**100% FAIR**



# FAIR in practice





# Today



# Step 1: data cleaning

- Making data computer readable
- Reconciling data entry mistakes
- Often only needed with human generated data:
  - Manually collected
- Or non calibrated machines



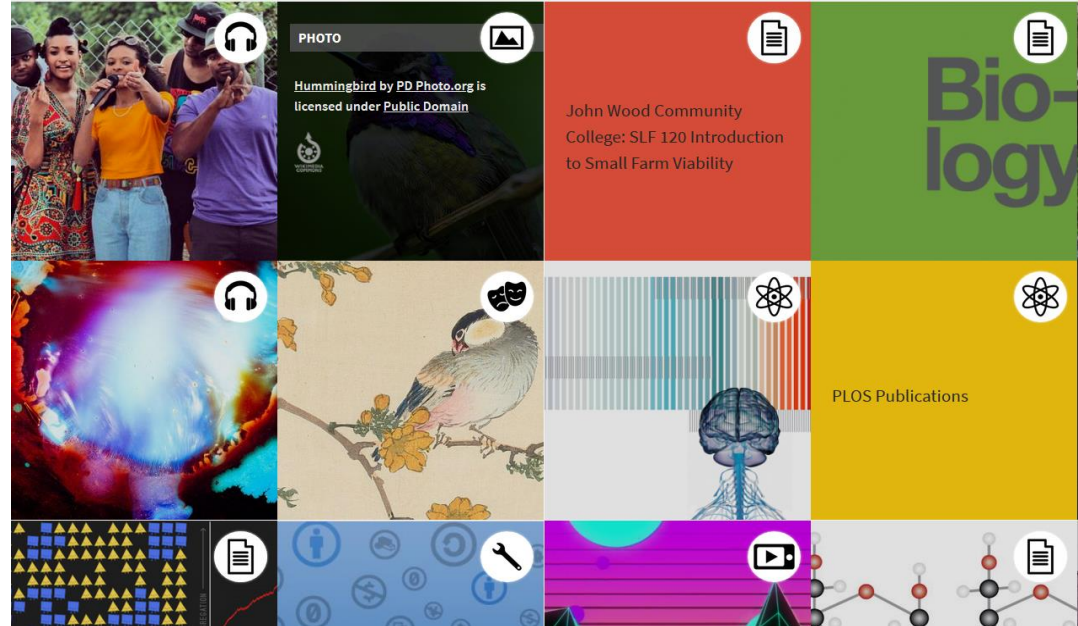
# Step 2: Linking data

- Annotating measurements with linked data concepts
- Helps with findability and metadata standardisation
- Improves interoperability for the future



# Step 3: Uploading data

- Depositing the data to a secure, sustainable system
  - FAIR  $\neq$  Open
- Does not need to be public
- Makes data findable, accessible
- Attaching a proper license improves reusability



# Let's sprint!

