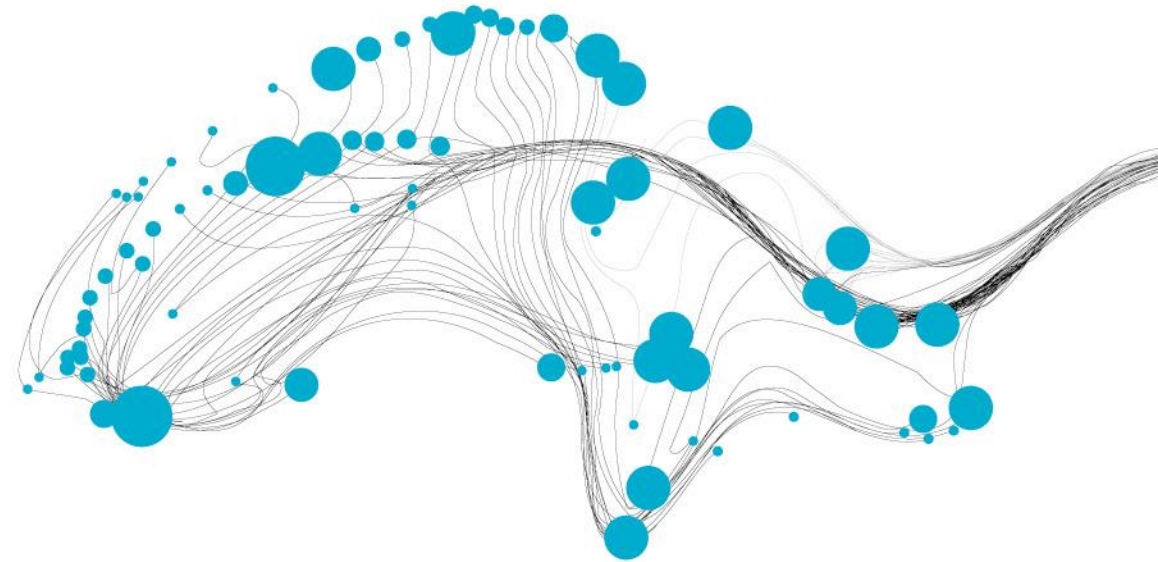


FAIRIFICATION PROCESS ACTIVITY TRACKER DATASET

VICTOR ALECU & VLAD GOSA



FITNESS TRACKER DATASET

DESCRIPTION

- .csv formatted synthetic fitness data.
- Sourced from Kaggle (CC0 license)
- 12 columns (ints, doubles, strings, dates)
- 1 million rows (reduced to 10k)
- Metadata on column descriptions, provenance, license, use cases

REUSE STAKEHOLDERS

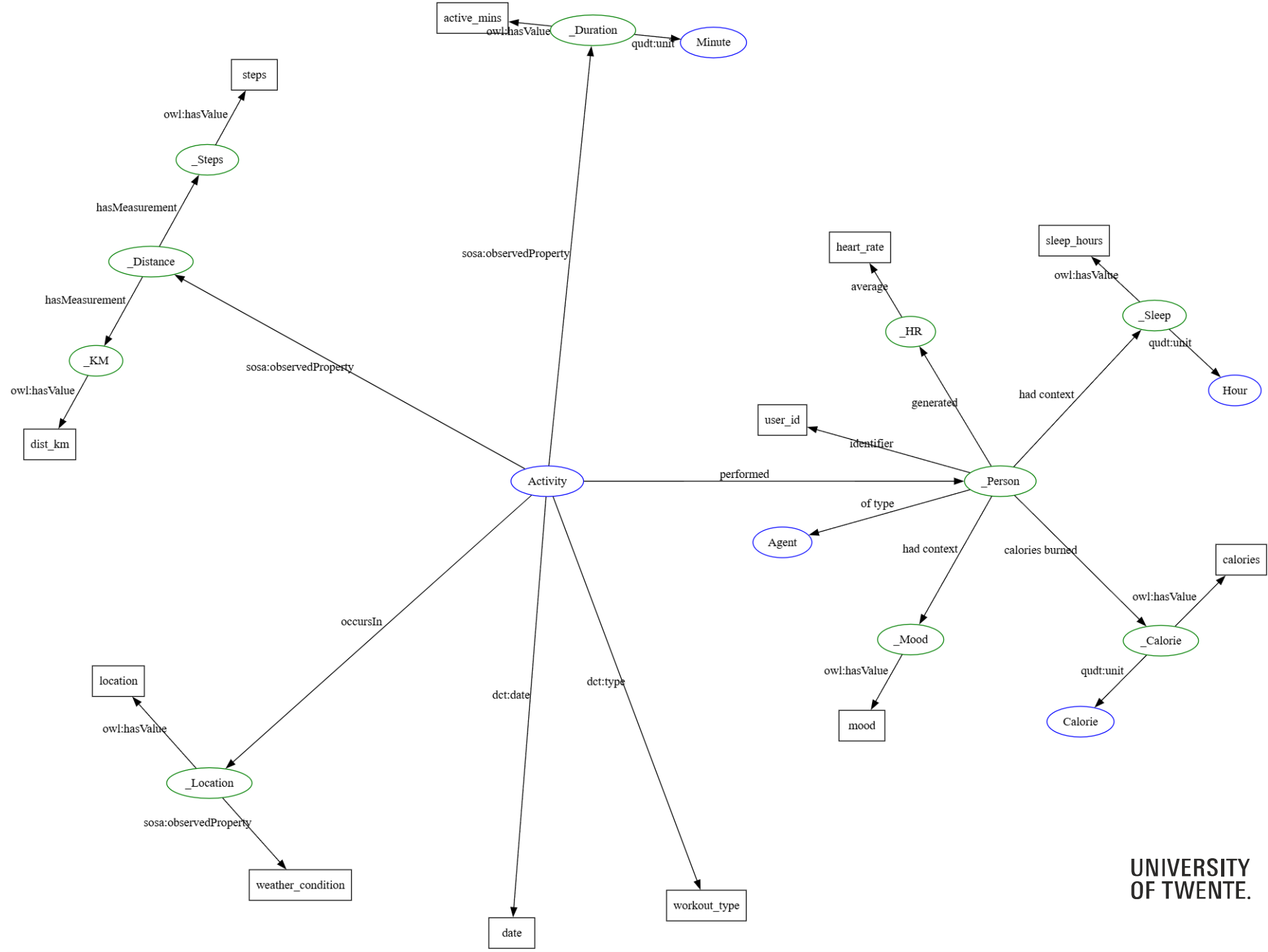
- Researchers
- Companies (marketing departments)
- Students
- Semantic Web users

FAIR ISSUES

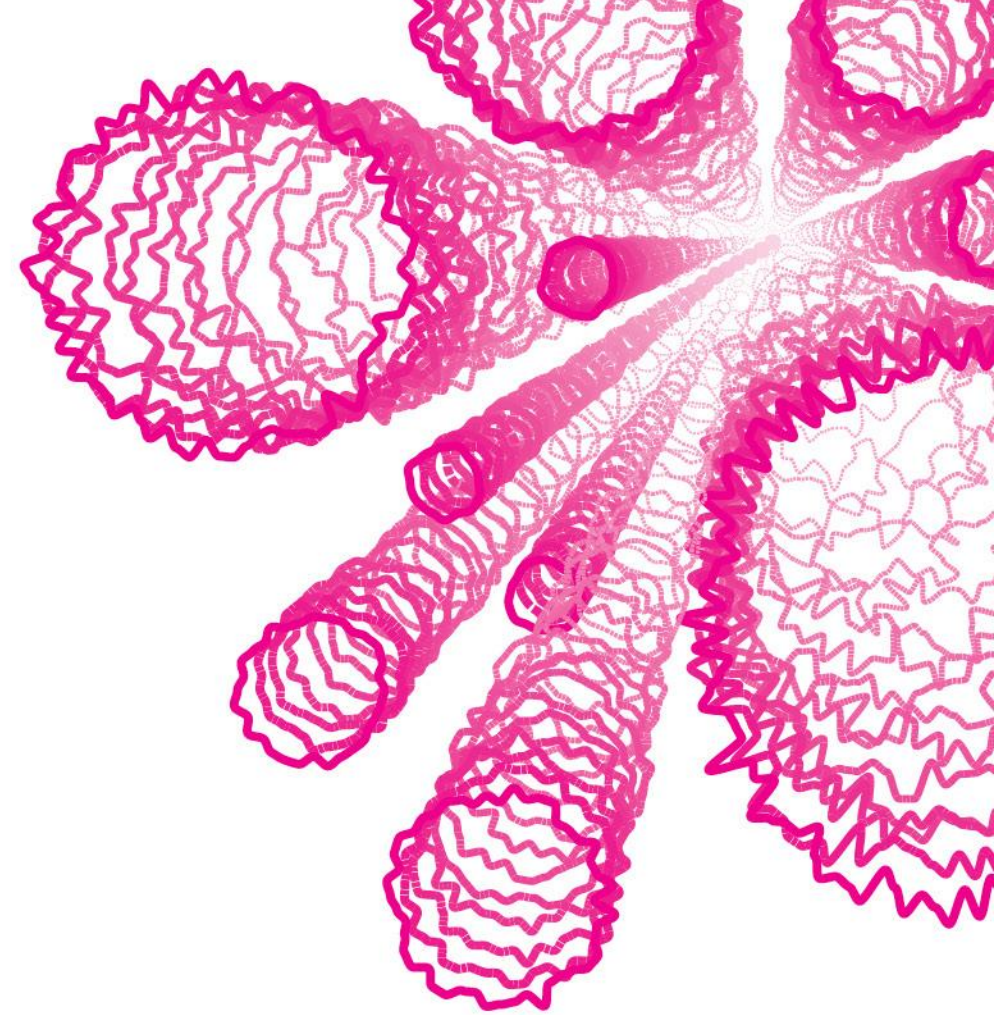
- **F1:** Non-persistent identifier
- **F2:** No rich metadata
- **F3:** No data identifier in metadata
- **A2:** Metadata doesn't have longer persistency than data
- **I2:** Metadata represented in a semi-structured way
- **I2:** (Meta)data does not use vocabularies following the FAIR principles
- **I3:** (Meta)data does not include qualified references to other (meta)data. Only Kaggle recommendations



SEMANTIC DATA MODEL



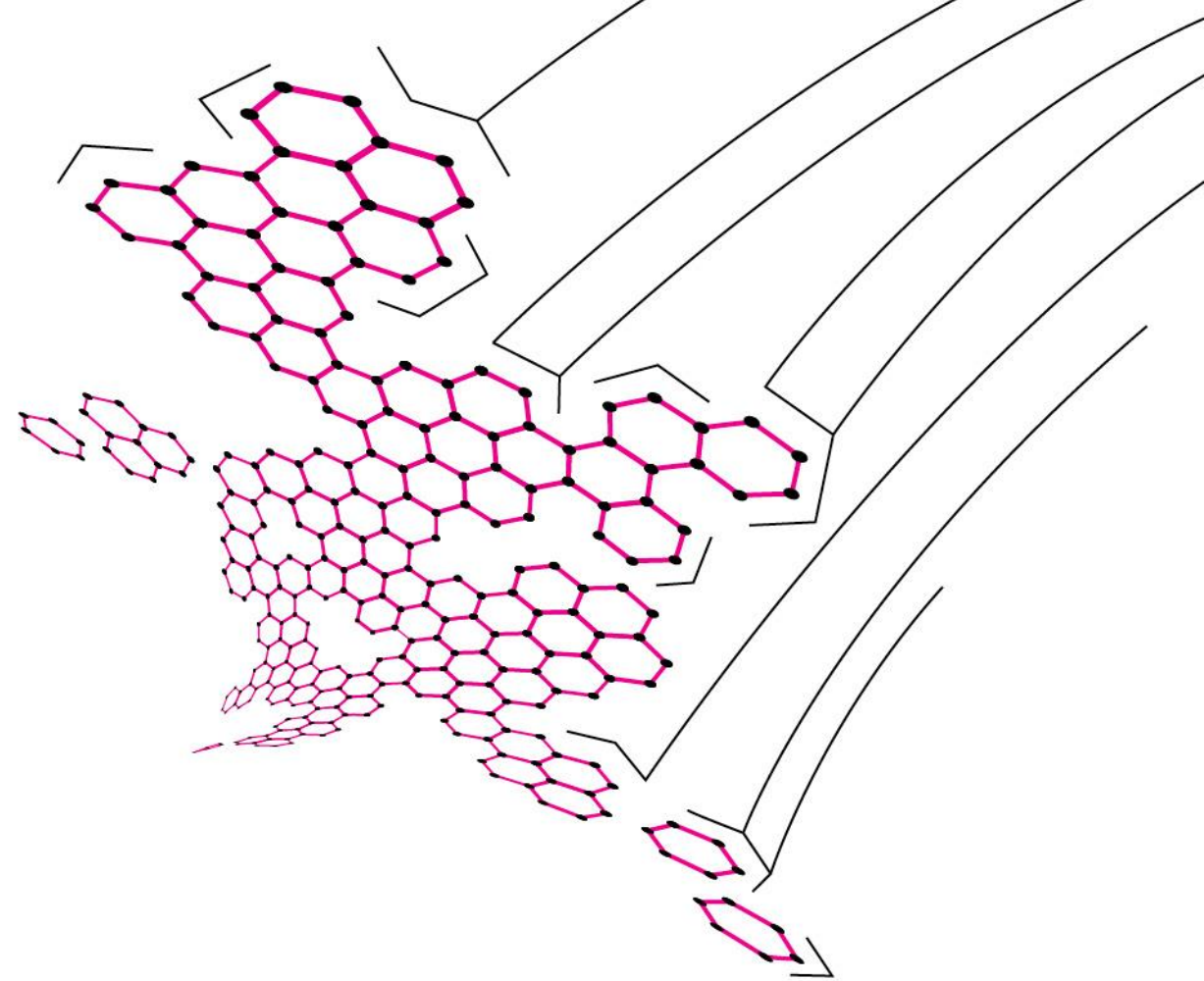
METADATA PROPERTIES BEYOND FDP SCHEMA



- **Title**
- **Description**
 - Plaintext – enhances human readability
- **Source (dct:source)**
 - Location from which the original dataset was obtained (in our case Kaggle)

IMPROVEMENT IN FAIRNESS

- **F1:** Globally unique and persistent identifier through w3id.org/permalid
- **F2:** Rich metadata
- **F3:** Data identifier with clear predicate in metadata
- **F4:** Can now also be accessed through w3id.org/permalid GitHub repo (not only Google)
- **I1:** (Meta)data is now triplified, has public schema, and uses text/turtle IANA media type
- **I2:** (Meta)data is linked to resources which follow the FAIR principles
- **I3:** (Meta)data points to qualified references of resources which follow the FAIR principles





**THANK YOU FOR YOUR
ATTENTION!**