Semantic Interoperability of Metadata

Oya Beyan
Fraunhofer Institute for Applied Information Technology (FIT)
RWTH Aachen University, Informatik 5, Germany
Learning from Data

- quality of care
- patient recruitment
- pragmatic trials
- patient safety
- comparative effectiveness
- observational studies
- pharmaceutical vigilance
- health utilization
- safety surveillance
- research

Government
Industry
Hospital
Private Practice
Emergency Response

Oya Beyan, GO FAIR Personal Health Train IN, GO BUILD Workshop, Hamburg, 11.11.2019
… reality
Sharing Analytics Tasks
PHT Framework

- Persistently identifiable, auditable trains
- Traceability and provenance
- Reproducibility
- Standardized interfaces

Oya Beyan, GO FAIR Personal Health Train IN, GO BUILD Workshop, Hamburg, 11.11.2019
I designed a great algorithm to predict heart diseases but need data to train it.

Are there any patients aged 65+ who use beta-blockers in combination with verapamil?

What is the average weight of patients diagnosed with cardiovascular diseases?

PHT Approach

Data Consumers

- Intends to access privacy-sensitive data from multiple curators
- Formulates queries to describe data requirements
- Specifies the analytics algorithm
- Sends request to execute them

Tools & Services

- Metadata Discovery of Data
- Model/Train Repositories
- Train Metadata Schema and Annotation
- Broker Services
### PHT Framework

#### Data Stations

- Provides computing resources
- Integrates data from multiple data sources
- Maintains private data repositories
- Publishes schemas/metadata describing available data sets and access protocols
- Executes analytic tasks in secure enclave
- Communicates with other components
- Ideally: Acts as FAIR data points
PHT Framework

Trains
A collection of Digital Objects and workflows
Adaptable, reproducible, identifiable, versioned, self containing

Algorithm.py

```python
import ml_lib as ml

def main():
    model = ml.load('/model1')
    model.fit_transform('/data')
    model.save('/model2')
```

query.{cql / sparql}
A train encapsulates all digital objects during the lifecycle of execution, including results.
PHT Framework

Handling Station

- Acts as a gateway between consumer and data curators
  - Broker between parties
  - Evaluates & monitors trains
  - Facilitates train routing
  - Registry & Indexing
  - Maintains persistency
  - Aggregates results

- Each process in a PHT framework is described by rich metadata and communicated between entities by following community standards
Applying FAIR Principles to PHT

STATIONS:
• Discoverable to data consumers
• Publish data standards and computational resources
• Defined access policies
• Standardized interfaced to interact with trains

TRAINS:
• Find and Reuse Trains
• Interpretable content and requirements
• Repeats executed Trains
• Guarantees access to the same Train over the lifecycle

Oya Beyan, GO FAIR Personal Health Train IN, GO BUILD Workshop, Hamburg, 11.11.2019
Main components of the PHT architecture

TRAIN as a FAIR Digital Object

Core Data Model for Digital Objects

- a bit sequence being stored, managed and served by some repositories,
- is referenced by a globally resolvable and persistent identifier, and
- is described by metadata

Ref: RDA’s Data Foundation & Terminology Group (DFT) 2014: Core Model
TRAIN as a FAIR Digital Object

FAIR Digital Objects

Ref: Luiz Bonino, Fair digital framework FAIR digital object + LDP
Train is defined by a persistent identifier, provides links to digital resources (including executed code and generated outcomes) and has a metadata generated before and during the execution process.
TRAIN as a FAIR Digital Object

Oya Beyan, GO FAIR Personal Health Train IN, GO BUILD Workshop, Hamburg, 11.11.2019
Train Registries & Repositories

Train Registries & Repositories

1. Develop Research Code

2. Describe it
   - Name
   - Type
   - Description
   - Creator
   - Version
   - Format
   - Others...

3. Persistently Identify and Save into Repository
   - Persistent
   - Unique

4. Assign a DOI
   - DOI: 10.1234/exampledata

5. Find
   - Access
   - Interoperate within services
   - Reuse
Train Registries & Repositories

https://www.train4fair.com/
Train Registries & Repositories
Train Registries & Repositories
Train Registries & Repositories

DataCite Fabrica Test
Fraunhofer FIT / DOIs

10.20408/k3ck-xd46

URL

Metadata

BMI Calculation Workflow
Joao Bosco Jares MSc.
Workflow published 2019 via DataCite

https://handle.test.datacite.org/10.20408/k3ck-xd46

Citation

Joao Bosco Jares MSc. BMI Calculation. DataCite. https://doi.org/10.20408/K3CK-XD46

Oya Beyan, GO FAIR Personal Health Train IN, GO BUILD Workshop, Hamburg, 11.11.2019
The Train4FAIR: Experiment Landing Page

An auto generated by the Train Platform Project to give you a brief introduction about the project as well as make it easy to find information and resources related to each new Train experiment. (You can find below the information regarding the experiment)

**Experiment Info**

- **Experiment Name:** UC03 - Check Rare Diseases Train
- **Experiment Description:** Train to Check Rare Diseases Train
- **Experiment Publish Version:** 1.0
- **Experiment Publish Year:** 2019
- **Experiment Publisher:** DataCita

**Experiment Metadata:**
- [http://134.202.64.5:6001/repositoryService/train:5dc41871a7ba1ed6f32655ef2](http://134.202.64.5:6001/repositoryService/train:5dc41871a7ba1ed6f32655ef2)
- [http://134.202.64.5:6001/repositoryService/train:5dc41871a7ba1ed6f32655ef2](http://134.202.64.5:6001/repositoryService/train:5dc41871a7ba1ed6f32655ef2)
- [http://train-platform.de:12008/exflow/b01f829b-af70](http://train-platform.de:12008/exflow/b01f829b-af70)
- **Experiment Source Repository:** Fraunhofer
- **Experiment Language:** EN, ENGLISH
- **Experiment Metadata Format:** application/json
- **Experiment Rights:** CC0 1.0 Universal
- **Experiment Rights URI:** [http://creativecommons.org/publicdomain/zero/1.0/](http://creativecommons.org/publicdomain/zero/1.0/)

**Author Info**

- **Author Name:** yuliu
- **Author Affiliation:** Fraunhofer FIT
- **Author Name Identifier:** [https://orcid.org/0000-0001-6566-2802](https://orcid.org/0000-0001-6566-2802)

**DOI Info**

- **DOI Prefix:** 10.20408
GO FAIR PHT Implementation Network

We are working together

.......... be part of it!

Please contact us:

Oya Beyan
beyan@fit.fraunhofer.de

Fraunhofer Institute for Applied Technology FIT, Schloss Birlinghoven, Konrad-Adenauer-Straße, 53754 Sankt Augustin, Germany

https://www.go-fair.org/implementation-networks/overview/personal-health-train/