BETTER
BIG DATA EARTH OBSERVATION TECHNOLOGY

Semantic Analytics In BETTER

Dr. Simon Scerri (Fraunhofer IAIS)
Mohamed Nadjib Mami (Fraunhofer IAIS)

BETTER Hackathon
08th November 2019
Frascati, Rome

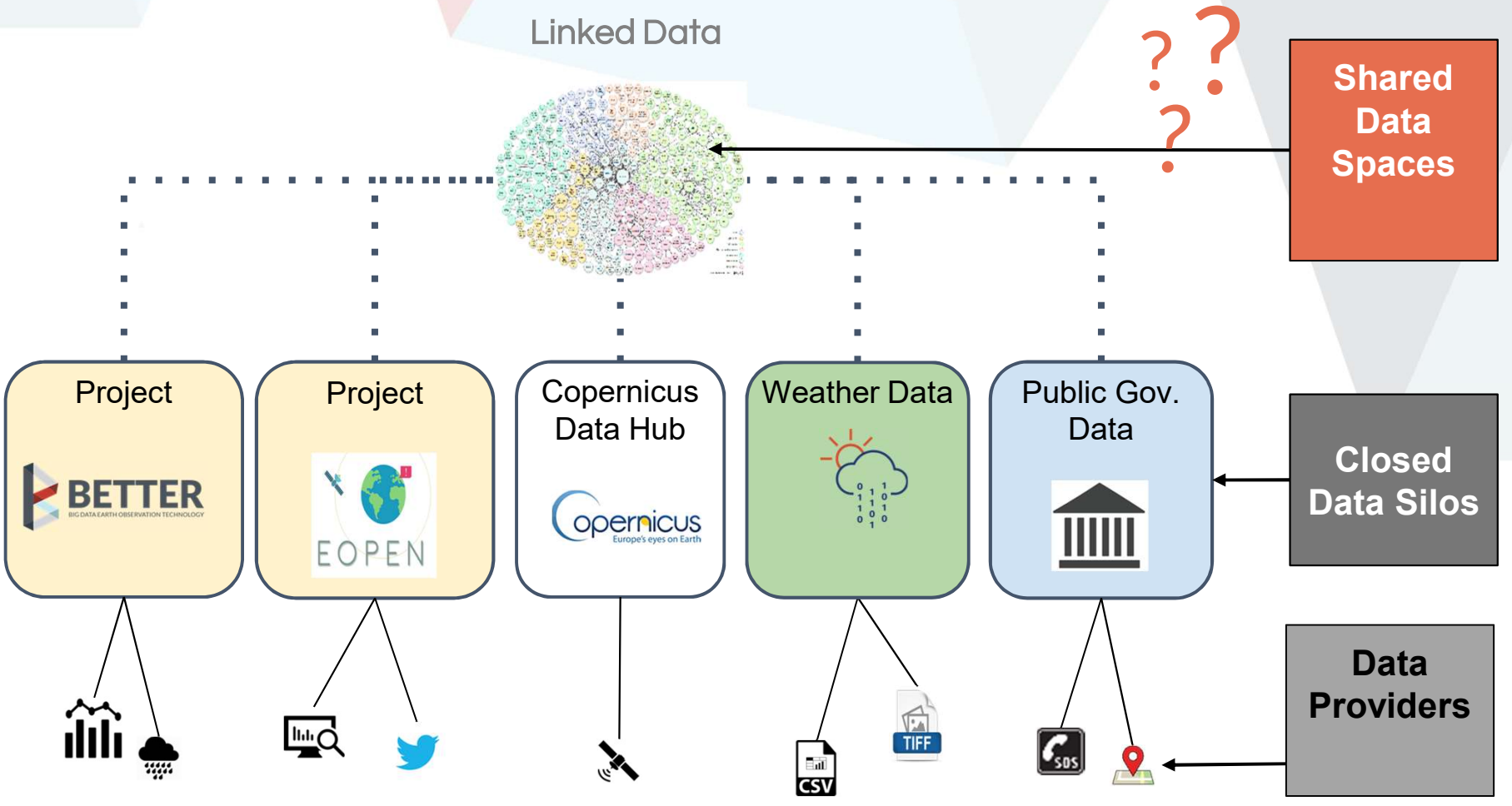
VIEW these slides here
https://bit.ly/better-semantic-analysis

Download material here
https://bit.ly/better-semantic-material

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no 776280

deimos elecnor group | TERRADUE | Fraunhofer IAIS | SatCen | WFP | ETH zürich

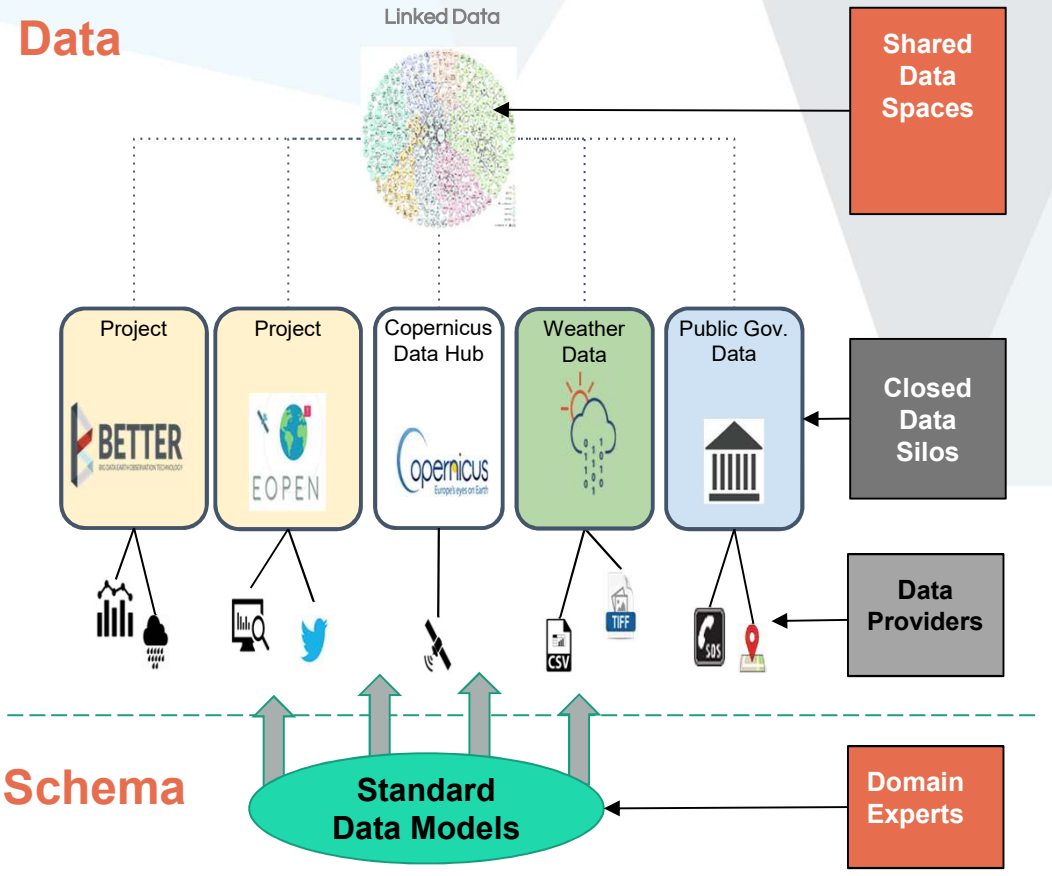
Data Heterogeneity: Limiting the Potential!



VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

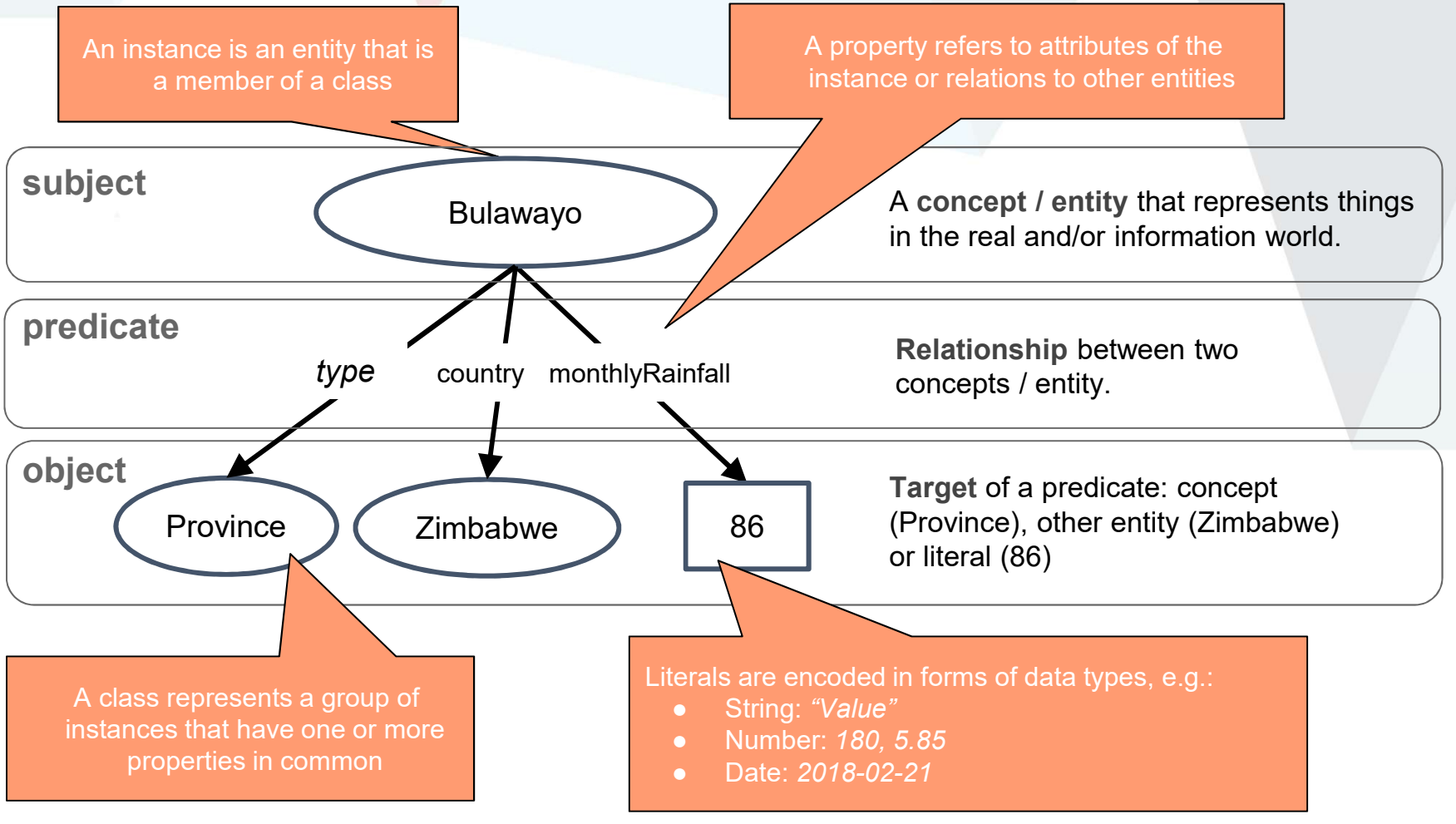
What is needed to Enable Linking and Maximise Re-use?

- Highly-structured Data Format
- Using Common Domain Models
- Porting existing data onto highly-structure format
- Using Universal Identifiers for Things



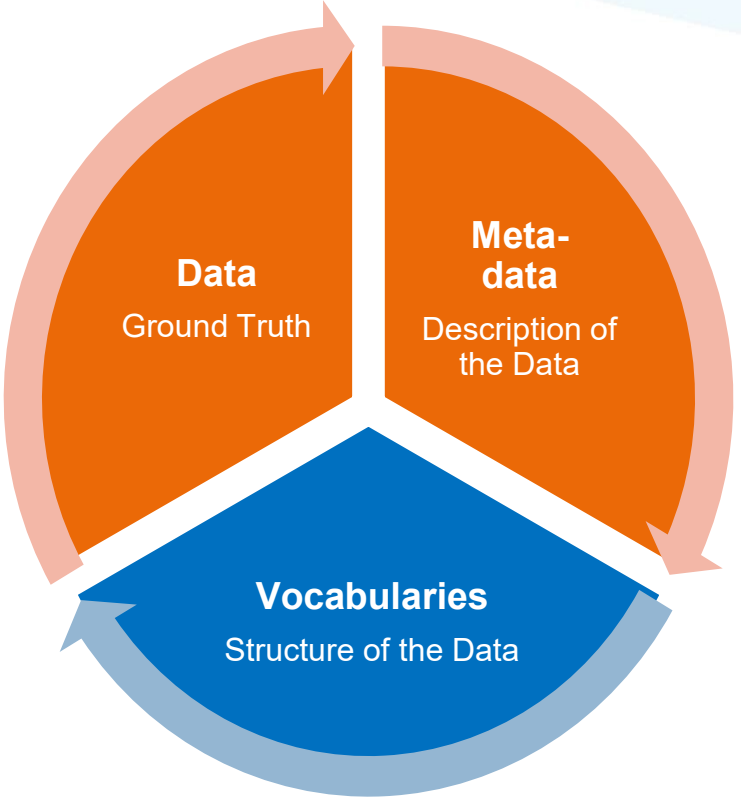
VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

Encoding Knowledge with RDF



VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

Ontologies: Enabling a common ‘Language’



- **Raw data / ground truth**
  - People, Places, Organisations, Sensor data, Production data, etc.
- **Metadata**
  - License information, Provenance, Versioning, Documentation, etc.
- **Vocabularies**
  - Domain Models: Definitions of Class and Property(-hierarchies)
  - Define Metadata that describes Raw Data (entities) using
  - T-box enabling Knowledge Representation

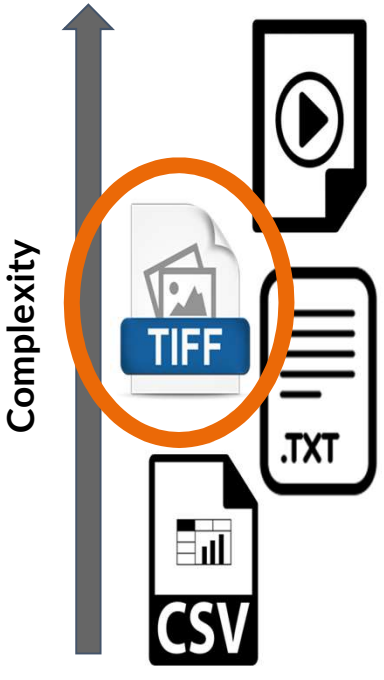
VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

Transformation tools: Un/Semi/Structured data to RDF

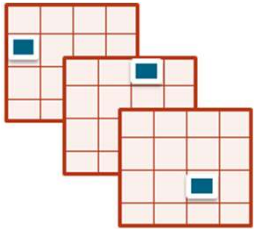
Un/Structured Text, Media, etc.

Relational data models

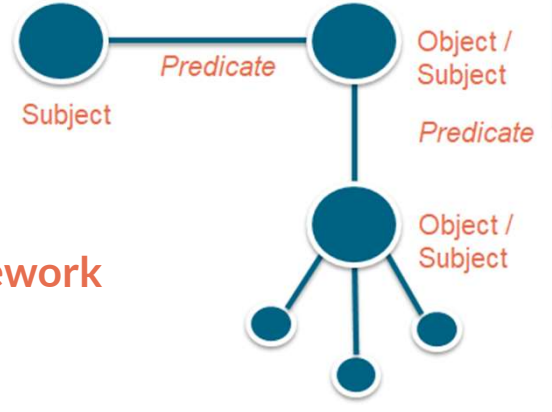
Graph based data model



?



R2RML
SILK Framework
....

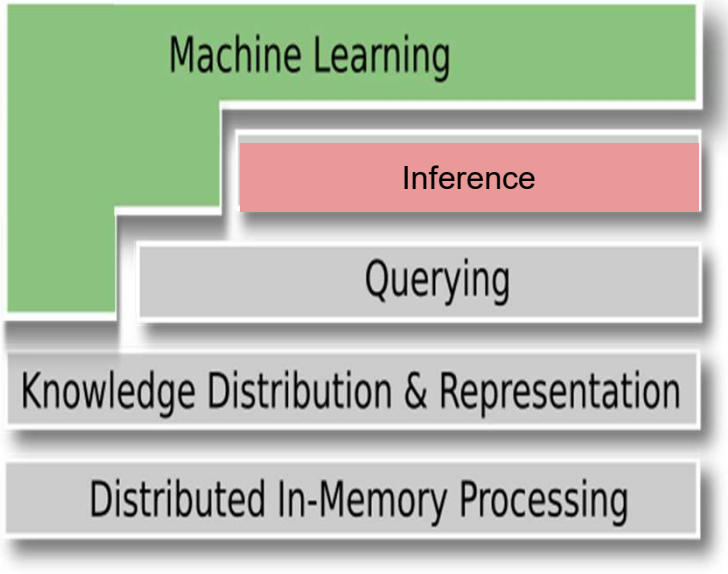


VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

deimos elecnor group | TERRAQUE | Fraunhofer IAIS | SatCen | WFP | ETH zürich

SANSA: Innovative KG-based Inference & ML

- **Machine Learning Distributed algorithms**
  - KG embeddings for KB completion, link prediction
  - Graph Clustering
  - Association Rule Mining
  - Semantic Decision Trees



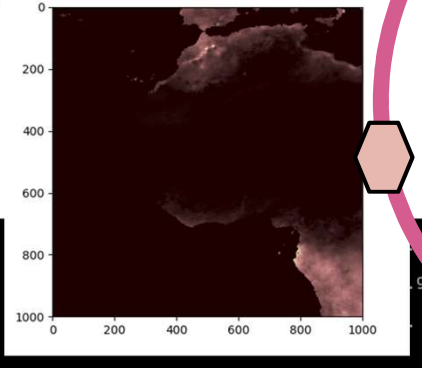
- **Inference**
  - In-memory via **rule-based forward chaining**
  - Dynamically build **Rule dependency graphs**
  - Based on RDF/OWL fragments

VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

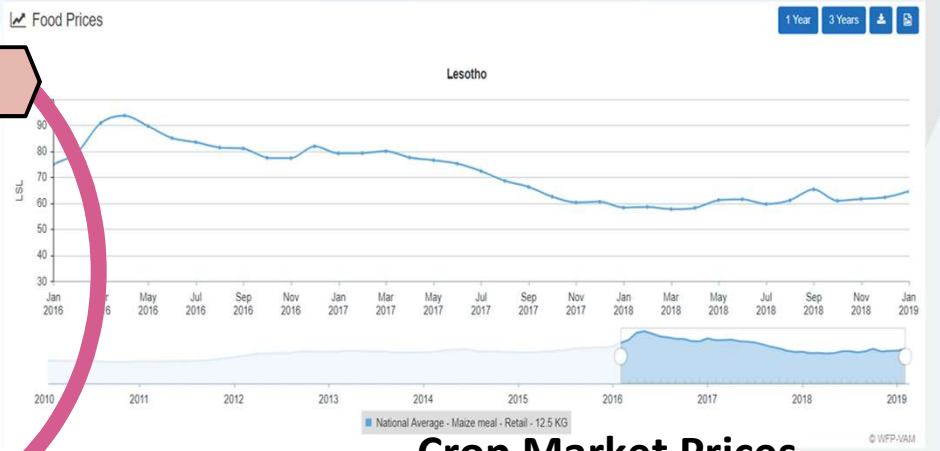
deimos elecnor group | TERRAQUE | Fraunhofer IAIS | SatCen | WFP | ETH zürich

Today's Exercise!
Rainfall to Market Price Correlation Detection

CHIRPS Rainfall aggregates




```
>>> array
array([[  0.      ,
        134.87132],
       [  0.      ,
        114.61279],
       [  0.      ,
        101.95908],
       ...,
       [  0.      ,   0.      ,   0.      , ..., 167.36545, 170.28468,
        176.22345],
       [  0.      ,   0.      ,   0.      , ..., 163.13359, 166.25623,
        173.71649],
       [  0.      ,   0.      ,   0.      , ..., 161.90408, 167.51051,
        168.20584]], dtype=float32)
>>> array[989,934]
171.32156
>>>
```



Crop Market Prices

VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

deimos elecnor group | TERRAQUE | Fraunhofer IAIS | SatCen | WFP | ETH zürich

CHIRPS Knowledge Graph Linking & Pattern Discovery

- **Ingestion of CHIRPS-derived Knowledge Graphs/RDF**
  - Ontology Mapping for semi-automatic RDF transformation
- **Interlinking with relevant open data**
  - Crop Market Prices from WFP
- **Implicit Discovery of Patterns**
  - Identify *correlations* between Rainfall and Crop (Maize) Market prices

VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

Required Libraries

- Python 3+
  - Particular packages requirement: GDAL, Shapely
- Java 7+
- Maven
- Apache Spark 2+
- Linux OS

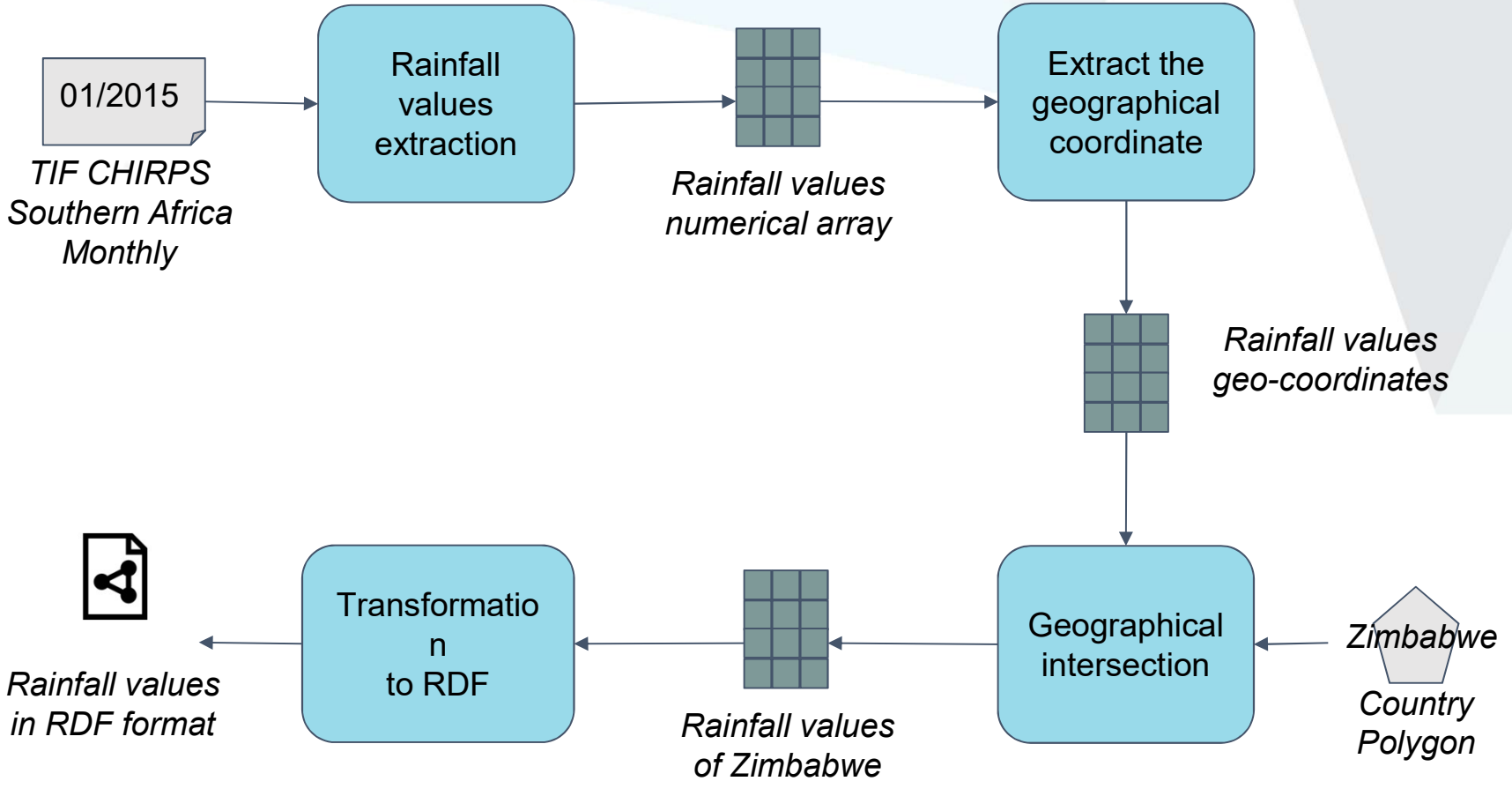
Download Material Here

https://bit.ly/better-semantic-material

VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

deimos elecnor group | TERRA DUE | Fraunhofer IAIS | SatCen | WFP | ETH zürich

Task 1: From CHIRPS TIF to RDF



VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

Task 1: From CHIRPS TIF to RDF

- **Required Libraries:** Python 3 + GDAL and Shapely packages
- **Material Location:** App > 1_Rainfall_TIF_to_RDF
- **Application Execution:** run the command

```
python3 Main.py ../../Data/CHIRPSv2_SouthernAfrica_N30_daystotal_2015-01-01_2015-01-31.tif tif-rainfall-output.ttl
```

VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

deimos elecnor group | TERRAQUE | Fraunhofer IAIS | SatCen | WFP | ETH zürich

Task 1: From CHIRPS TIF to RDF

- **Output:** RDF multidimensional data (aka Data Cube Vocabulary)

```
@prefix qb: <http://purl.org/linked-data/cube#> .
@prefix eg: <http://example.org/ns#> .
@prefix dbpedia: <http://dbpedia.org/ontology/> .
@prefix cf-feature: <http://purl.oclc.org/NET/ssnx/cf/cf-feature#> .

eg:obs1 a qb:Observation ;
    qb:dataSet eg:dataset-prices ;
     dbpedia:month "1" ;
     dbpedia:year "2015" ;
     cf-feature:rainfall_amount "789587" ;
.
```

Namespaces (domains)

An Observation

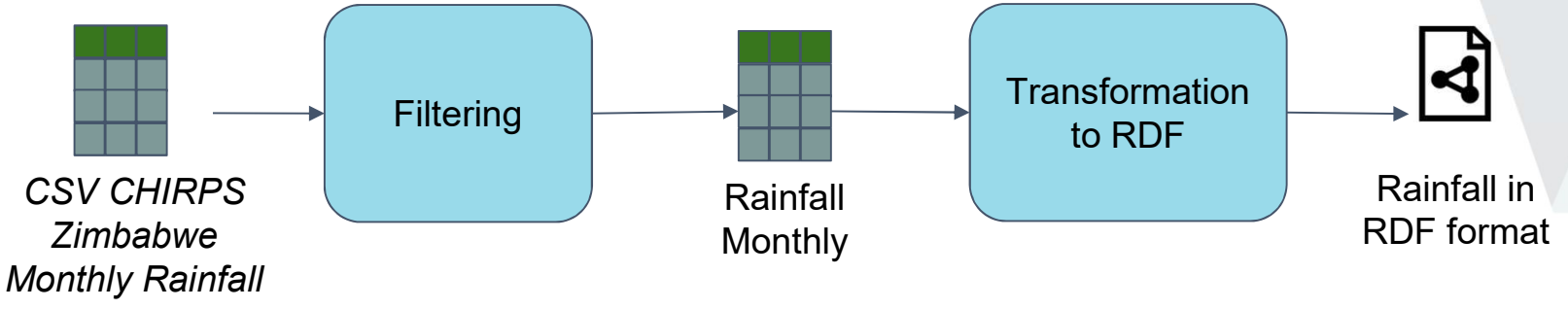
A DataSet

Dimensions

The Measure

VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

Task 2: From CHIRPS CSV to RDF



VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

Task 2: From CHIRPS CSV to RDF

- **Required libraries:** Python 3
- **Material location:** App > 2_Rainfall_CSV_to_RDF
- **Application Execution:** run the command

```
python3 Main.py ../../Data/Zimbabwe_Rainfall.csv rainfall-output.ttl
```

Task 2: From CHIRPS CSV to RDF

- **Output:** RDF multidimensional data (aka Data Cube Vocabulary)

```
@prefix qb: <http://purl.org/linked-data/cube#> .
@prefix eg: <http://example.org/ns#> .
@prefix dbpedia: <http://dbpedia.org/ontology/> .
@prefix cf-feature: <http://purl.oclc.org/NET/ssnx/cf/cf-feature#> .

eg:obs1 a qb:Observation ;
    qb:dataSet eg:dataset-prices ;
     dbpedia:month "1" ;
     dbpedia:year "2015" ;
     cf-feature:rainfall_amount "789587" ;
.
```

Namespaces (domains)

An Observation

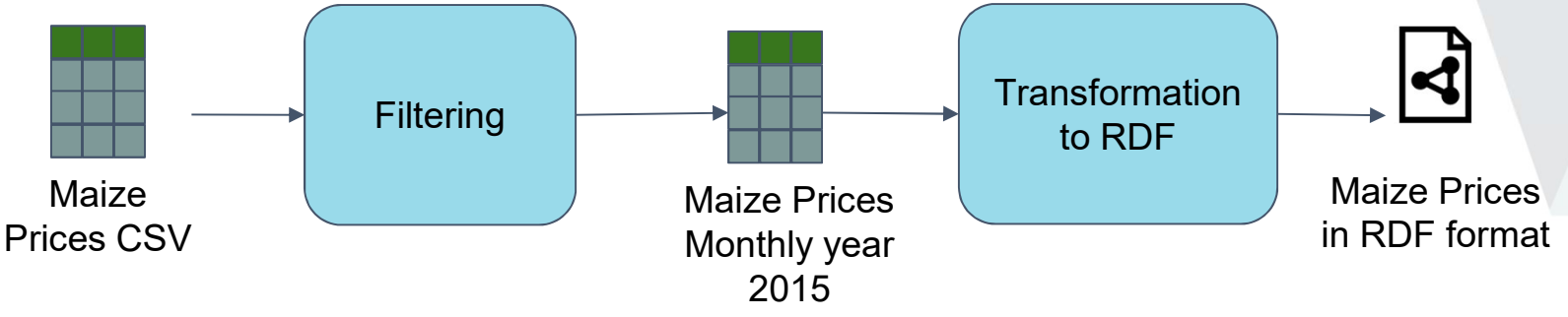
A DataSet

Dimensions

The Measure

VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

Task 3: From WFP Maize prices to RDF



VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

Task 3: From WFP Maize prices to RDF

- **Required Libraries:** Python 3
- **Material Location:** App > 3_Prices_CSV_to_RDF
- **Application Execution:** run the command

```
python3 Main.py ../../Data/Zimbabwe_Maize_Prices-2015.csv prices-output.ttl
```

Task 3: From WFP Maize prices to RDF

- **Output:** RDF multidimensional data (aka Data Cube Vocabulary)

```
@prefix qb: <http://purl.org/linked-data/cube#> .
@prefix eg: <http://example.org/ns#> .
@prefix dbpedia: <http://dbpedia.org/ontology/> .
@prefix cbo: <http://comicmeta.org/cbo/> .

eg:obs1 a qb:Observation ;
    qb:dataSet eg:dataset-prices ;
     dbpedia:month "1" ;
     dbpedia:year "2015" ;
     cbo:price "789587" ;
.
```

Namespaces (domains)

An Observation

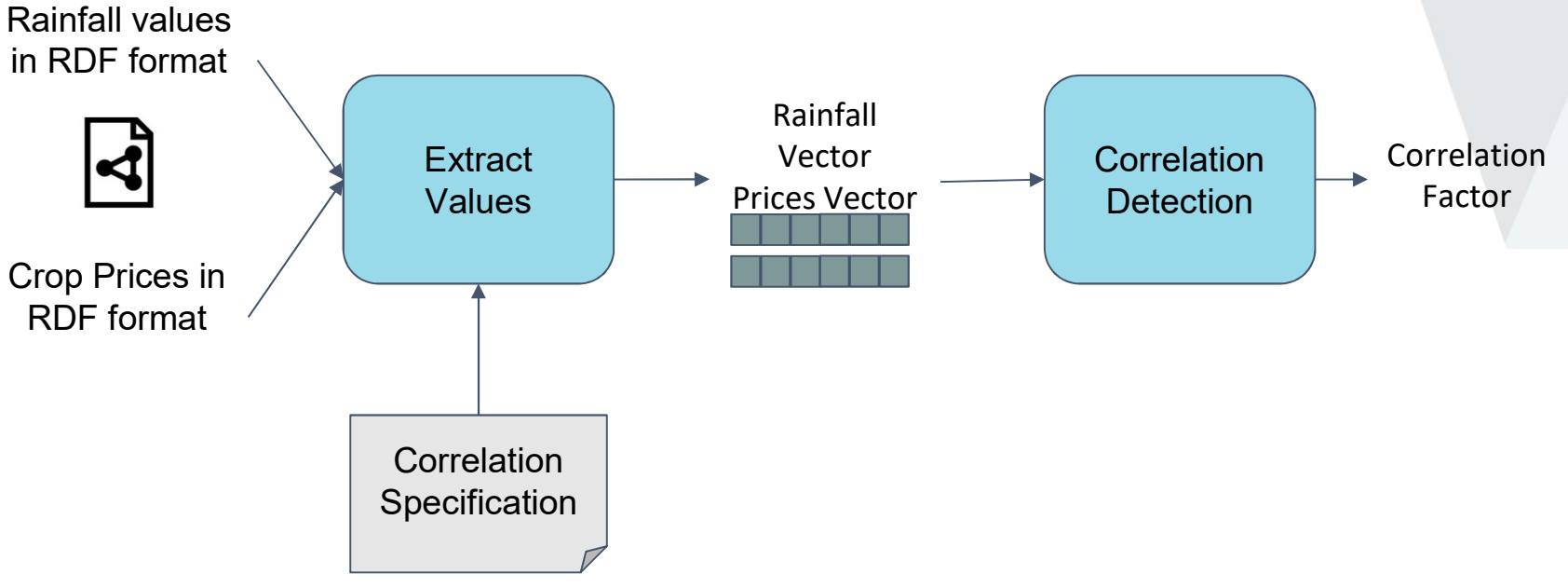
A DataSet

Dimensions

The Measure

VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

Task 4: Correlation Detection



VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

Task 4: Correlation Detection

```
{
  "properties": [
    {
      "source": "/home/user/.../Apps/2_Rainfall_CSV_to_RDF/rainfall-output.ttl",
      "target": "<http://purl.oclc.org/NET/ssnx/cf/cf-feature#rainfall_amount>",
      "filters": {
        "<http://dbpedia.org/ontology/year>": "2015",
        "<http://dbpedia.org/ontology/month>": "1,2,3,4"
      }
    },
    {
      "source": "/home/user/.../Apps/3_Prices_CSV_to_RDF/prices-output.ttl",
      "target": "<http://comicmeta.org/cbo/price>",
      "filters": {
        "<http://dbpedia.org/ontology/month>": "{02,03,04}{03,04,05}{04,05,06}{05,06,07};sum",
        "<http://dbpedia.org/ontology/year>": "2015"
      }
    }
  ]
}
```

Filter based on other properties

The two properties to correlated

VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

Task 4: Correlation Detection

- **Required Libraries:** Python 3, Maven, Apache Spark
- **Material Location:** App > 4_SANSA_Correlation_Detection
- **Application Execution:**
  - Run the command `mvn package` to package the App
  - Run the command (if SPARK_HOME not set, navigate to Spark folder)

```
spark-submit --class org.ml.test.App --master local[*] --executor-memory 5G
target/correlation-1.0-SNAPSHOT.jar ../../Data/correlation.conf local[*]
```

Task 4: Correlation Detection

- **Output:** Correlation Value using *Pearson* method
  - Value range between -1 and 1
  - The closer to 1 = positive correlation
    - When rainfalls change, prices change in the same direction (up/down)
  - The closer to -1 = negative correlation
    - When rainfalls change, prices change in the opposite direction (up/down)

VIEW these slides here: https://bit.ly/better-semantic-analysis
Download material here: https://bit.ly/better-semantic-material

deimos elecnor group | TERRAQUE | Fraunhofer IAIS | SatCen | WFP | ETH zürich