A Provenance Model for Quantified Self Data

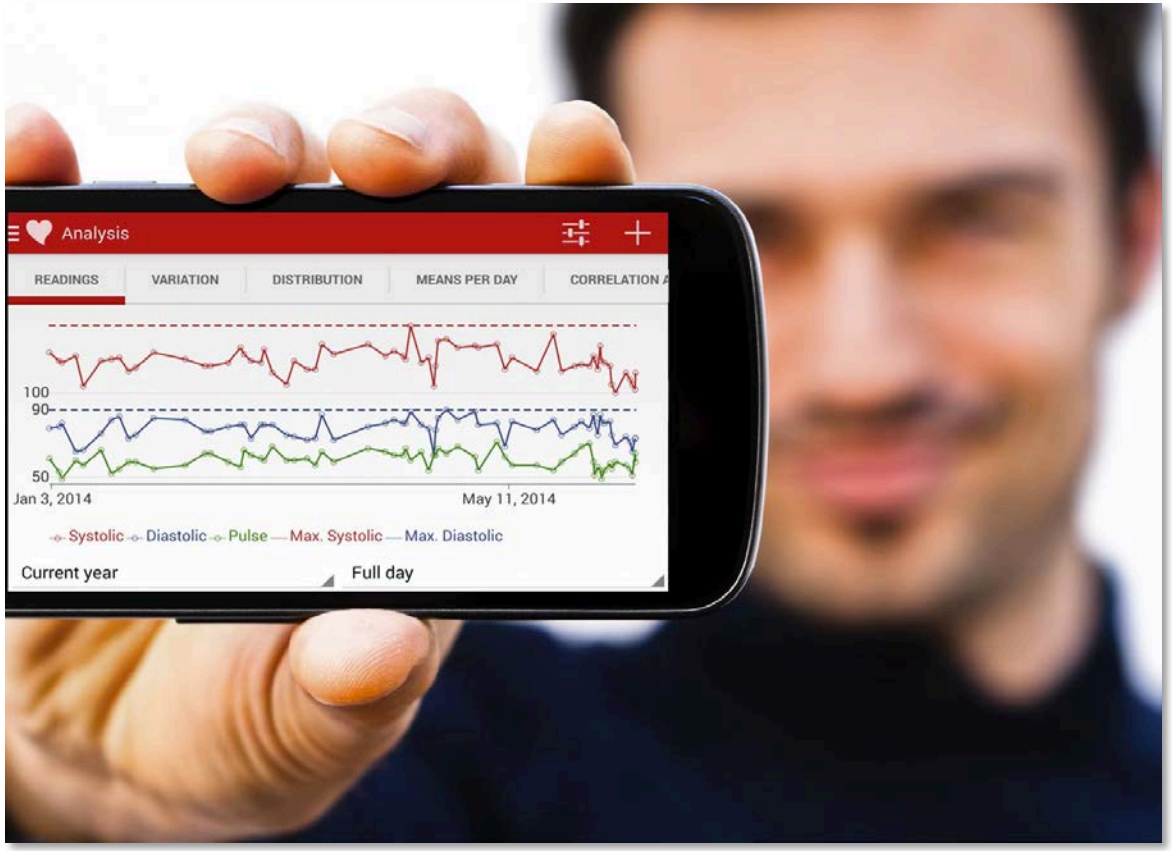
Andreas Schreiber

Department for Intelligent and Distributed Systems
German Aerospace Center (DLR), Cologne/Berlin

DLR

Motivation
Use Cases

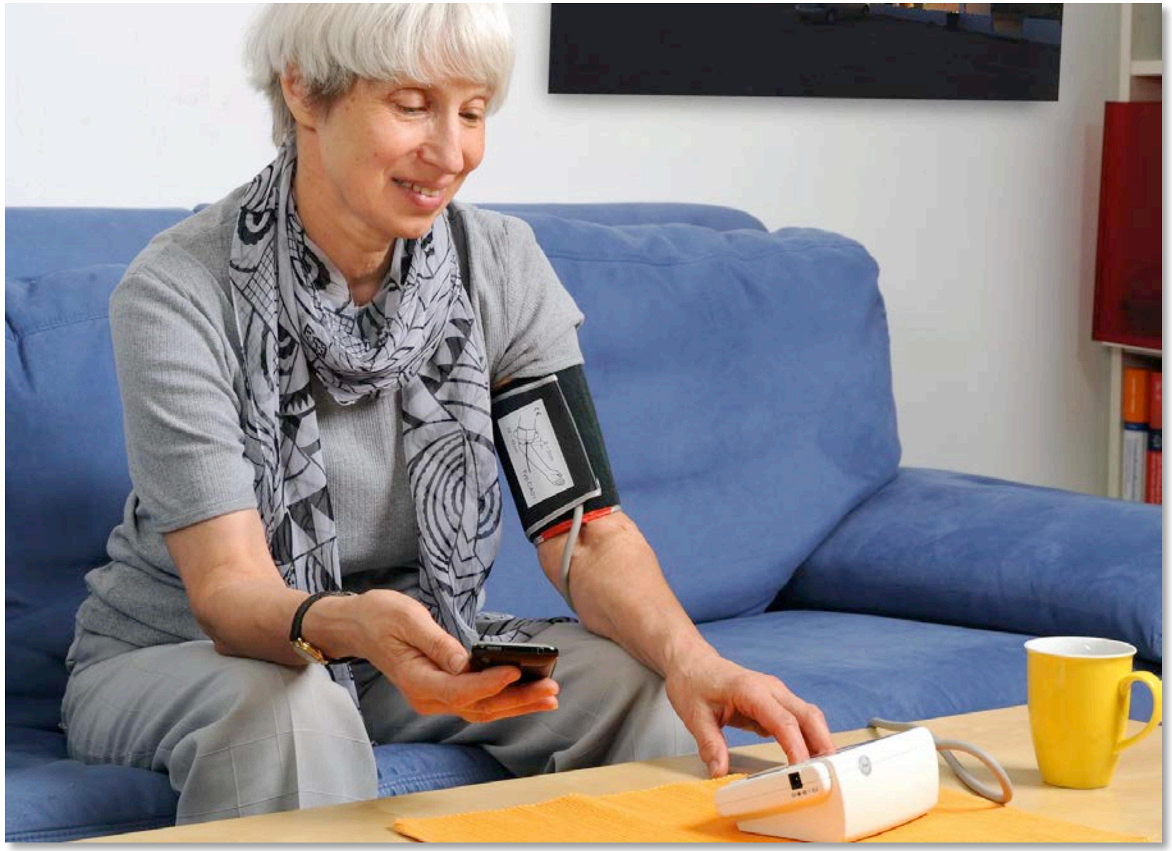
Quantified Self (n = 1)



Medical Trials (n > 1)

Motivation
More Use Cases

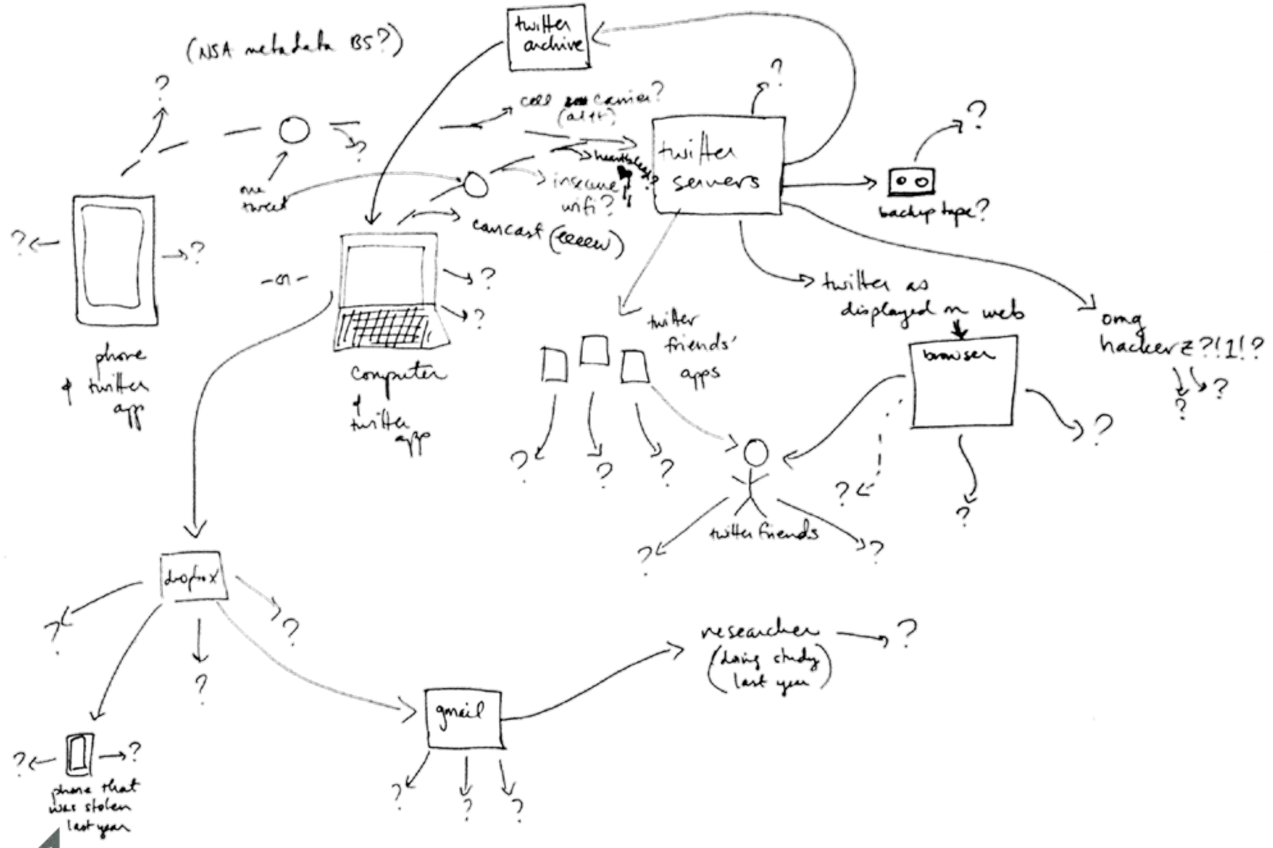
Telemedicine

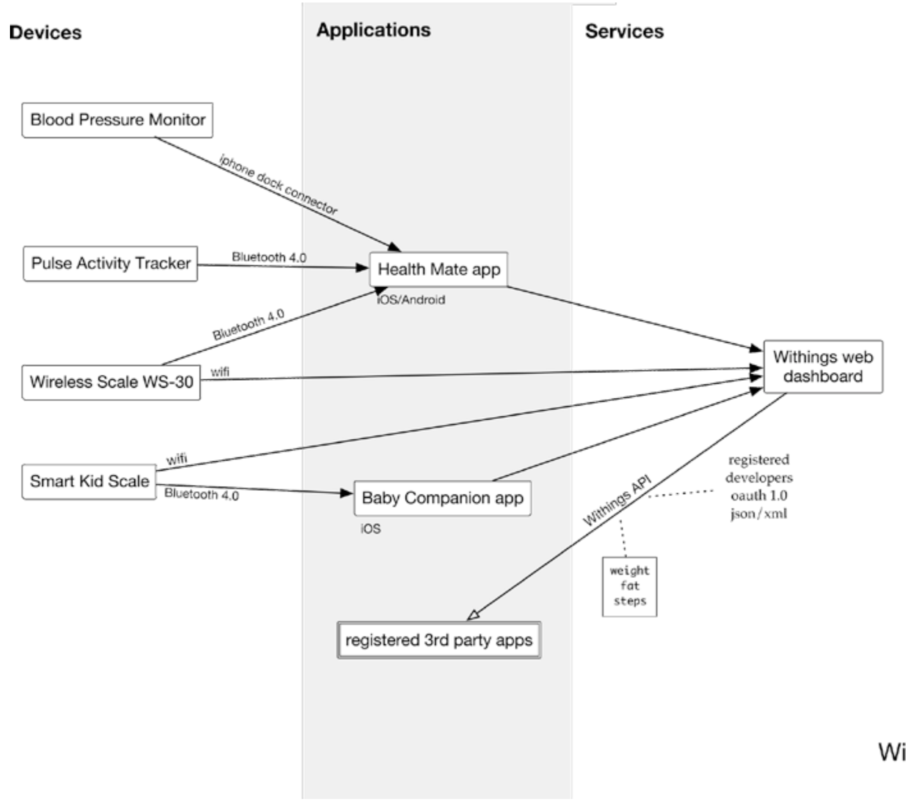


Medical experiments

Motivation
Understand, how data has been produced, processed, stored, accessed, ...

It started at the *Breakout Session on Mapping Data Access* (2014 QS Europe Conference, Amsterdam)
https://forum.quantifiedself.com/t/breakout-mapping-data-access/995





Withings

Motivation
(Provenance) Questions

Entity-focused

- What data about the user were created during the activity X?
- What data about the user were automatically generated?
- What data about the user were derived from manual input?

Activity-focused

- Which activities support visualization of the users data?
- In which activities can the user input data?
- What processes are communicating data?

Agent-focused

- What parties were involved in generating data X?
- What parties got access on data X?
- Can other parties see user’s data X?

➢ Data

➢ Apps and Services

➢ Access and Privacy

Provenance
Definition

Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness.

PROV W3C Working Group
https://www.w3.org/TR/prov-overview

PROV
W3C Specification for Provenance

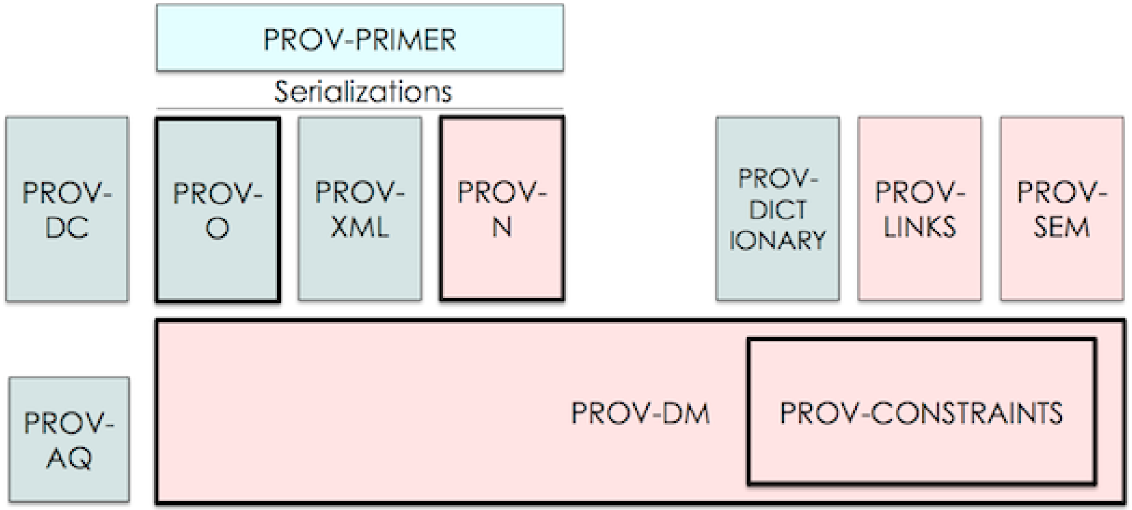
W3C Provenance Working Group: `https://www.w3.org/2011/prov`

PROV

- The goal of PROV is *to enable the wide publication and interchange of provenance* on the Web and other information systems
- PROV enables one *to represent and interchange provenance information using widely available formats* such as RDF and XML

PROV Family of Documents

- **PROV-O**: PROV Ontology
- **PROV-DM**: PROV Data Model
- **PROV-N**: PROV Notation
- **PROV-CONSTRAINTS**: PROV Constraints
- **PROV-AQ**: Accessing and Querying



Overview of PROV
Key Concepts

Entities

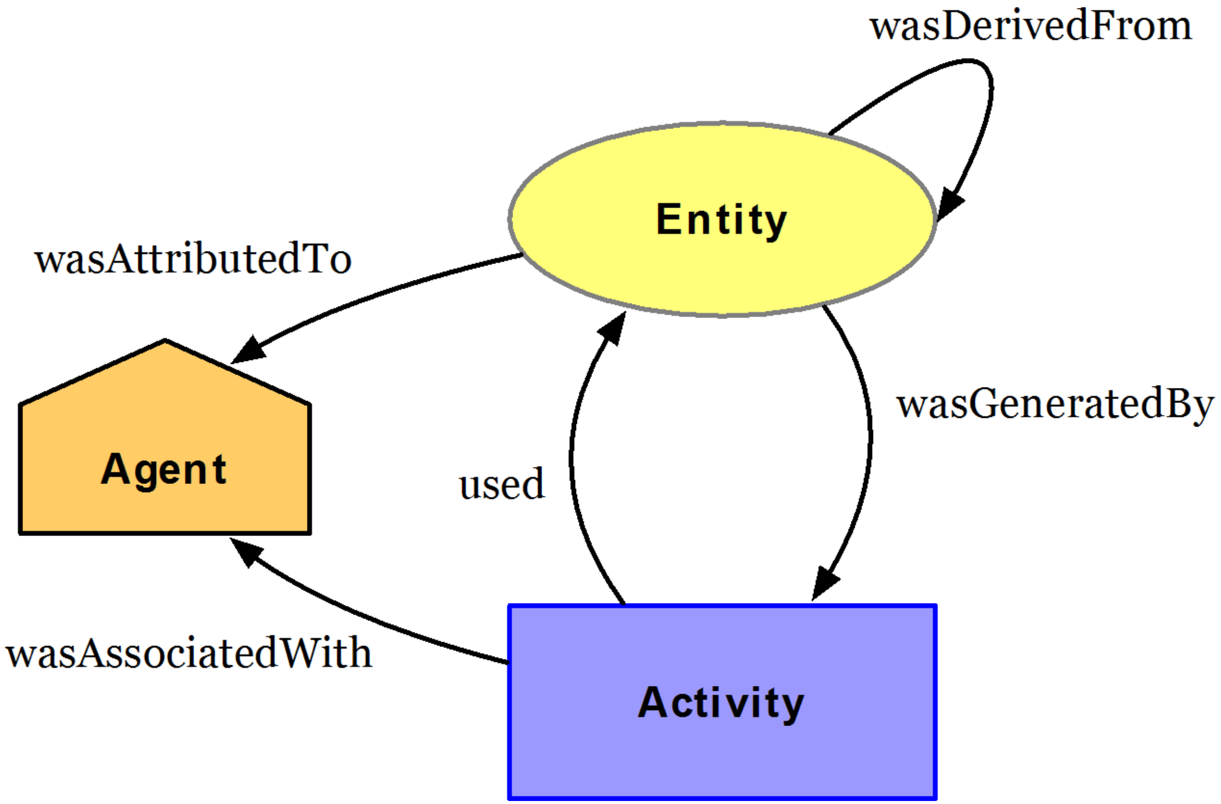
- Physical, digital, conceptual, or other kinds of things
- For example, documents, web sites, graphics, or data sets

Activities

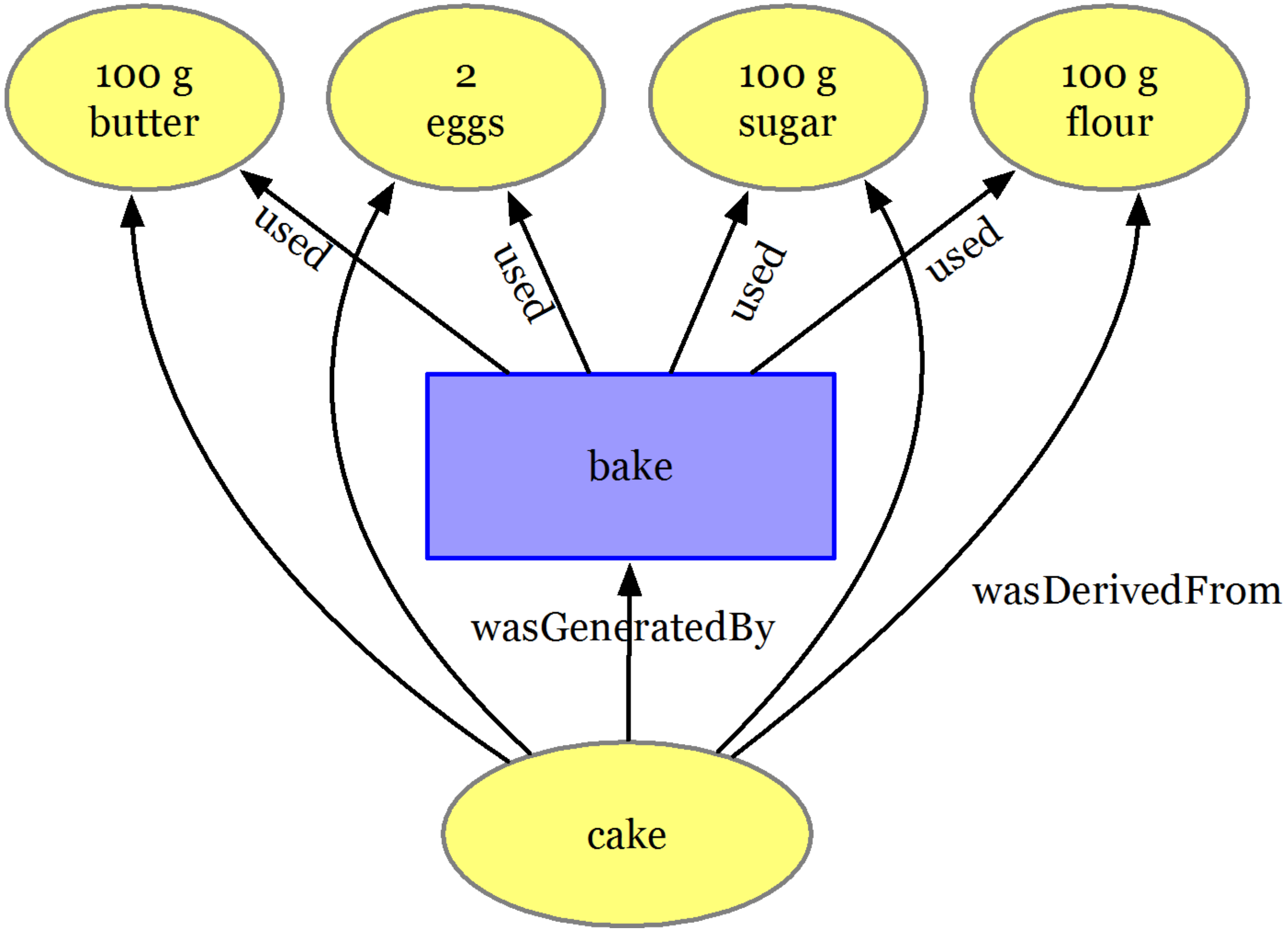
- Activities *generate* new entities or make *use* of existing entities
- Activities could be actions or processes

Agents

- Agents takes a role in an activity and have the responsibility for the activity
- For example, persons, pieces of software, or organizations



PROV Example
Baking a Cake



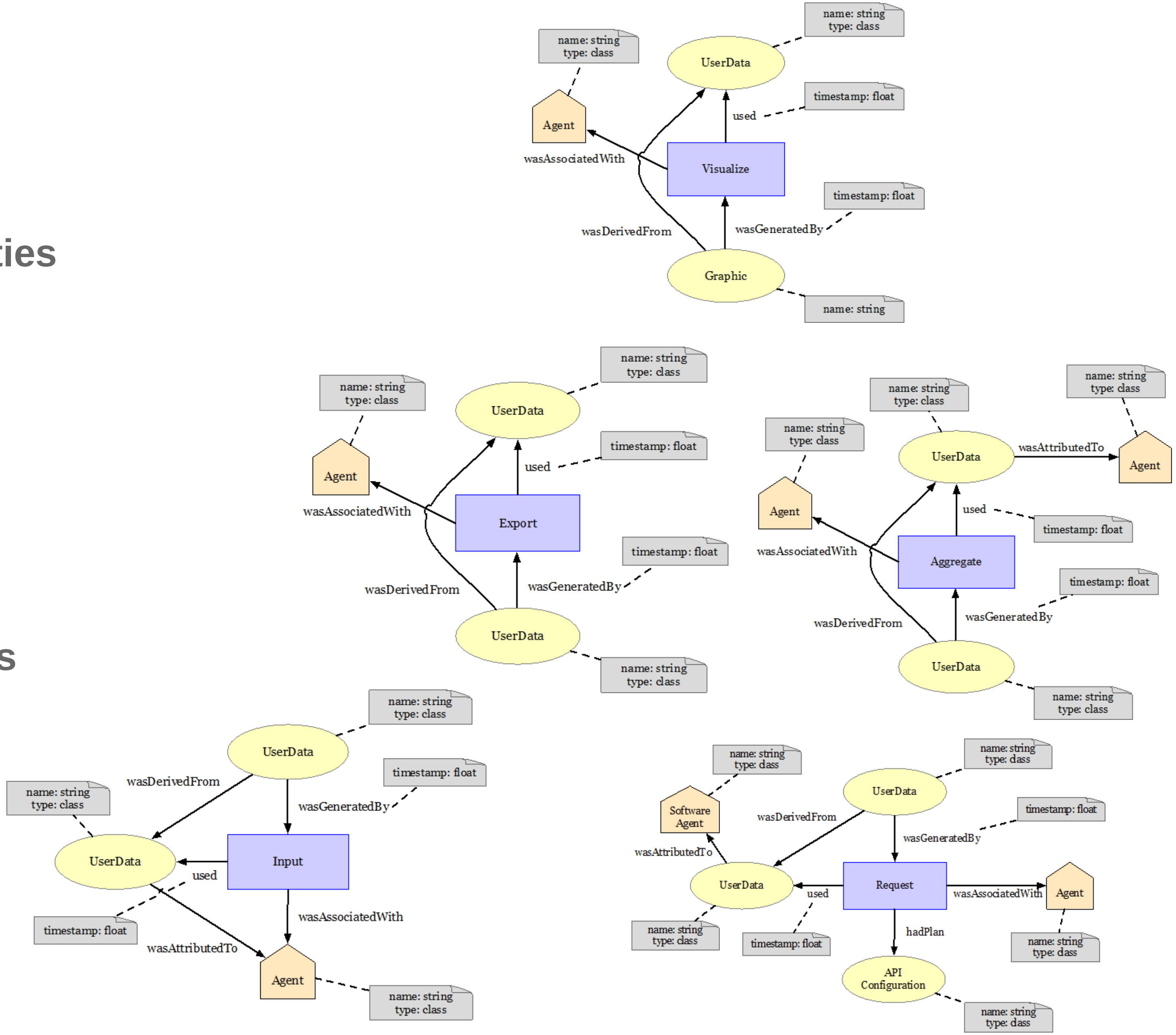
PROV Model for Quantified Self
Overview

Five sub models (templates) for basic activities

- Input
- Export
- Request
- Aggregate
- Visualize

The activities generate or change data that is associated or attributed to Agents

- Users
- Software



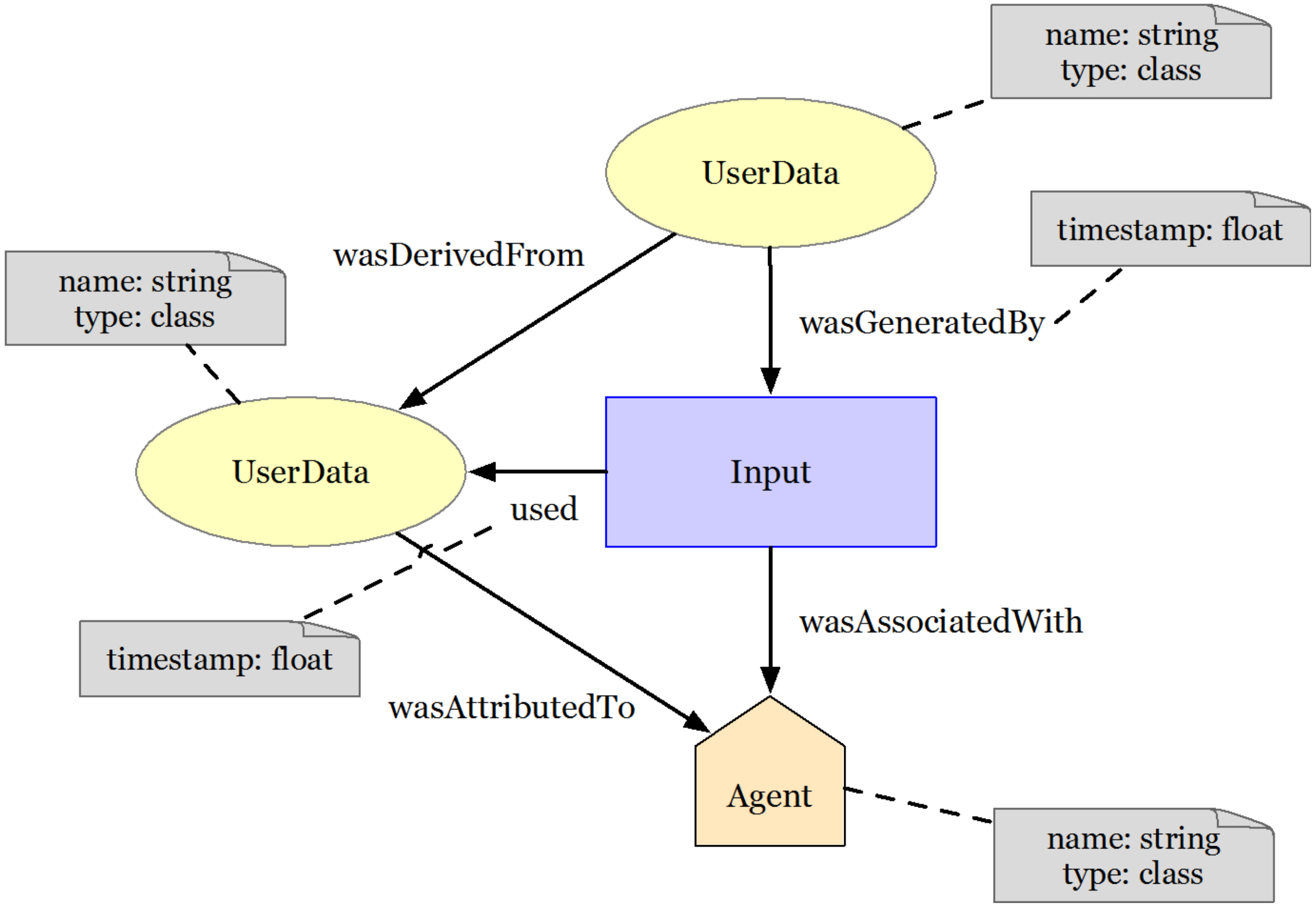
DLR

PROV Model for Quantified Self
Input

Getting (raw) data from the user

- Manual input
- Via sensors

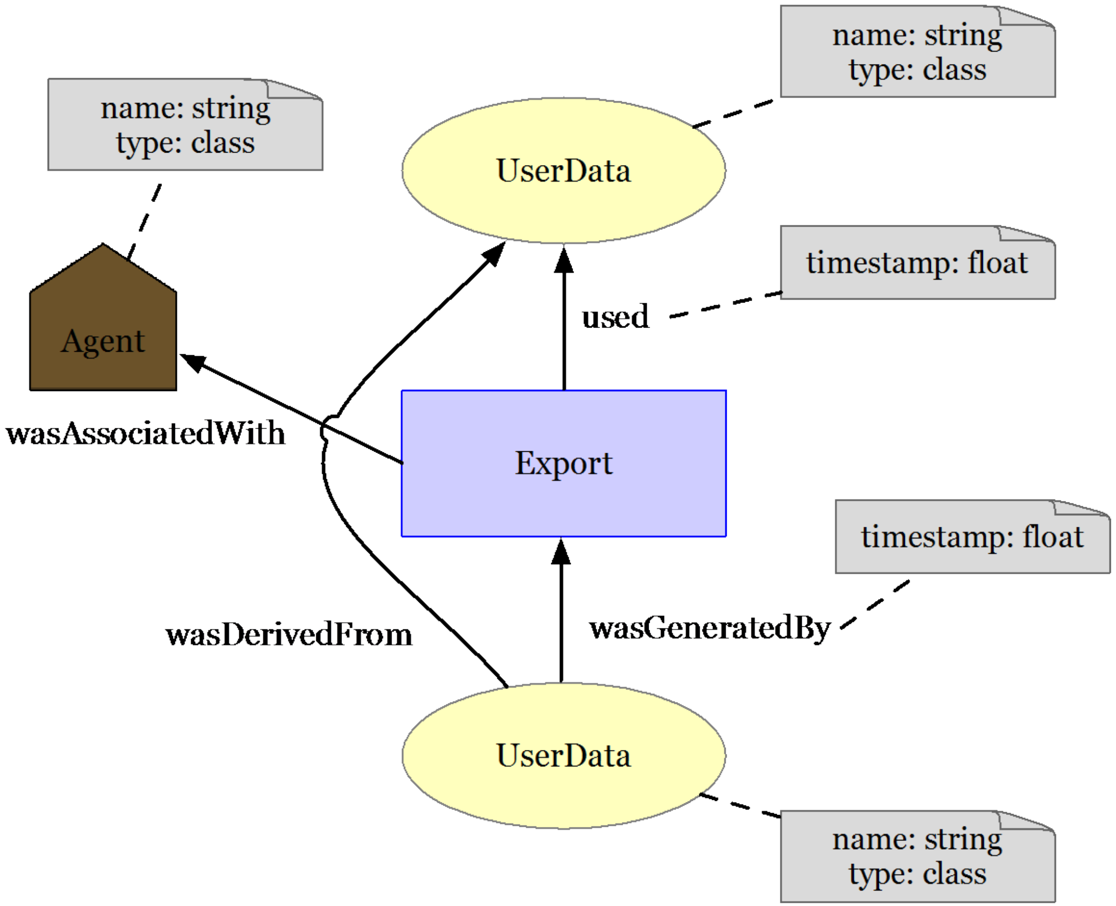
Usually every QS workflow needs such a functionality to record data from the user



PROV Model for Quantified Self
Export

Exports data to other formats

- Transfer to other systems
- Readable by people



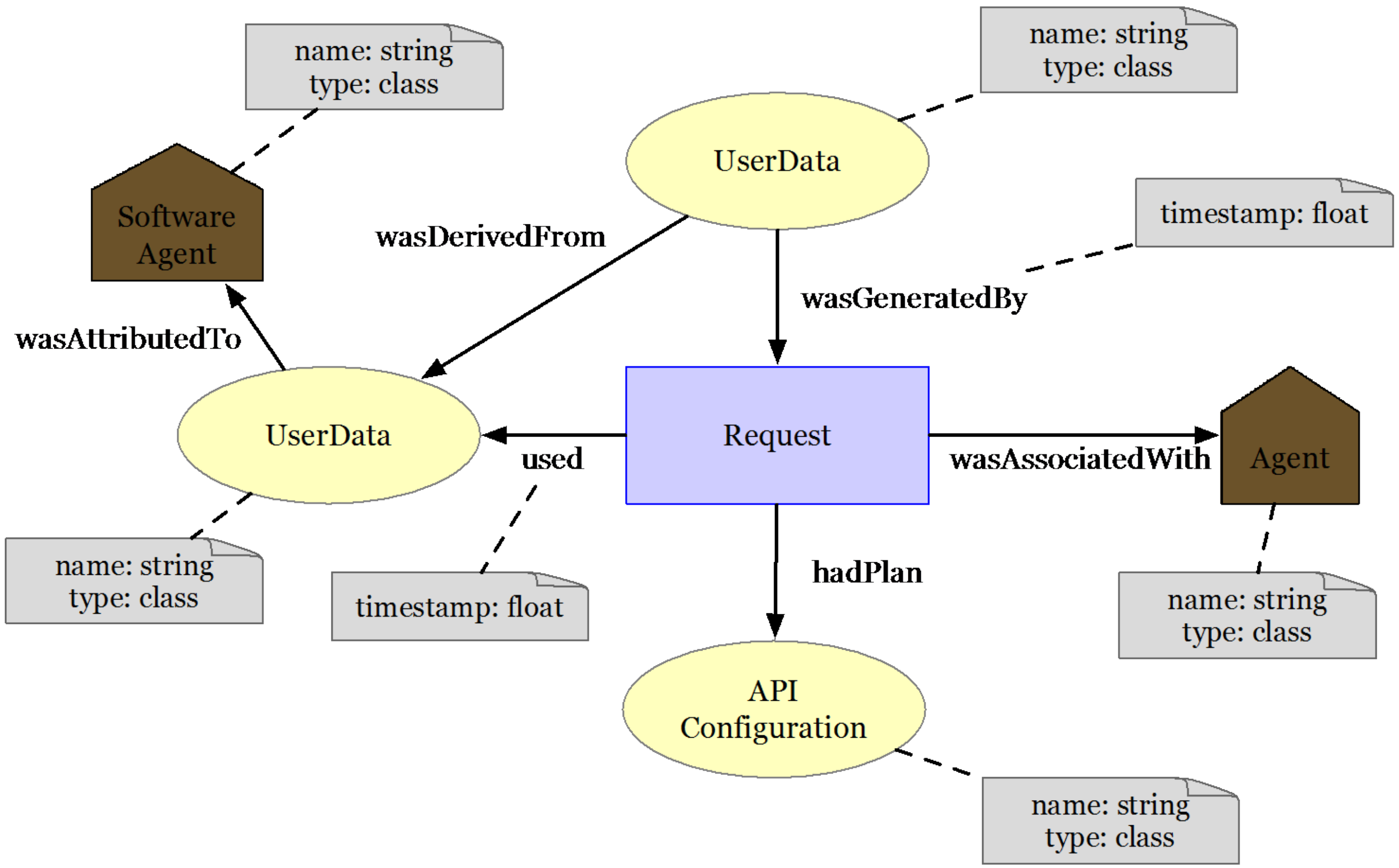
PROV Model for Quantified Self
Request

Getting or storing data

- Web Services
- Cloud Storage

Access via APIs

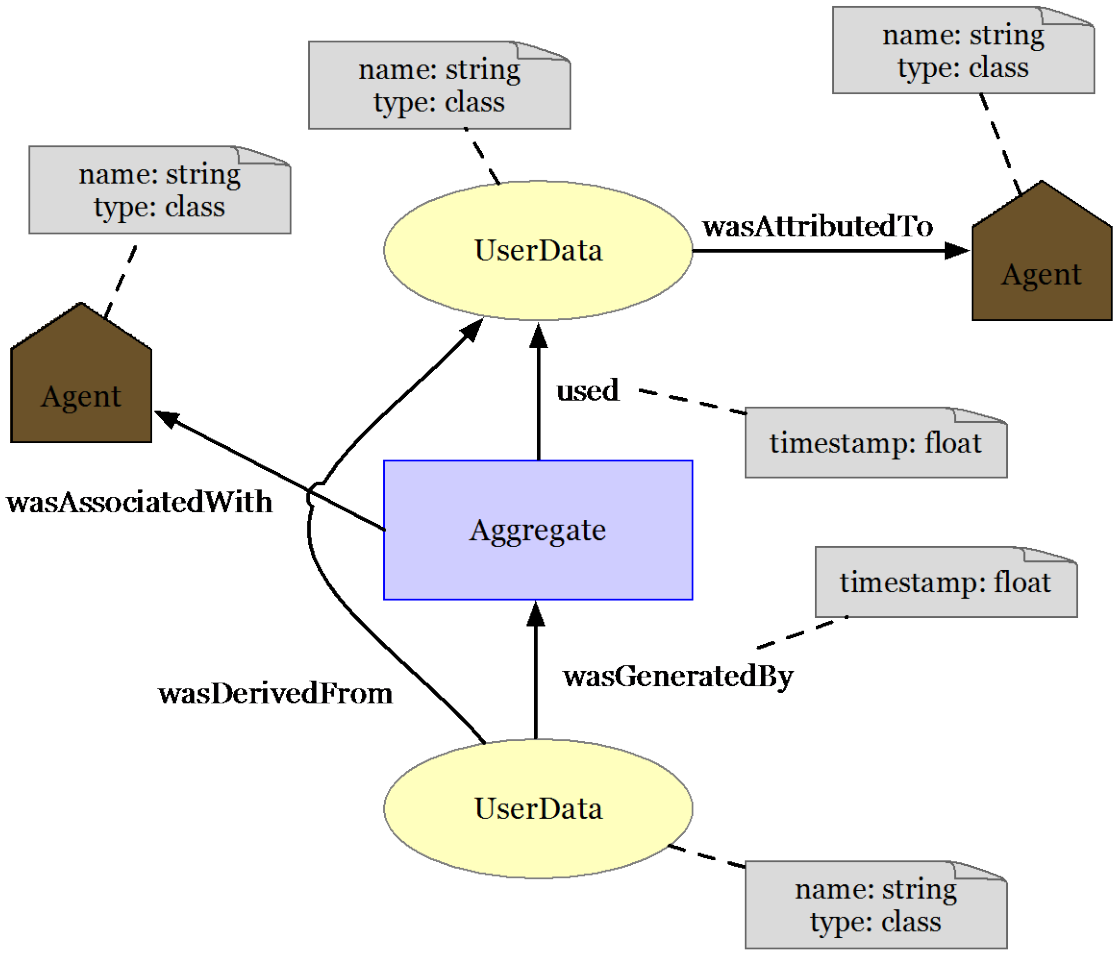
- Vendor APIs (Fitbit, Withings, Microsoft etc.)



PROV Model for Quantified Self
Aggregate

Combine or analyze data

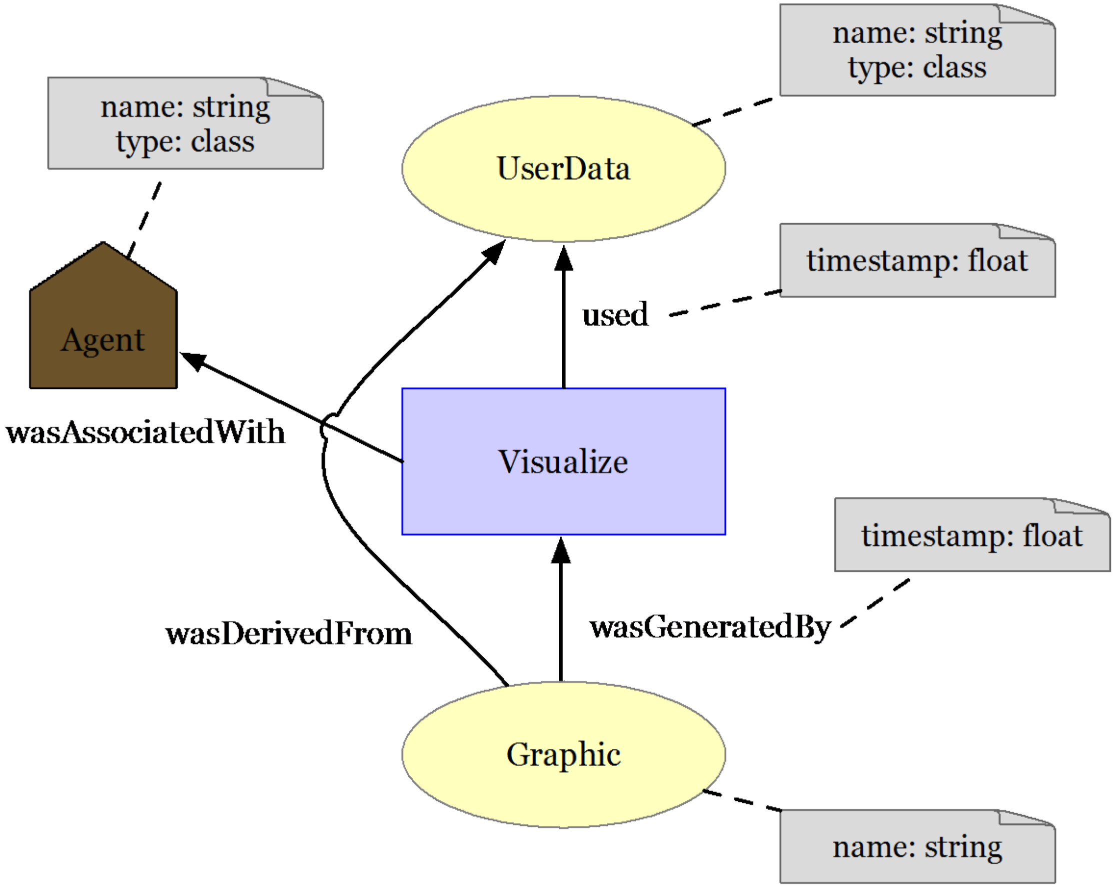
- Aggregate different data sets
- Statistical analysis
- Extract knowledge



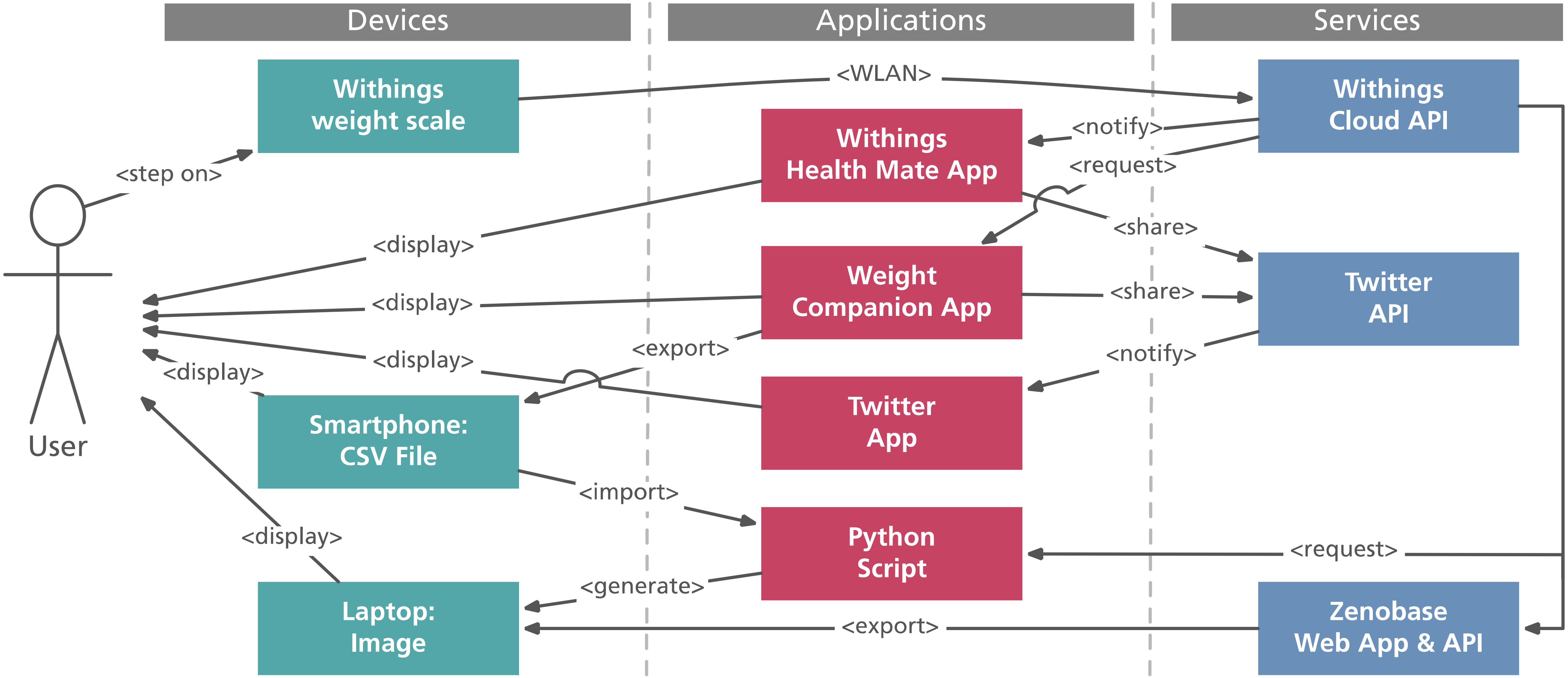
PROV Model for Quantified Self
Visualize

Visualize data

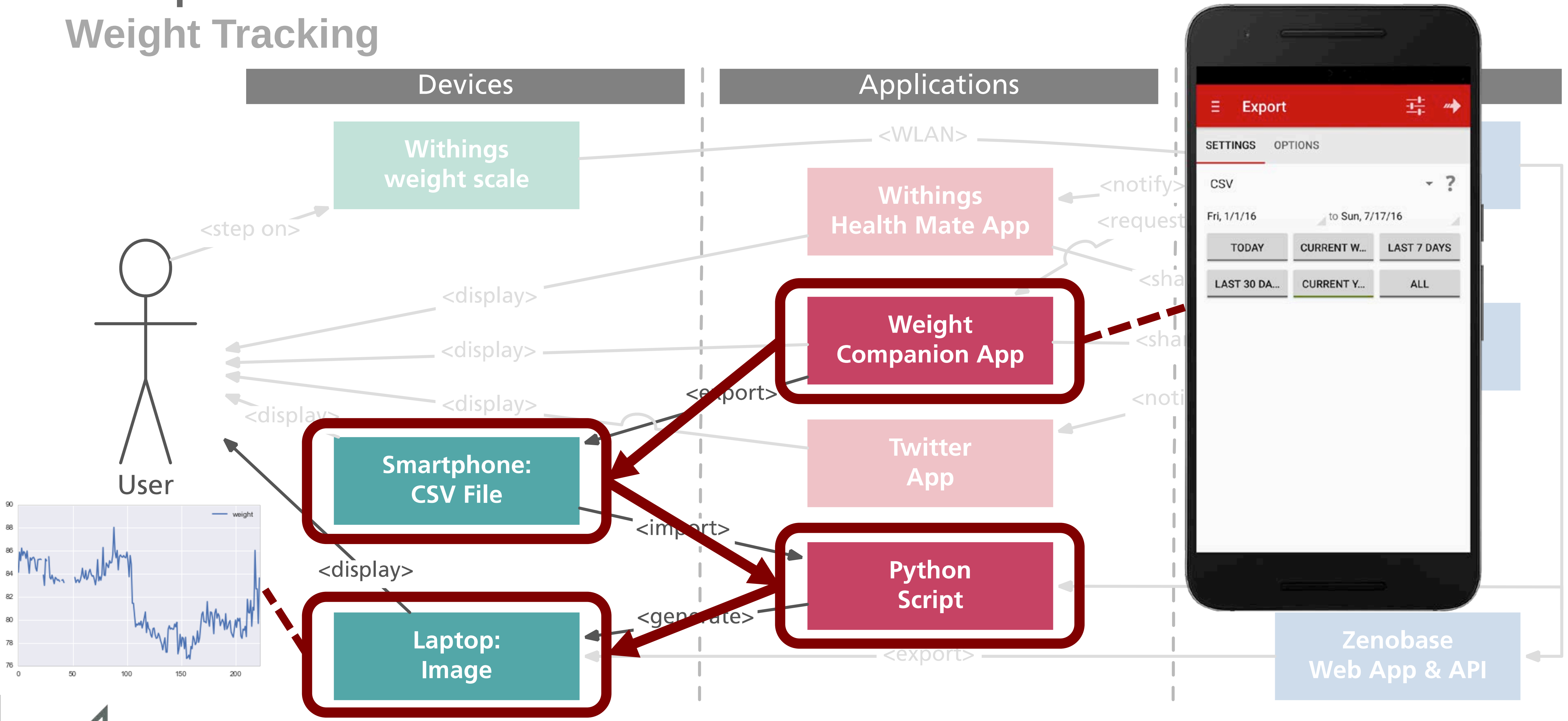
- Time series plots
- Representation of statistical information
- Historical information



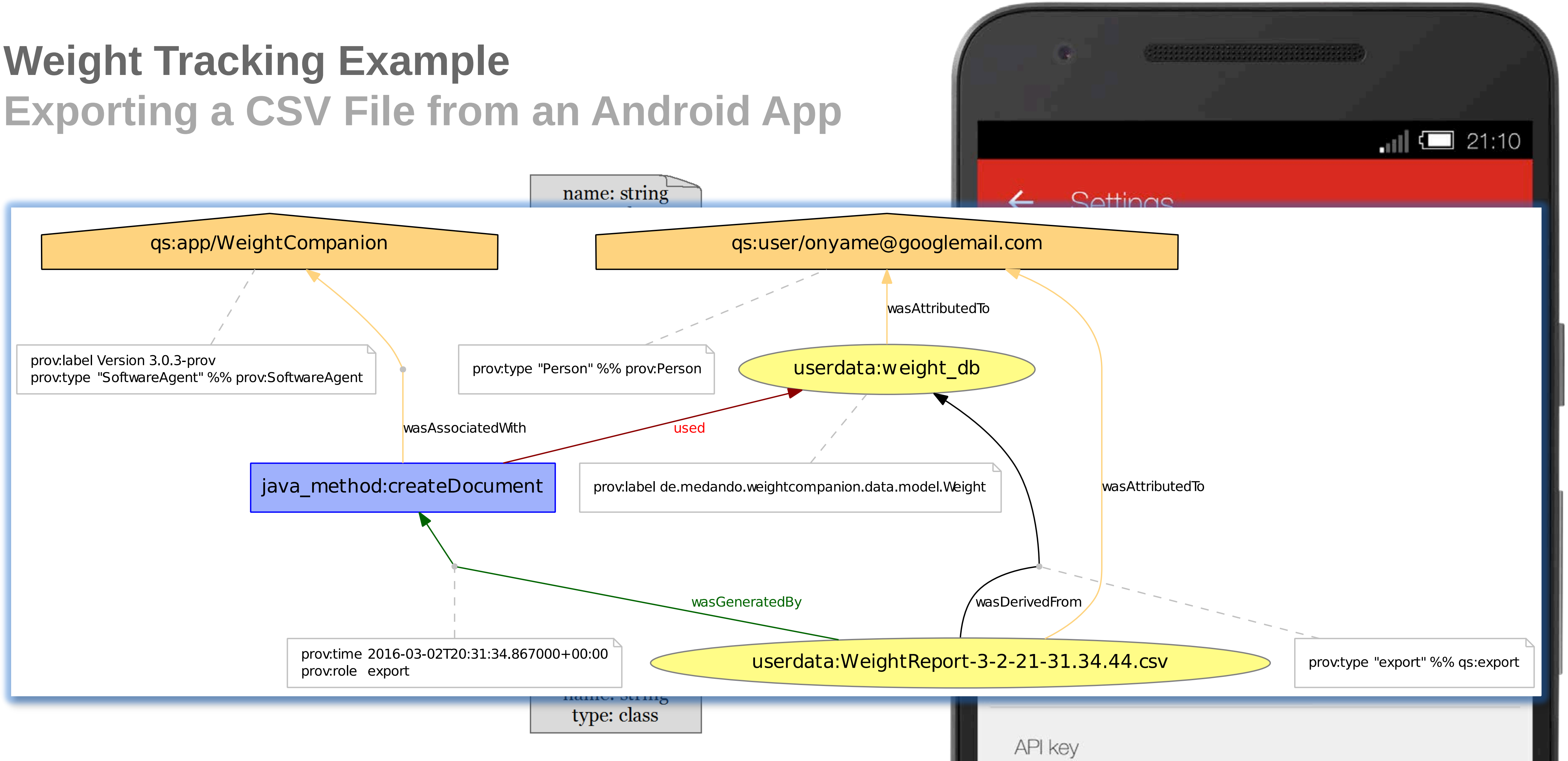
Example
Weight Tracking



Example
Weight Tracking



Weight Tracking Example
Exporting a CSV File from an Android App



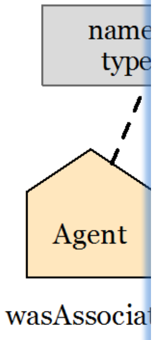
https://play.google.com/store/apps/details?id=de.medando.weightcompanion

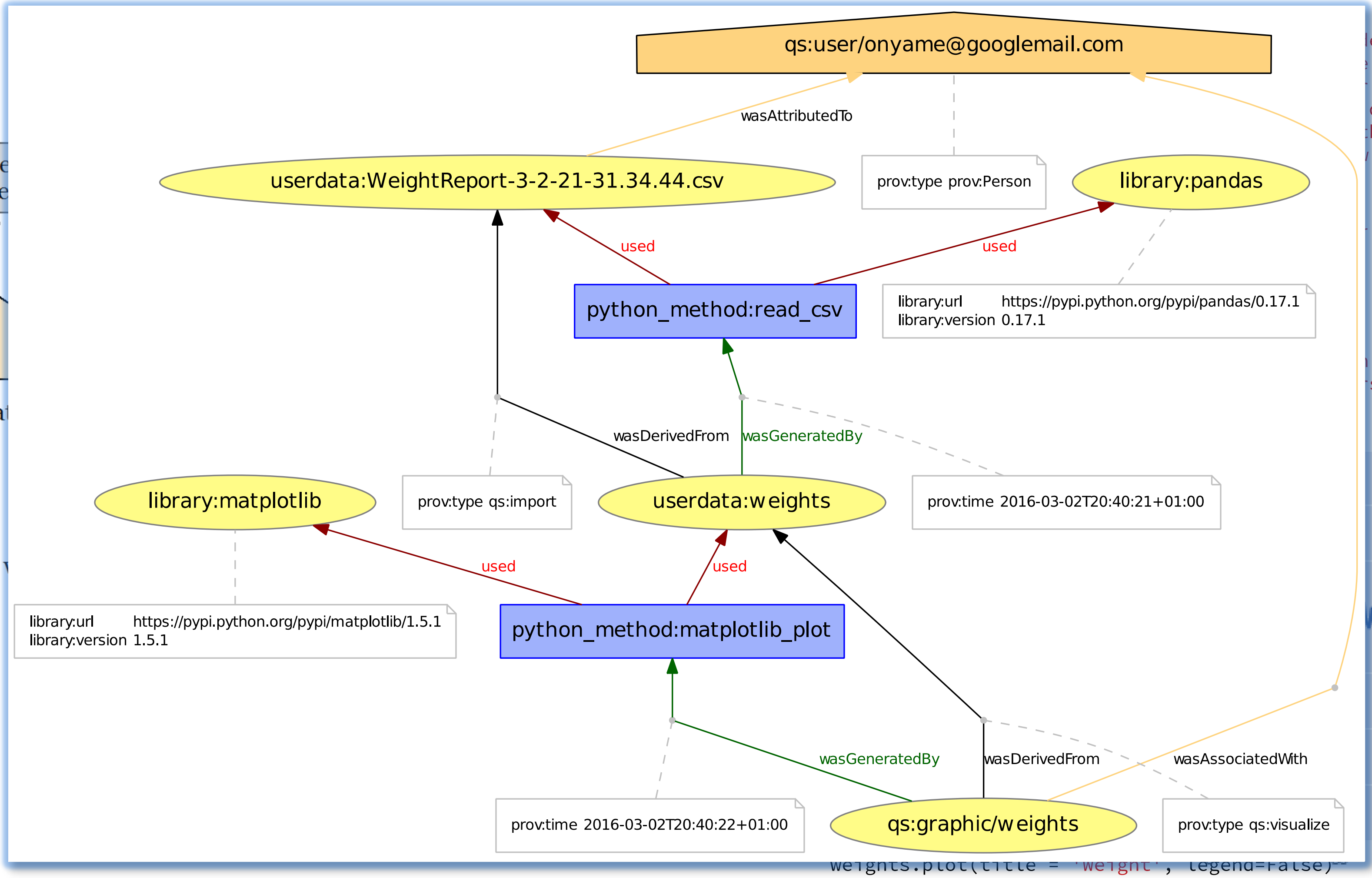
Weight Tracking Example

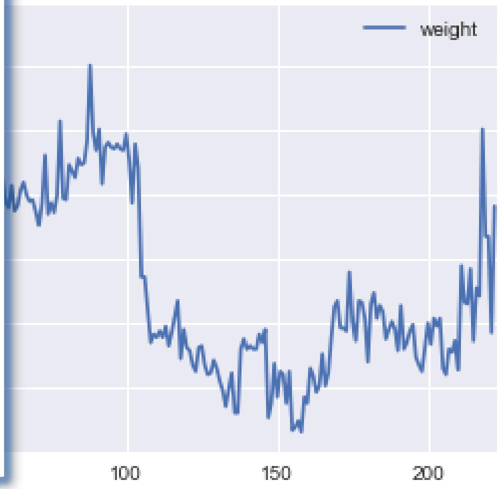
Visualizing Data with a Python Script

```
# Provenance-related Imports
from prov.model import ProvDocument, PROV
from provstore.api import Api
from time import gmtime, strftime

# Create a new provenance document
prov = ProvDocument()
```

```
e/qs/')
.dlr.de/qs/userdata/')
.de/qs/user/')
dlr.de/qs/graphic/')
hon.org/pypi/')
.python.org')

.com', {'prov:type': PROV['Person']})

': pd.__version__)
sion': matplotlib.__version__)
```



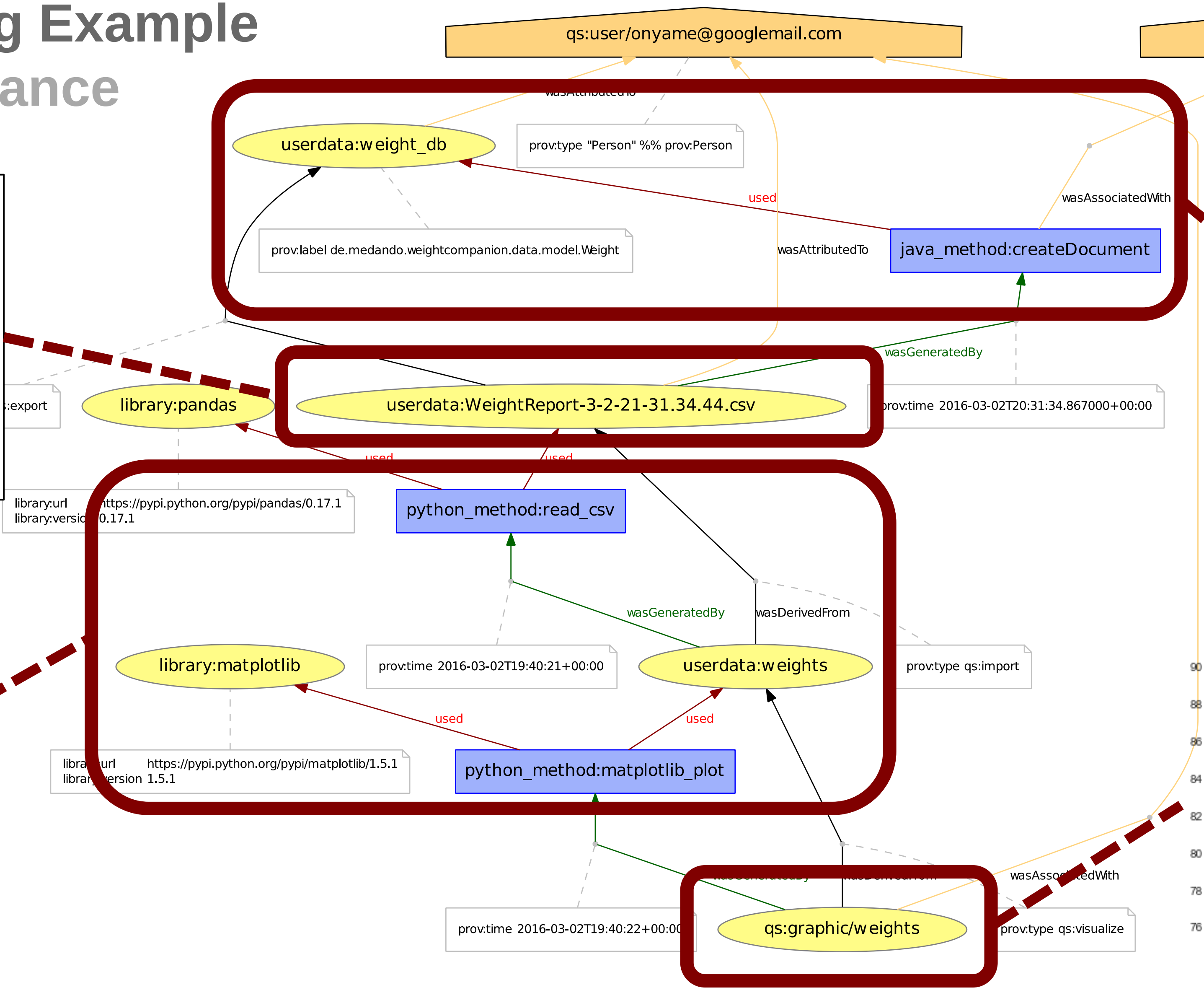




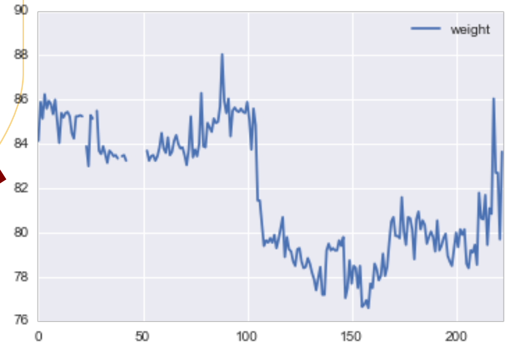

```
weights.plot(title = 'weight', legend=False)
entity_plot_weights = prov.entity('graphic:weights')
```

Weight Tracking Example
Merged Provenance

```
Date,Time,Weight,Waist,Hip,Device,Comment
"Jun 13, 2012",14:00,83.7,,,Withings,
"Jun 13, 2012",14:08,79.7,,,Withings,
"Jun 15, 2012",21:59,82.7,,,Withings,
"Jun 15, 2012",22:04,82.7,,,Withings,
"Jun 24, 2012",18:32,86.1,,,Withings,
"Jun 26, 2012",07:42,80.8,,,Withings,
"Jun 27, 2012",07:40,81.1,,,Withings,
"Jun 29, 2012",07:34,79.4,,,Withings,
"Jun 30, 2012",22:12,81.7,,,Withings,
"Jul 1, 2012",11:21,80.6,,,Withings,
"Jul 7, 2012",17:04,80.7,,,Withings,
"Jul 10, 2012",07:46,81.8,,,Withings,
"Jul 11, 2012",07:32,78.6,,,Withings,
"Jul 12, 2012",07:26,79.4,,,Withings,
```




```
# Application Import
from pandas import DataFrame, Series, read_c
import matplotlib.pyplot as plt

prov.entity('library:pandas', {'library:vers
prov.entity('library:matplotlib', {'library:

# Import weights from CSV file
WC_FILE = 'WeightReport-3-2-21-31.34.44.csv'
entity_csvfile = prov.entity('userdata:%s' %
prov.wasAttributedTo(entity_csvfile, agent_us
wc_data = read_csv(WC_FILE, parse_dates={'Ti
                   index_col='Timestamp')

# Get just the weights
weights = wc_data[['Weight']]
entity_weights = prov.entity('userdata:weigh
prov.activity('python_method:read_csv')
prov.wasGeneratedBy(entity_weights, 'python_
    strftime('%Y%m%dT%H%M%S%Z', gmtime()))
prov.used('python_method:read_csv', entity_c
prov.wasDerivedFrom(entity_weights, entity_c
    other_attributes={'prov:type': 'qs:impor
prov.used('python_method:read_csv', 'library

# Plot the weights
weights.plot(title = 'Weight', legend=False)
entity_plot_weights = prov.entity('graphic:w
```



Conclusions

Summary

- We collected QS workflows at QS Meetups and QS Conferences
- We used the *standardized data model PROV* for modeling the provenance of QS and medical workflows
- We evaluated the feasibility with selected QS workflows

Intentionally not mentioned in this talk

- *Storing provenance*
- *Querying provenance*

Current and Future Work

Provenance Templates and Standardization

- PROV templates for developer of tools and services
- Guidelines and APIs for developers
- Writing a specification

Provenance Analytics and Visualization

- Traceability of data
- Reproducibility of studies
- Visualization

Thank You!

Questions?

Andreas.Schreiber@dlr.de

www.DLR.de/sc | @onyame