



ĐẠI HỌC BÁCH KHOA HÀ NỘI  
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

# Mashup-based Linked Data Integration

# Agenda

1. Introduction
2. Linked Widgets Framework
3. Mashup Models
4. Conclusion and Future Work

# 1. Introduction

# Motivation

**CSV**(lat, long, name, etc.)

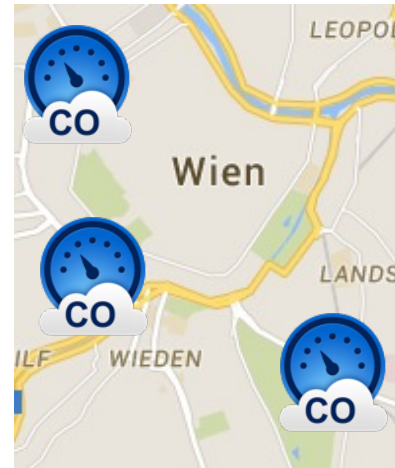
**XML**(lat, long, Image URL)

**JSON**(lat, long, CO, NO<sub>2</sub>, etc.)

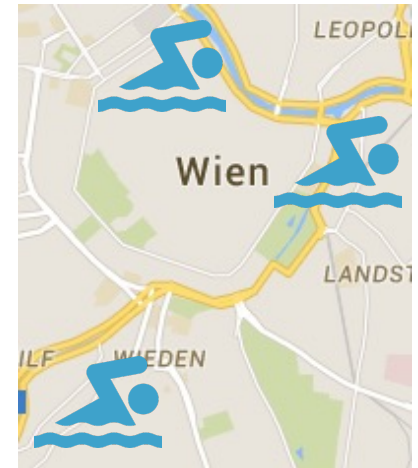
**CSV**(lat, long, name, etc.)



Flickr/Google  
Image Search



Air quality data



51 public swimming  
pools in Vienna

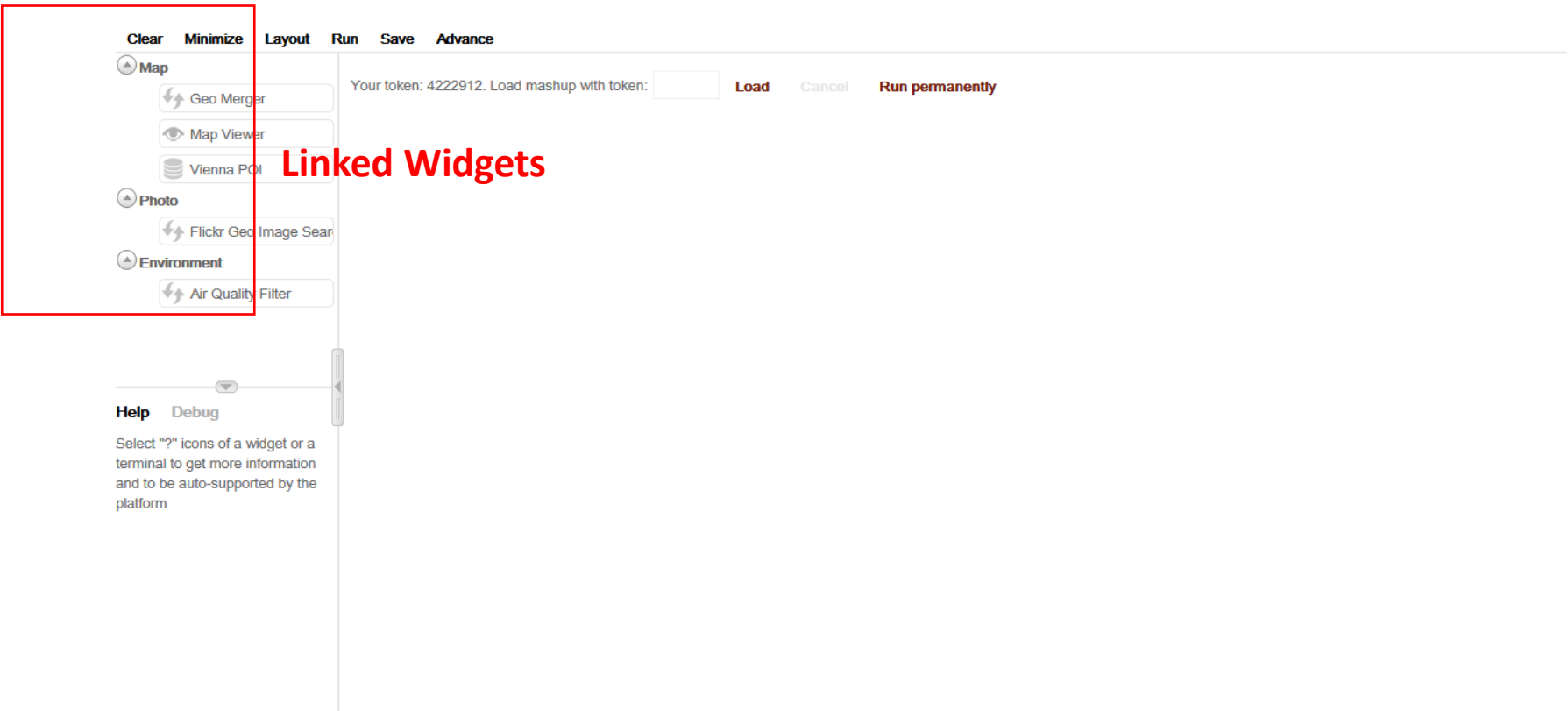


?

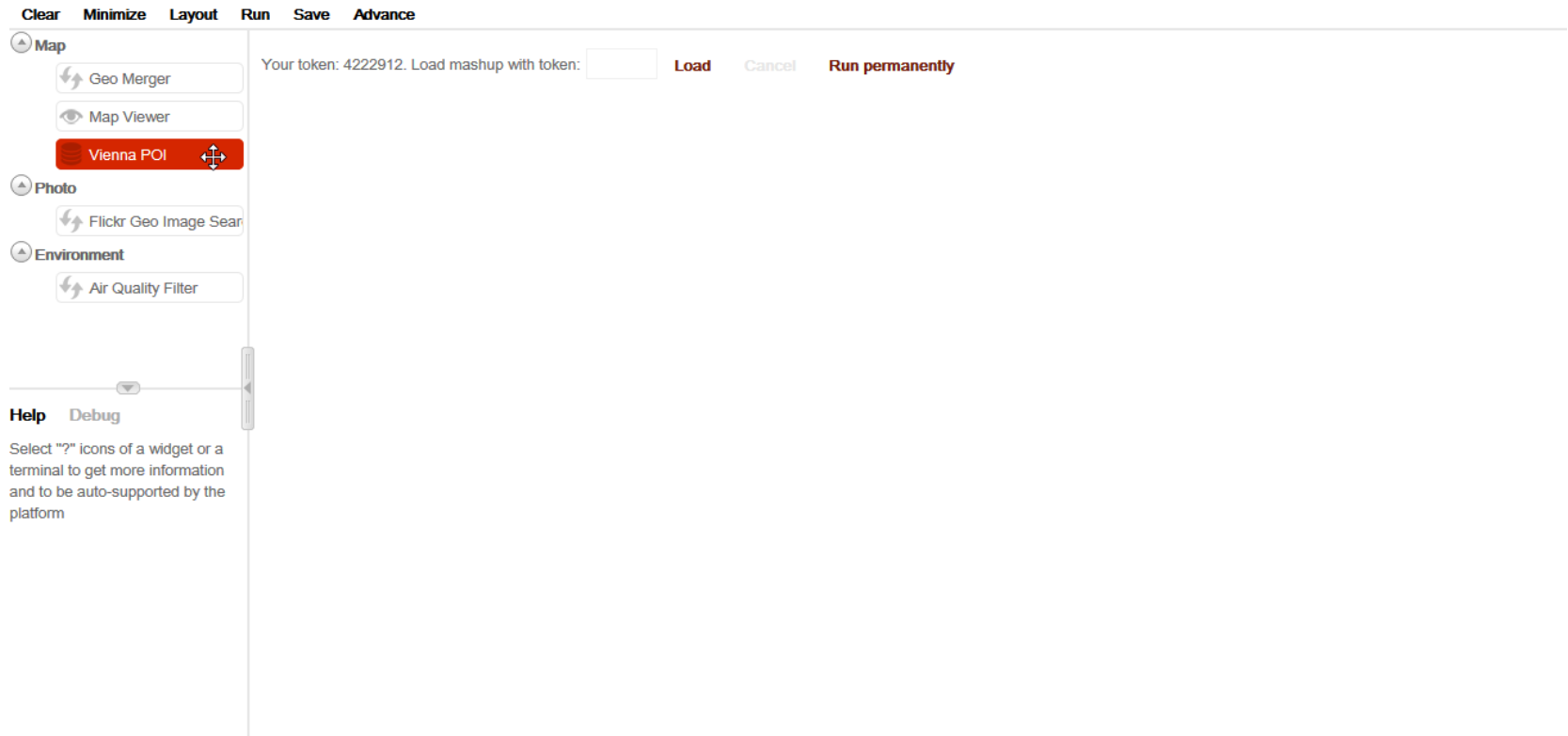
# Motivation



# Modular approach for data integration

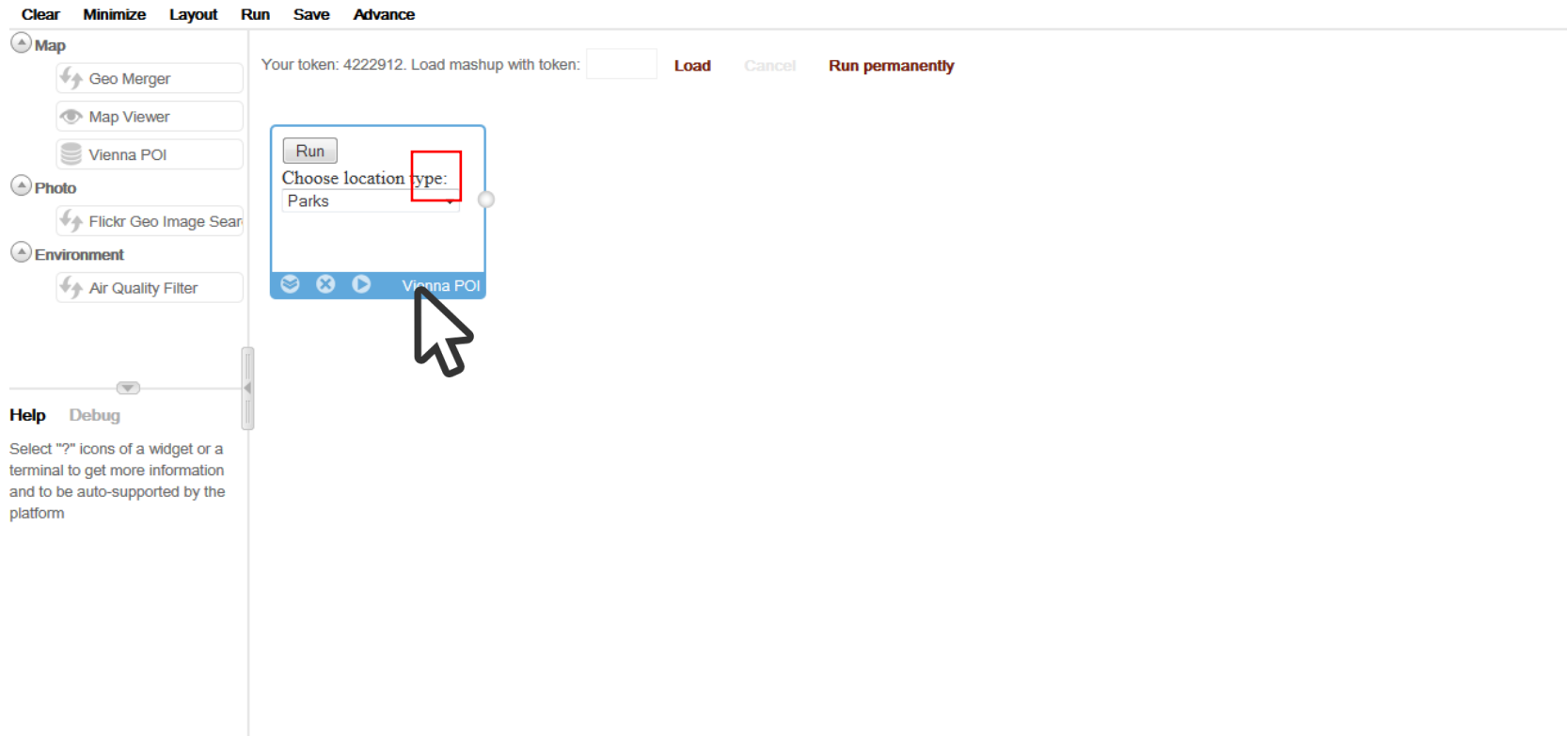


# Modular approach for data integration





# Modular approach for data integration





# Modular approach for data integration

The screenshot displays a web-based mashup development environment. At the top, a menu bar includes 'Clear', 'Minimize', 'Layout', 'Run', 'Save', and 'Advance'. On the left sidebar, there are three main categories: 'Map' (containing 'Geo Merger', 'Map Viewer', and 'Vienna POI'), 'Photo' (containing 'Flickr Geo Image Search'), and 'Environment' (containing 'Air Quality Filter'). Below these is a 'Help' section with a 'Debug' button and a text box explaining that '?' icons provide more information and auto-support. The main workspace shows a 'Run' button, a 'Choose location type:' dropdown menu with 'Parks' selected, and a 'Vienna POI' button. Above the workspace, a text field displays 'Your token: 4222912. Load mashup with token:' followed by 'Load', 'Cancel', and 'Run permanently' buttons.

# Modular approach for data integration

Clear Minimize Layout Run Save Advance

Map

- Geo Merger
- Map Viewer
- Vienna POI

Photo

- Flickr Geo Image Search

Environment

- Air Quality Filter

Your token: 4222912. Load mashup with token:  Load Cancel Run permanently

Run

Choose location type:

Parks

Run

Minimum Level:

Very good

Maximum Level:

None

Vienna POI

Air Quality Filter

Help Debug

Select "?" icons of a widget or a terminal to get more information and to be auto-supported by the platform

# Modular approach for data integration

The screenshot displays a web-based data integration platform interface. At the top, there is a navigation bar with buttons: Clear, Minimize, Layout, Run, Save, and Advance. Below this, a sidebar on the left contains a list of widgets categorized under 'Map', 'Photo', and 'Environment'. The 'Map' category includes 'Geo Merger', 'Map Viewer', and 'Vienna POI'. The 'Photo' category includes 'Flickr Geo Image Search'. The 'Environment' category includes 'Air Quality Filter'. A 'Help' section at the bottom left explains that '?' icons provide more information and auto-support. The main workspace shows a workflow with two widgets: 'Vienna POI' and 'Air Quality Filter'. The 'Vienna POI' widget has a 'Run' button and a 'Choose location type:' dropdown menu with 'Parks' selected. The 'Air Quality Filter' widget has a 'Run' button, a 'Minimum Level:' dropdown menu with 'Very good' selected, and a 'Maximum Level:' dropdown menu with 'None' selected. A mouse cursor is pointing at the 'Parks' option in the 'Choose location type:' dropdown. Above the widgets, there is a text field for 'Your token: 4222912. Load mashup with token:' and buttons for 'Load', 'Cancel', and 'Run permanently'.

Clear Minimize Layout Run Save Advance

Map

- Geo Merger
- Map Viewer
- Vienna POI

Photo

- Flickr Geo Image Search

Environment

- Air Quality Filter

Help Debug

Select "?" icons of a widget or a terminal to get more information and to be auto-supported by the platform

Your token: 4222912. Load mashup with token:  Load Cancel Run permanently

Run

Choose location type:

Parks

Run

Minimum Level:

Very good

Maximum Level:

None

Vienna POI

Air Quality Filter

# Modular approach for data integration

The screenshot displays a web-based interface for data integration. At the top, there is a menu bar with options: Clear, Minimize, Layout, Run, Save, and Advance. Below the menu, a sidebar on the left contains a list of widgets categorized under 'Map', 'Photo', and 'Environment'. The 'Map' category includes 'Geo Merger', 'Map Viewer', and 'Vienna POI'. The 'Photo' category includes 'Flickr Geo Image Search'. The 'Environment' category includes 'Air Quality Filter'. The main workspace shows a workflow with two widgets: 'Vienna POI' and 'Air Quality Filter'. The 'Vienna POI' widget has a 'Run' button and a 'Choose location type:' dropdown menu set to 'Parks'. The 'Air Quality Filter' widget has a 'Run' button, a 'Minimum Level:' dropdown menu set to 'Very good', and a 'Maximum Level:' dropdown menu set to 'None'. A blue line connects the output of the 'Vienna POI' widget to the input of the 'Air Quality Filter' widget. At the top of the main workspace, there is a text field for 'Your token: 4222912. Load mashup with token:' and buttons for 'Load', 'Cancel', and 'Run permanently'. At the bottom left, there is a 'Help' and 'Debug' section with a message: 'Select "?" icons of a widget or a terminal to get more information and to be auto-supported by the platform'.

# Modular approach for data integration

Clear Minimize Layout Run Save Advance

Map

- Geo Merger
- Map Viewer
- Vienna POI

Photo

- Flickr Geo Image Search

Environment

- Air Quality Filter

Help Debug

Select "?" icons of a widget or a terminal to get more information and to be auto-supported by the platform

Your token: 4222912. Load mashup with token:  Load Cancel Run permanently

Run

Choose location type:  
Parks

Run

Minimum Level:  
Very good

Maximum Level:  
None

Run

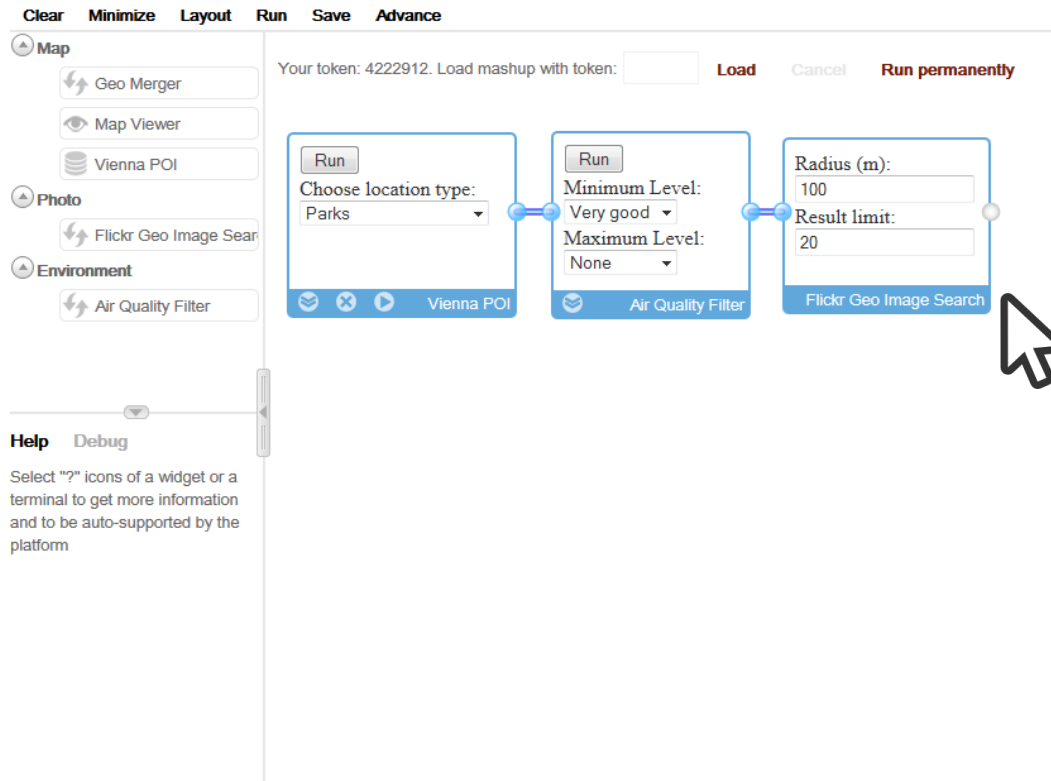
Radius (m):  
100

Result limit:  
20

Vienna POI

Air Quality Filter

Flickr Geo Image Search



# Modular approach for data integration

Clear Minimize Layout Run Save Advance

Map

- Geo Merger
- Map Viewer
- Vienna POI**

Photo

- Flickr Geo Image Search

Environment

- Air Quality Filter

Your token: 4222912. Load mashup with token:  **Load** **Cancel** **Run permanently**

Run

Choose location type:  
Parks

Run

Minimum Level:  
Very good

Maximum Level:  
None

Radius (m):  
100

Result limit:  
20

Vienna POI Air Quality Filter Flickr Geo Image Search

Help Debug

Select "?" icons of a widget or a terminal to get more information and to be auto-supported by the platform

# Modular approach for data integration

Clear Minimize Layout Run Save Advance

Map

- Geo Merger
- Map Viewer
- Vienna POI

Photo

- Flickr Geo Image Search

Environment

- Air Quality Filter

Your token: 4222912. Load mashup with token:  Load Cancel Run permanently

Run

Choose location type:  
Parks

Run

Minimum Level:  
Very good

Maximum Level:  
None

Radius (m):  
100

Result limit:  
20

Vienna POI

Air Quality Filter

Flickr Geo Image Search

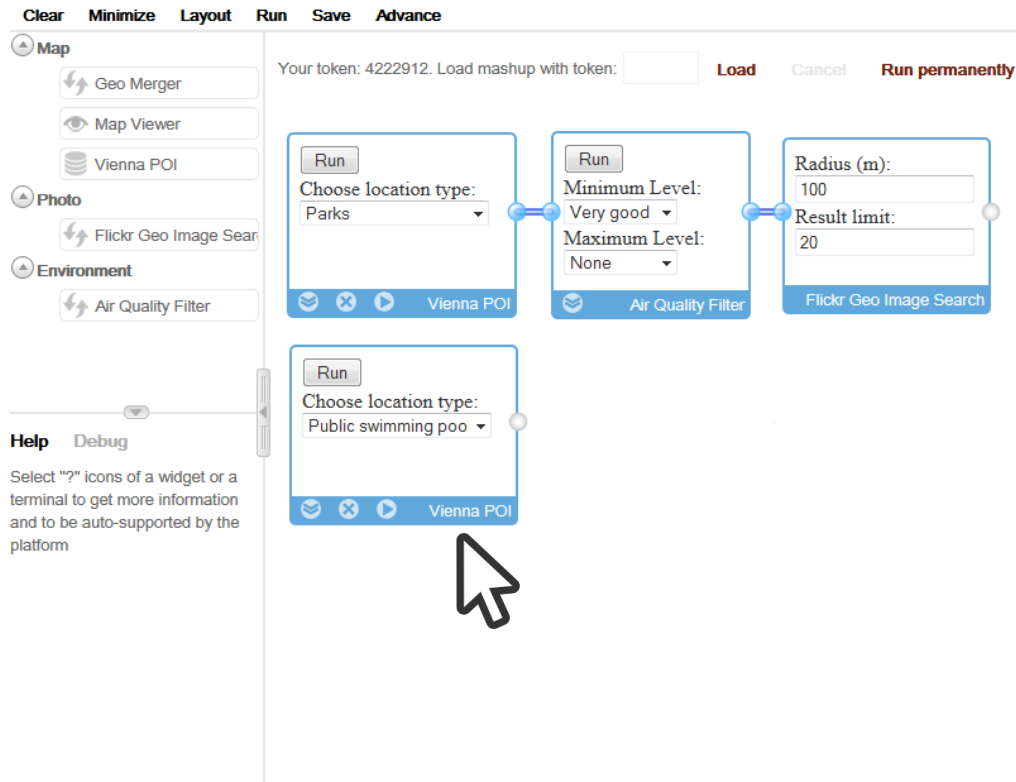
Run

Choose location type:  
Public swimming pool

Vienna POI

Help Debug

Select "?" icons of a widget or a terminal to get more information and to be auto-supported by the platform





# Modular approach for data integration

Clear Minimize Layout Run Save Advance

Map

Geo Merger

Map Viewer

Vienna POI

Photo

Flickr Geo Image Search

Environment

Air Quality Filter

Your token: 4222912. Load mashup with token:  Load Cancel Run permanently

Run

Choose location type:

Parks

Run

Minimum Level:

Very good

Maximum Level:

None

Run

Radius (m):

100

Result limit:

20

Run

Choose location type:

Public swimming pool

Vienna POI

Air Quality Filter

Flickr Geo Image Search

Help Debug

Select "?" icons of a widget or a terminal to get more information and to be auto-supported by the platform

# Modular approach for data integration

Clear Minimize Layout Run Save Advance

Map

- Geo Merger
- Map Viewer
- Vienna POI

Photo

- Flickr Geo Image Search

Environment

- Air Quality Filter

Your token: 4222912. Load mashup with token:  Load Cancel Run permanently

Run

Choose location type:  
Parks

Run

Minimum Level:  
Very good

Maximum Level:  
None

Run

Radius (m):  
100

Result limit:  
20

Run

Choose location type:  
Public swimming pool

Run

Maximum distances:  
200 meters

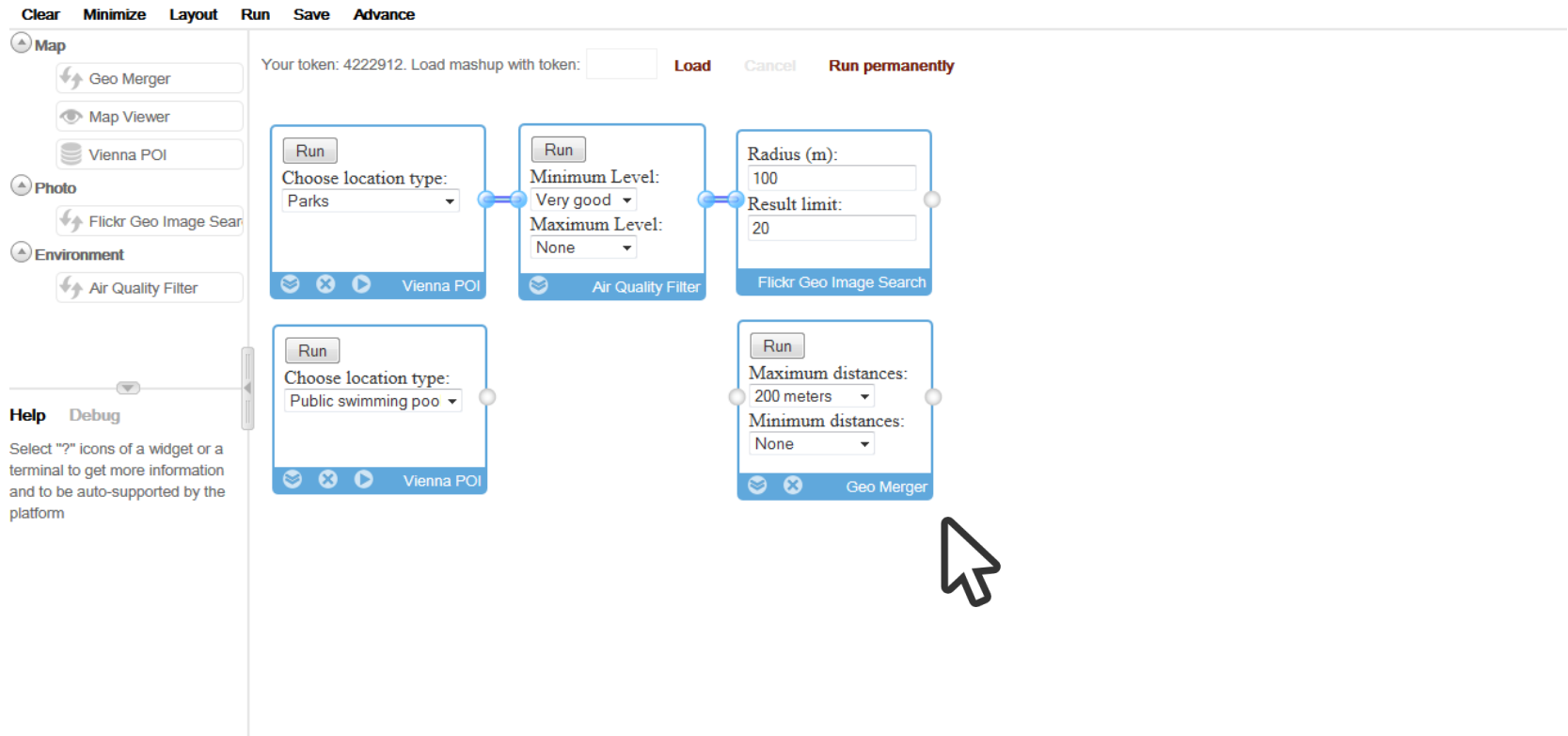
Minimum distances:  
None

Run

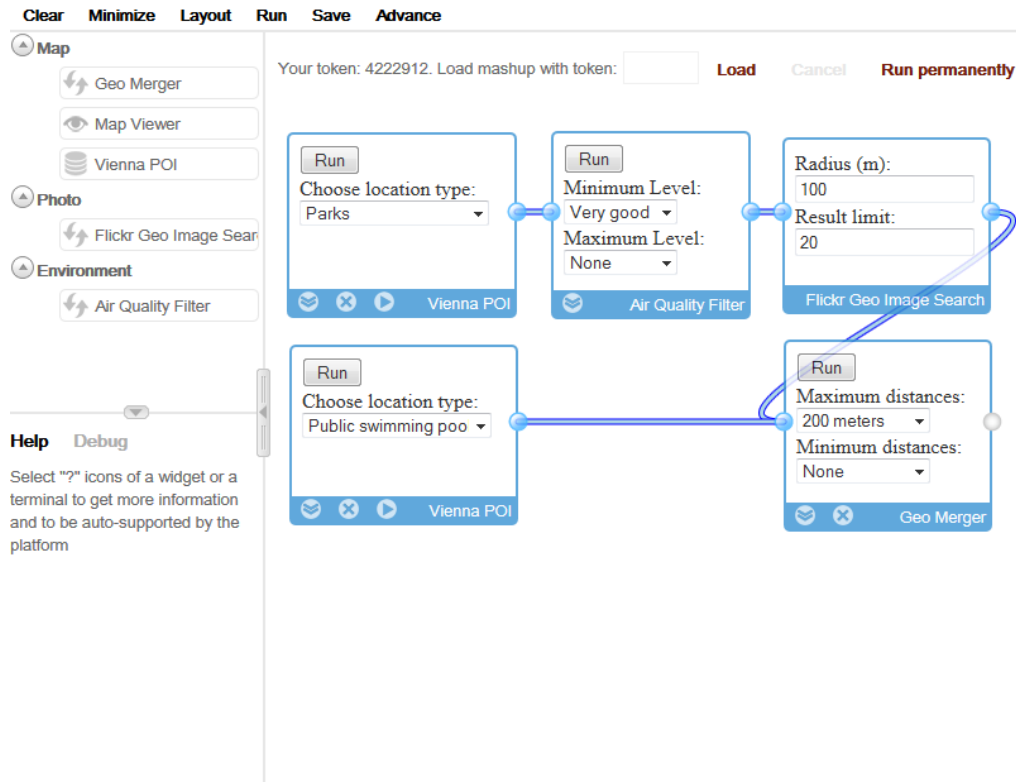
Geo Merger

Help Debug

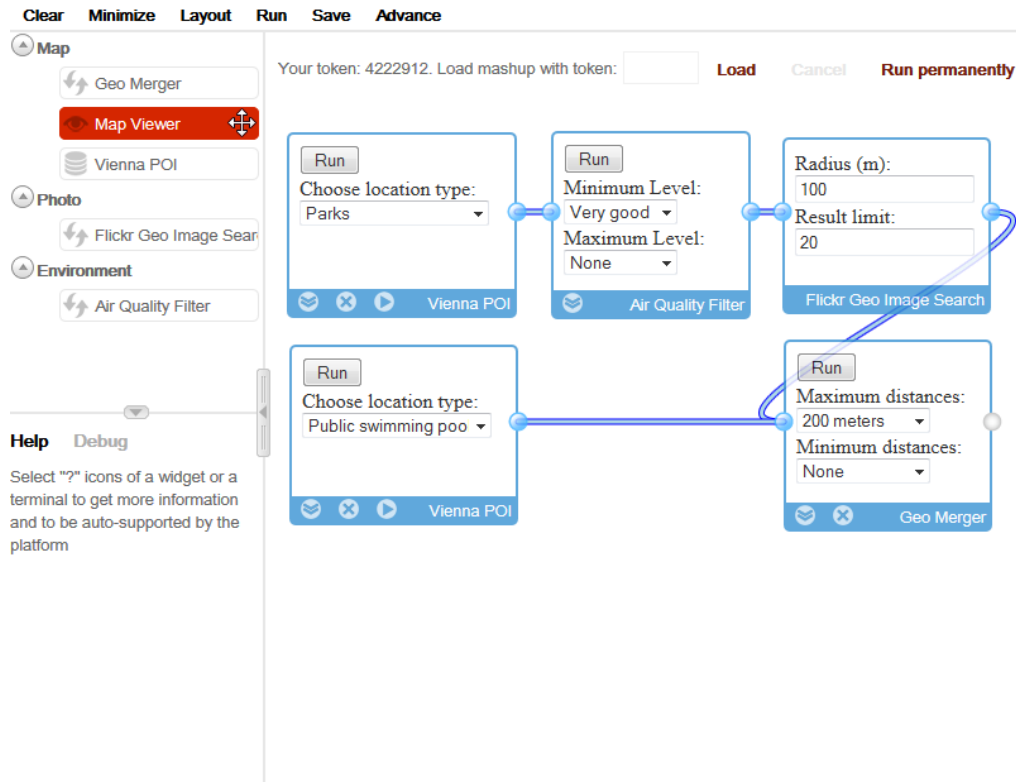
Select "?" icons of a widget or a terminal to get more information and to be auto-supported by the platform



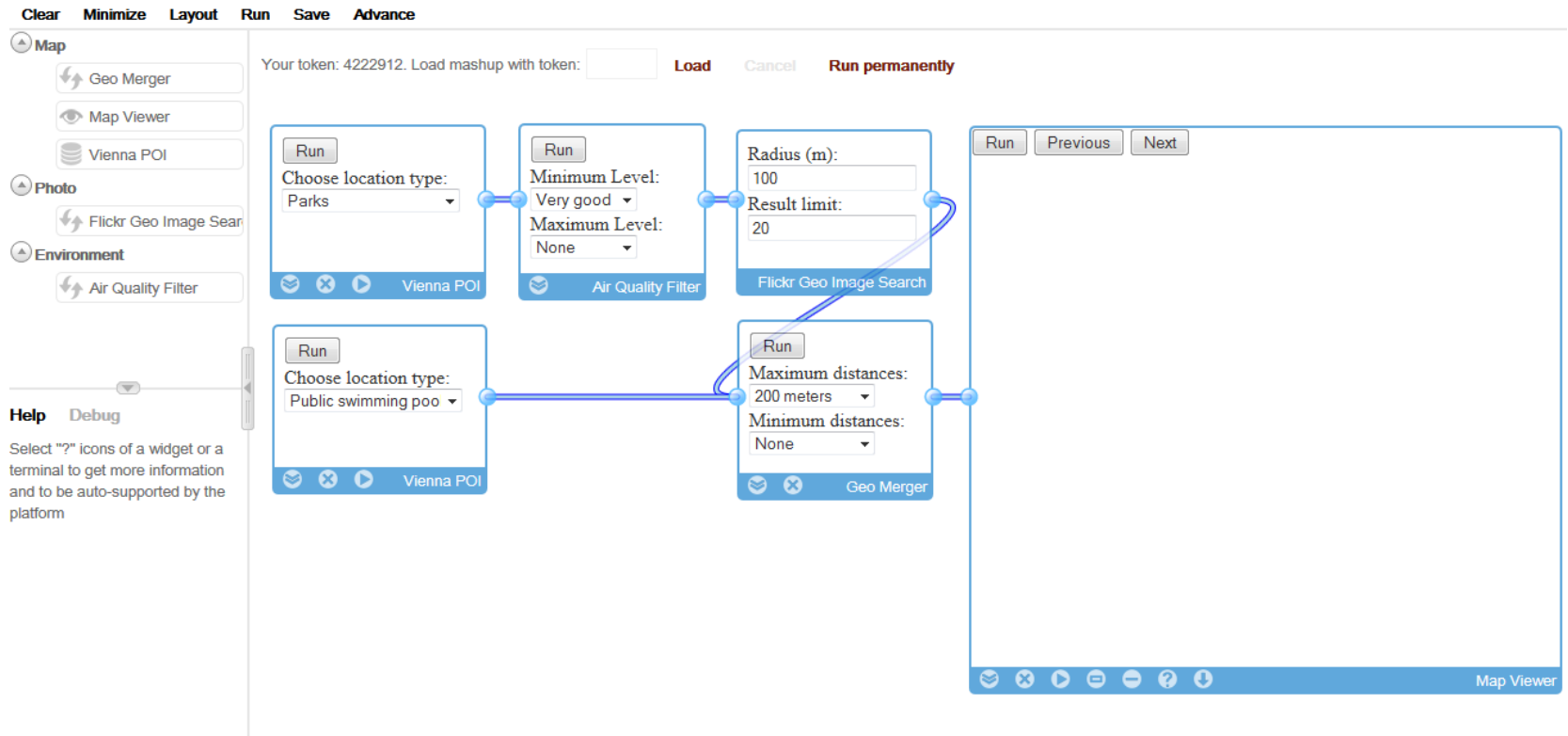
# Modular approach for data integration



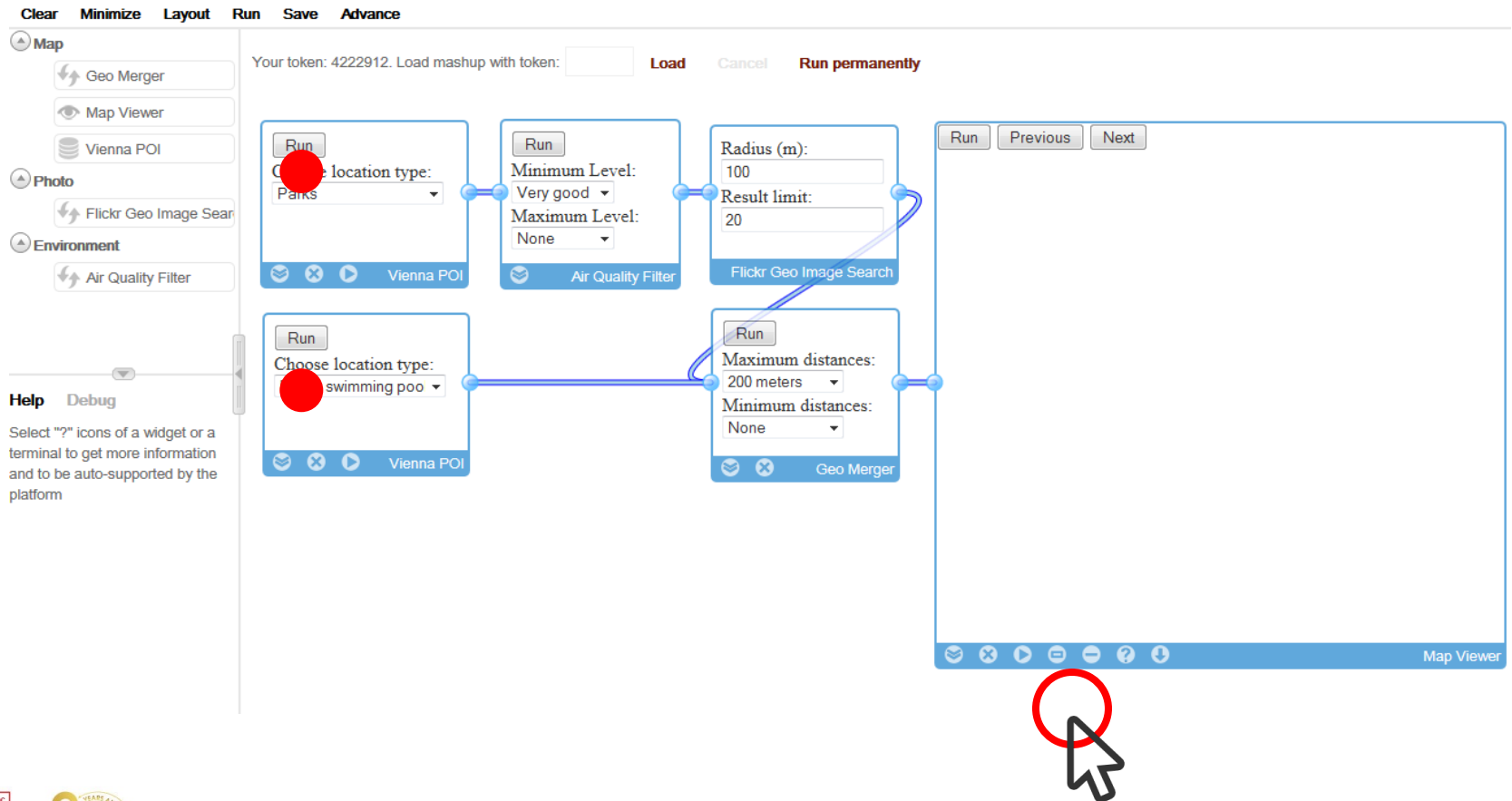
# Modular approach for data integration



# Modular approach for data integration



# Modular approach for data integration



# Modular approach for data integration

Clear Minimize Layout Run Save Advance

Map

- Geo Merger
- Map Viewer
- Vienna POI

Photo

- Flickr Geo Image Search

Environment

- Air Quality Filter

Help Debug

Select "?" icons of a widget or a terminal to get more information and to be auto-supported by the platform

Your token: 4222912. Load mashup with token:  Load Cancel Run permanently

Run

Choose location type: Parks

Run

Minimum Level: Very good

Maximum Level: None

Run

Radius (m): 100

Result limit: 20

Run

Choose location type: Public swimming pool

Run

Maximum distances: 200 meters

Minimum distances: None

Run

Left Top Bottom Right Full Map

@type: "http://ogd.ifs.tuwien.ac.at/vienna/PARKANLAGEOGD"  
@id: "http://ogd.ifs.tuwien.ac.at/vienna/PARKANLAGEOGD.74802"  
location: Object  
@type: "http://www.w3.org/2003/01/geo/wgs84\_pos#Point"  
lat: "48.179088632252956"  
long: "16.317551227631565"  
address: "Schönbrunner Schloßstraße 47, 1130 Wien, Austria"  
airQuality: 0.8257720583919212

Map Viewer



# Mashup-based data integration

- Mashups: innovative paradigm that “*combines data from multiple sources into an **integrated** and **single graphical interface***” [1]
  - are quick, flexible, and cost-effective
  - turn users from content consumers to content providers

# Challenges of mashup-based data integration research

1. Collaborative work among **end-user**, **data publisher** and **developer** communities is not considered
2. Integrating data that is **distributed** in different devices and not available on the web is not considered
3. **Semantic mashup** – “*A data mashup using RDF(S) as data model*” [2] – is still in its early stage

# Research Question

*How can **non-expert** users be enabled to explore and integrate **heterogeneous** data sources?*

**RQ1:** *How is it possible to support non-expert users in addressing data **heterogeneity**?*

**RQ2:** *How can non-expert users be enabled to **collaboratively** integrate data?*

**RQ3:** *How is it possible to **automate** the data exploration and integration process?*

## 2. Linked Widgets Framework

# Linked Widgets

- extends W3C standard web widgets with a semantic model
- lift data sources to a semantic level to facilitate data processing
- can be created by **independent** developers, hosted on different servers

## Processing function

### Semantic model

User Interface

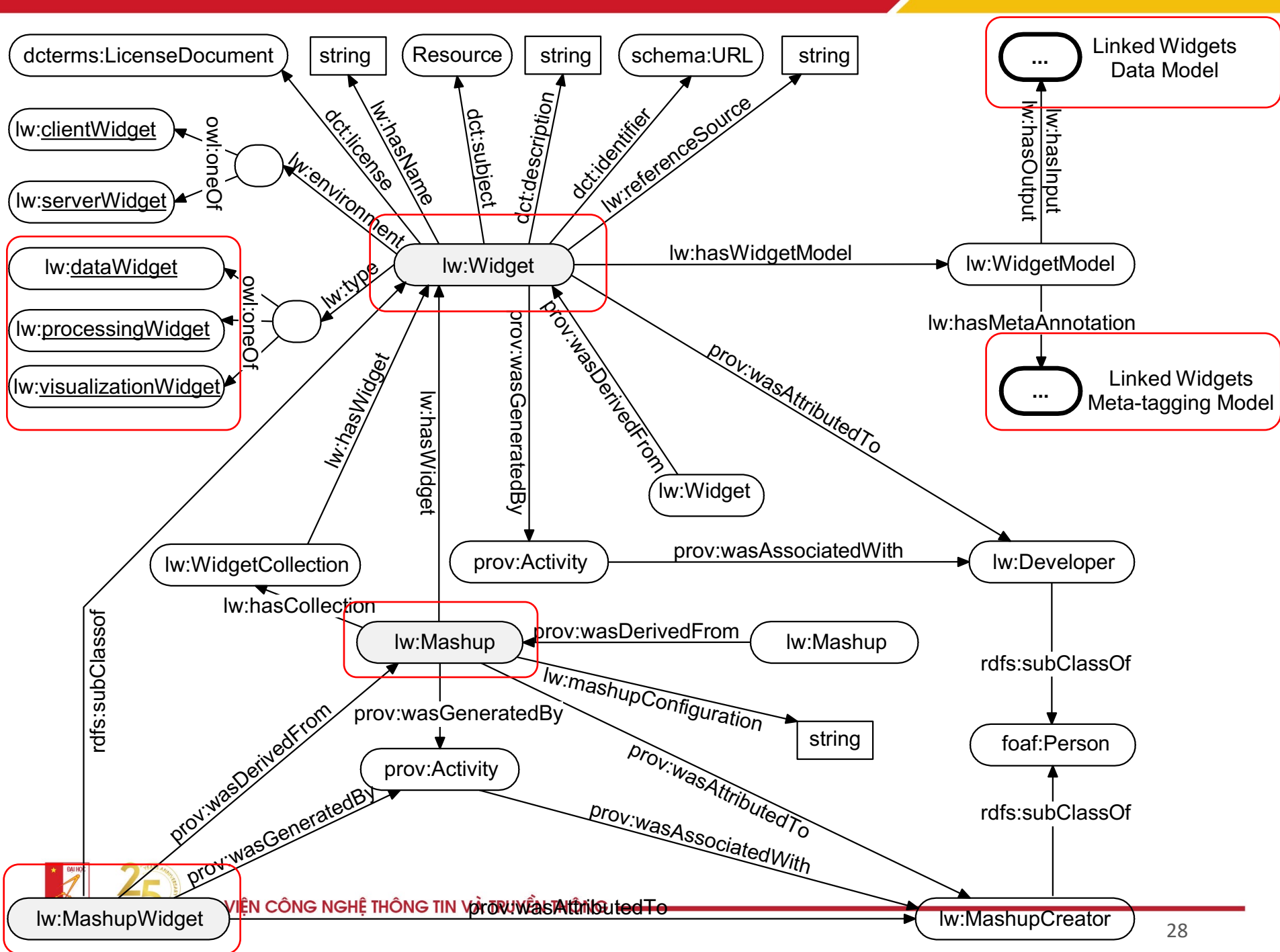
Core Interface

The diagram shows a widget titled "Point Of Interest Search". It has a "POI type:" input field with the value "park" and a "Radius (km)" dropdown menu with options: "MotorcycleParking", "NationalPark", "Park" (selected), "ParkBench", and "Parking". The widget has a blue border and a blue header bar with icons for search, close, play, and a "Point Of Interest Search" label. A red box highlights the "POI type:" and "Radius (km)" fields, labeled "Semantic model". A red arrow points to the "POI type:" field, labeled "User Interface". A red arrow points to the blue header bar, labeled "Core Interface".

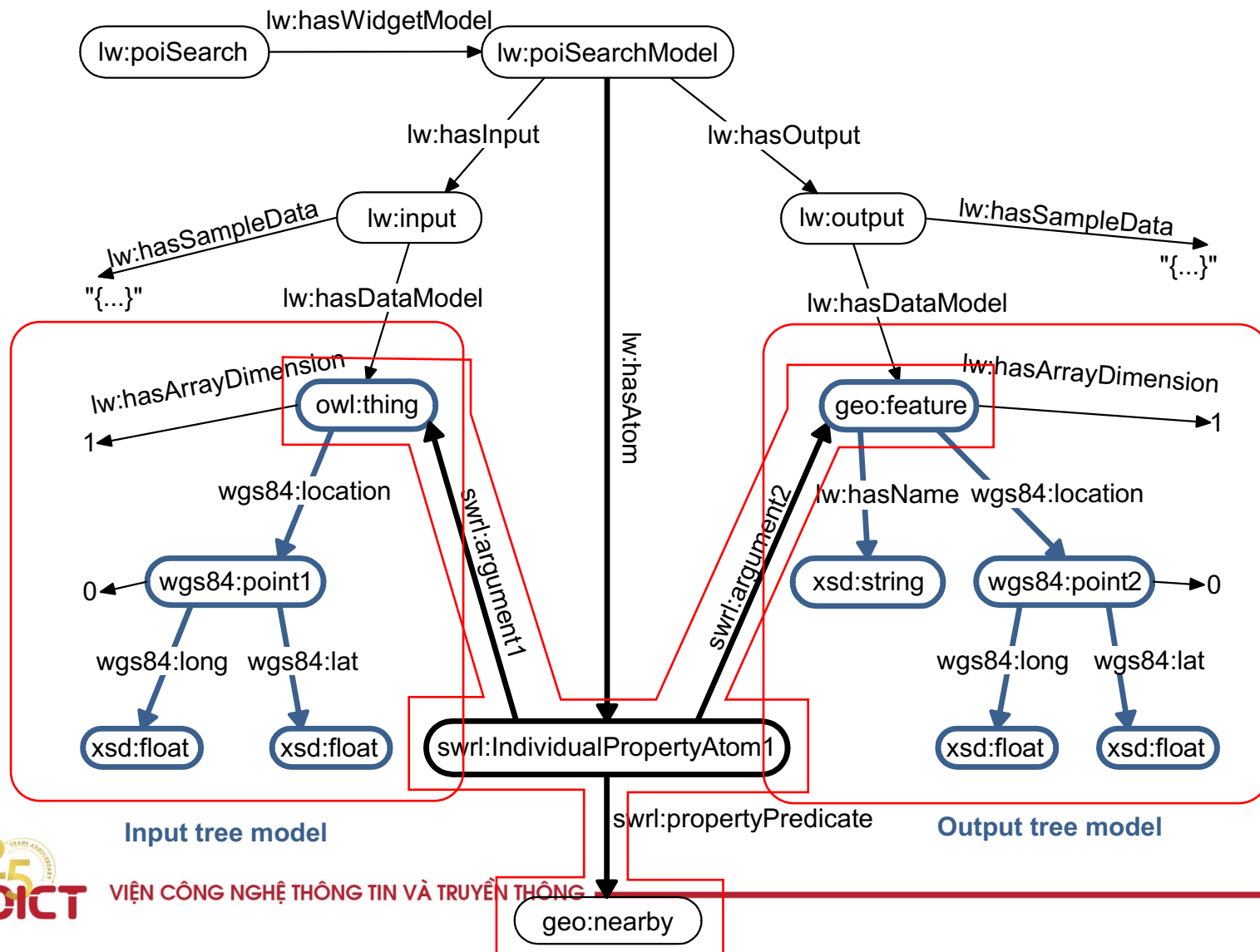
Client Linked Widgets  
run in a browser

The diagram shows a widget titled "Weather Observation". It has an "Indicator" section with three radio buttons: "Temperature" (selected), "Pressure", and "Wind". Below the indicators is a "Timer:" input field with the value "5 seconds" and a dropdown arrow. The widget has a green border and a green header bar with icons for search, close, play, and a "Weather Observation" label.

Server Linked Widgets run in a  
web server, PC, mobile phones,  
sensor, or embedded device



# Example: Data model of the POI search widget





# M1: Semantic Widget Discovery

Clear

Show previous results

Exact Search

Tolerant Search

☒ Short Name ☐ Full URI

To define widget's model, please drag & drop Data Model Relation and if necessary, use keywords

+ Terminal

Data Model

✗ Input

✗ Output

Type: Object of

+ with attributes & relations with others

location

Data Model

Type: Object of Point

+ with attributes & relations with others

lat

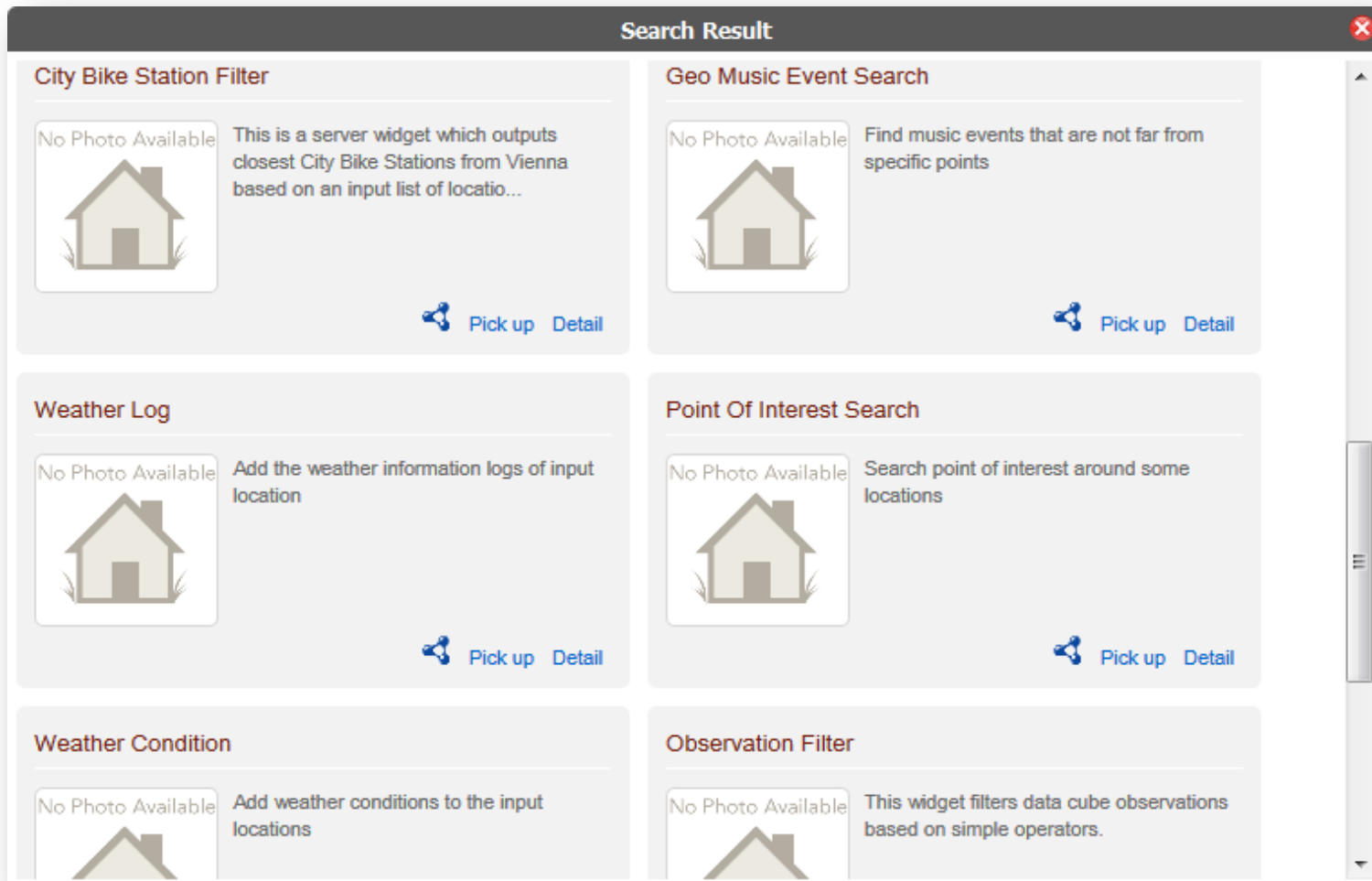
long

Data Model

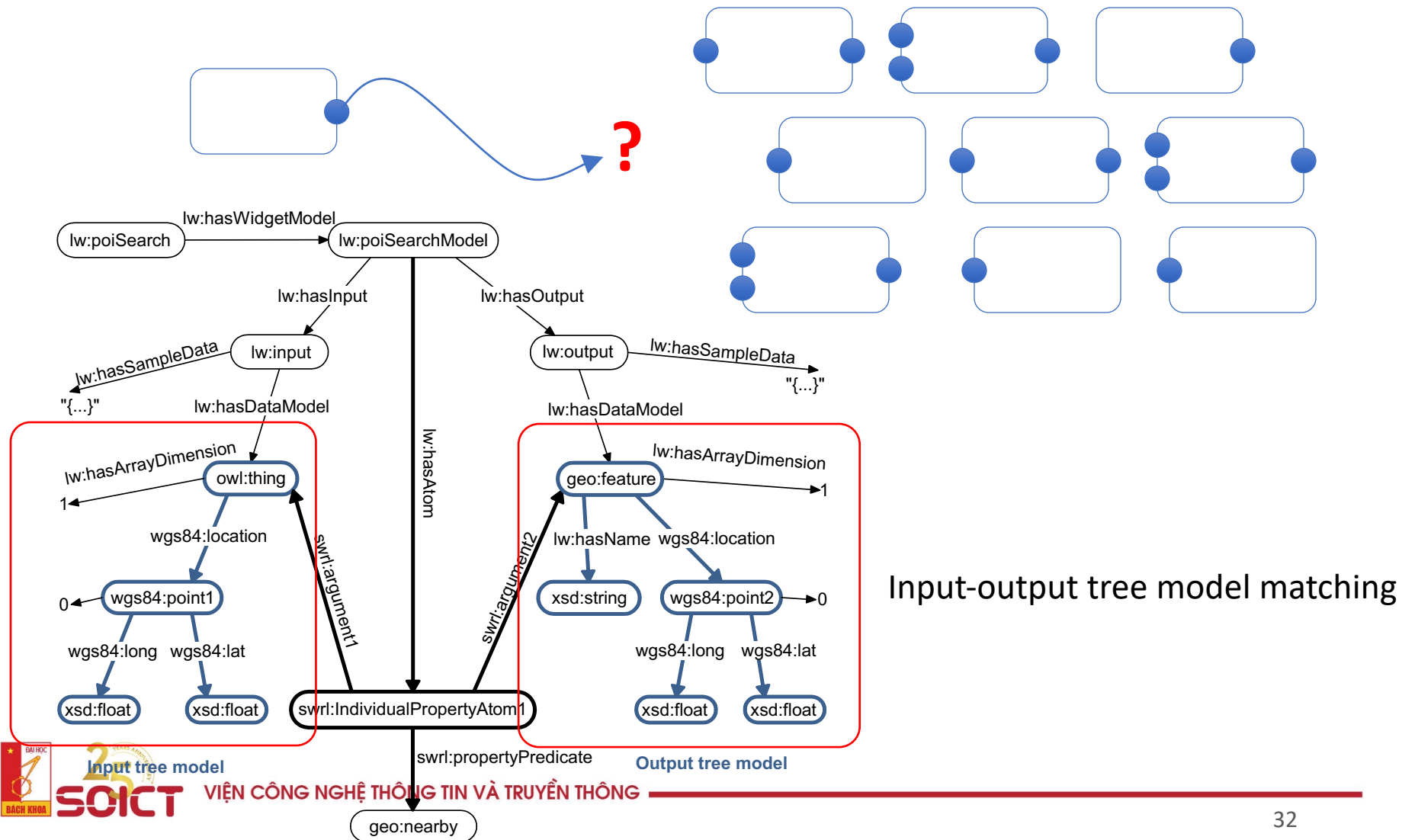
Ontology Alignment API:

<http://alignapi.gforge.inria.fr/>

# M1: Semantic Widget Discovery



# M2: Terminal matching

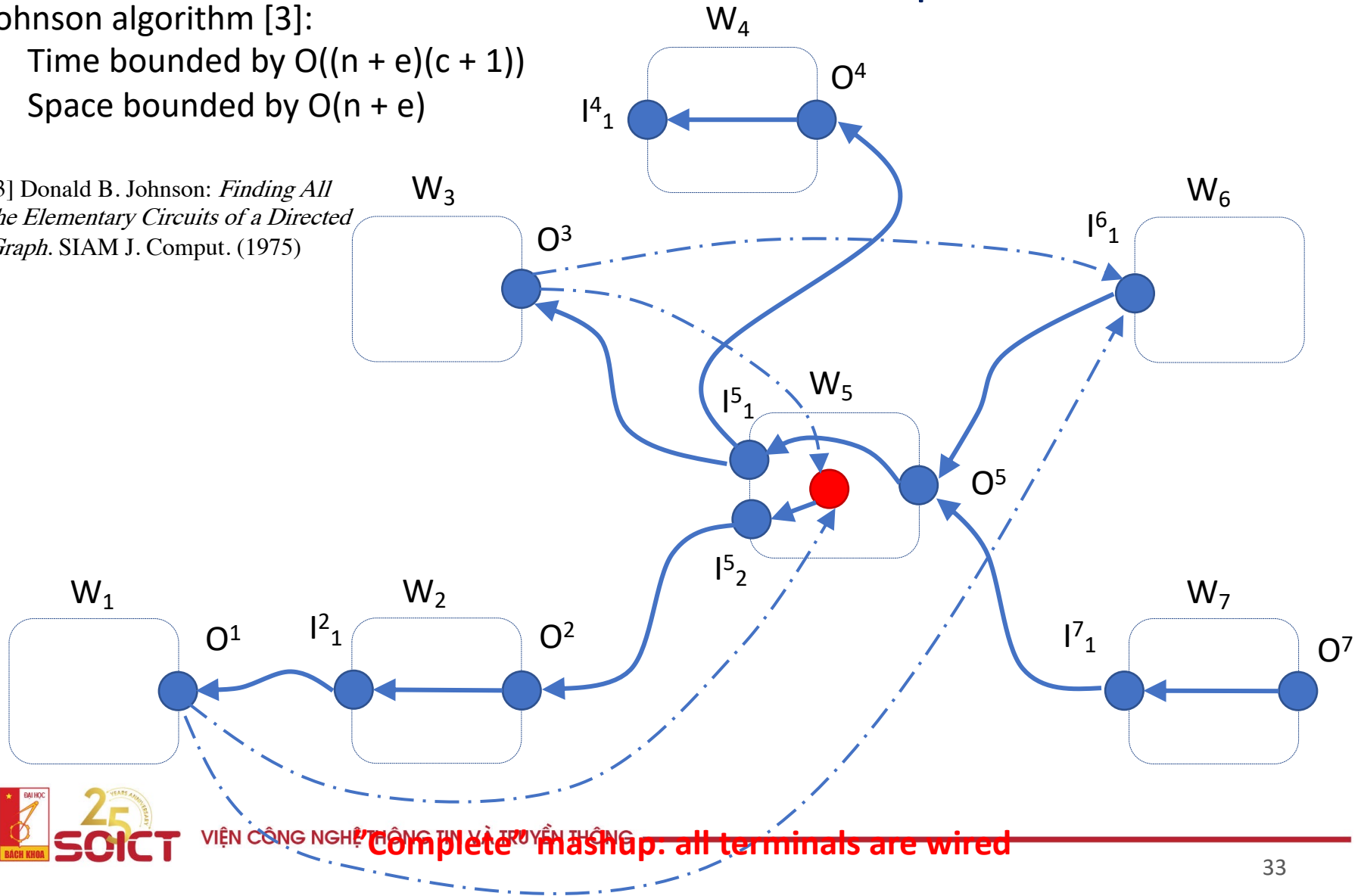


Reduce “automatic composition” to  
“list all cycles” algorithm

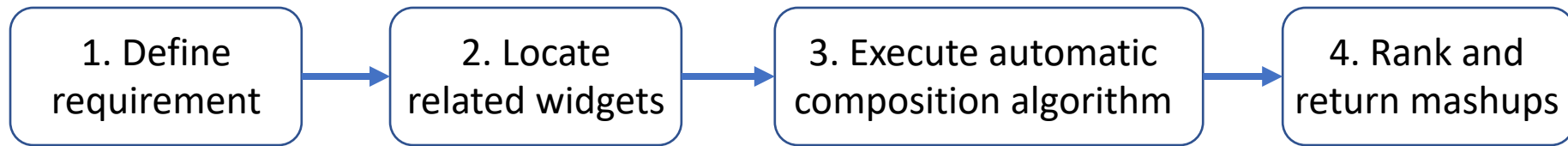
Johnson algorithm [3]:

- Time bounded by  $O((n + e)(c + 1))$
- Space bounded by  $O(n + e)$

[3] Donald B. Johnson: *Finding All the Elementary Circuits of a Directed Graph*. SIAM J. Comput. (1975)



# M4: Interactive tag-based composer



Enter text + tie words to  
DBpedia (Wikipedia) resources

**Restaurant nearby Bank**



<http://dbpedia.org/resource/Restaurant>

<http://www.geonames.org/ontology#nearby>

<http://dbpedia.org/resource/Bank>

A bank is a financial intermediary that creates credit by lending money to a borrower, thereby creating a corresponding deposit on the bank's balance sheet etc.

Benefits of the semantic tagging techniques

- No ambiguity
- Multiple language

# Define mashup requirement

☒ Free Style ☐ Syntax Style **Compose Mashup**

Restaurant nearby Bank

☒ Concept ☐ Relation ☐ Who ☐ Where ☐ When ☐ What

Bank (<https://en.wikipedia.org/wiki/Bank>)  
A bank is a financial institution that creates credit by lending money...

**Bank of America** ([https://en.wikipedia.org/wiki/Bank\\_of\\_America](https://en.wikipedia.org/wiki/Bank_of_America))  
Bank of America (abbreviated as BofA) is an American multinational bank...

**Bankruptcy** (<https://en.wikipedia.org/wiki/Bankruptcy>)  
Bankruptcy is a legal status of a person or other entity that cannot r...

**Bank and Monument stations** ([https://en.wikipedia.org/wiki/Bank\\_and\\_Monument\\_stations](https://en.wikipedia.org/wiki/Bank_and_Monument_stations))  
Bank and Monument are interlinked London Underground and Docklands Lig...

**Banksia** (<https://en.wikipedia.org/wiki/Banksia>)  
Banksia is a genus of around 170 species in the plant family Proteacea...

1. **Concept tagging**  
(DBpedia concepts)

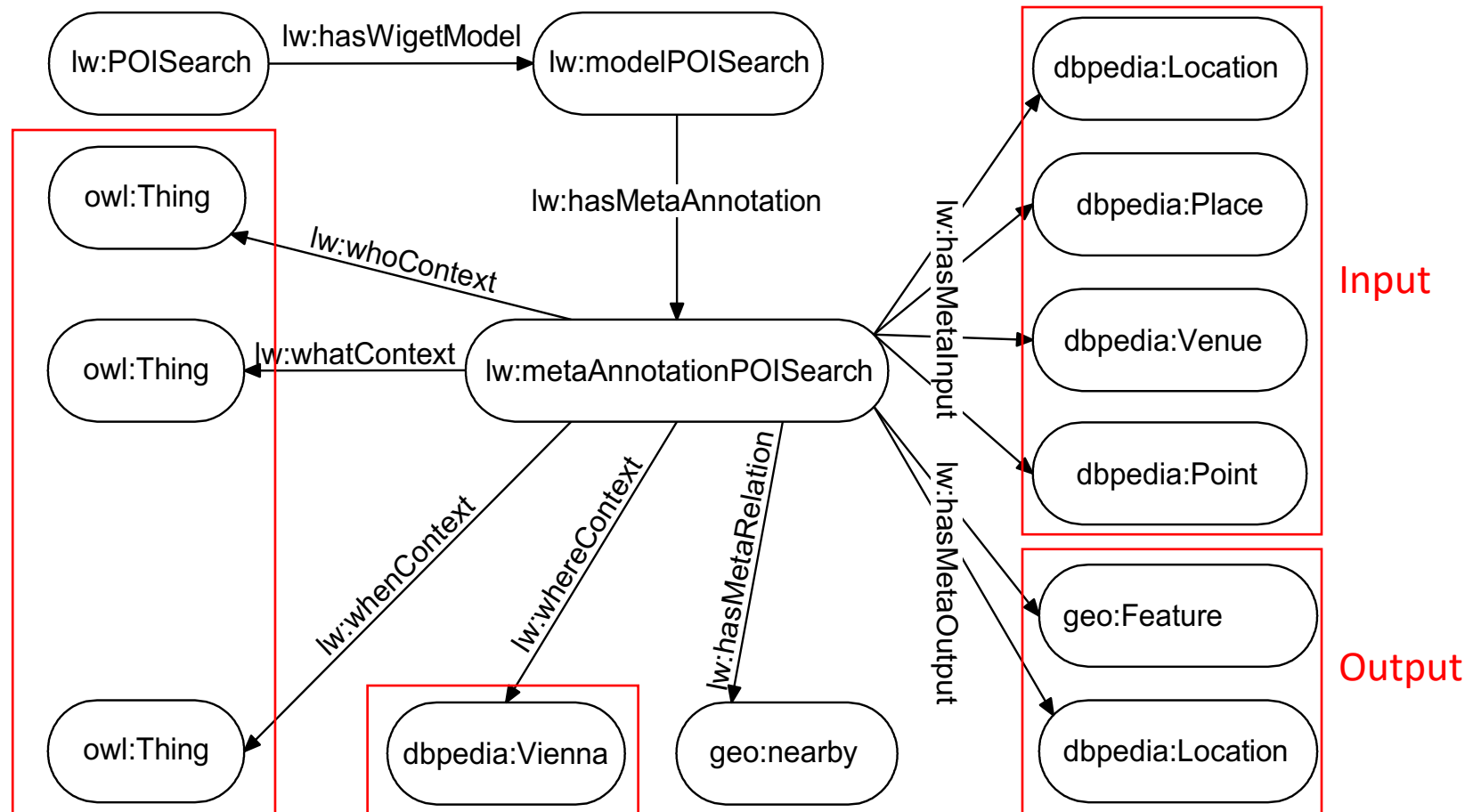
2. **Relation tagging**  
(LOD properties)

3. **Entity tagging**  
DBpedia entities

Example queries:

1. Point of interest
2. Vienna Castles
3. Playground with

# Meta-tagging model



**lw** <http://linkedwidgets.org/ontologies/>

**owl** <http://www.w3.org/2002/07/owl#>

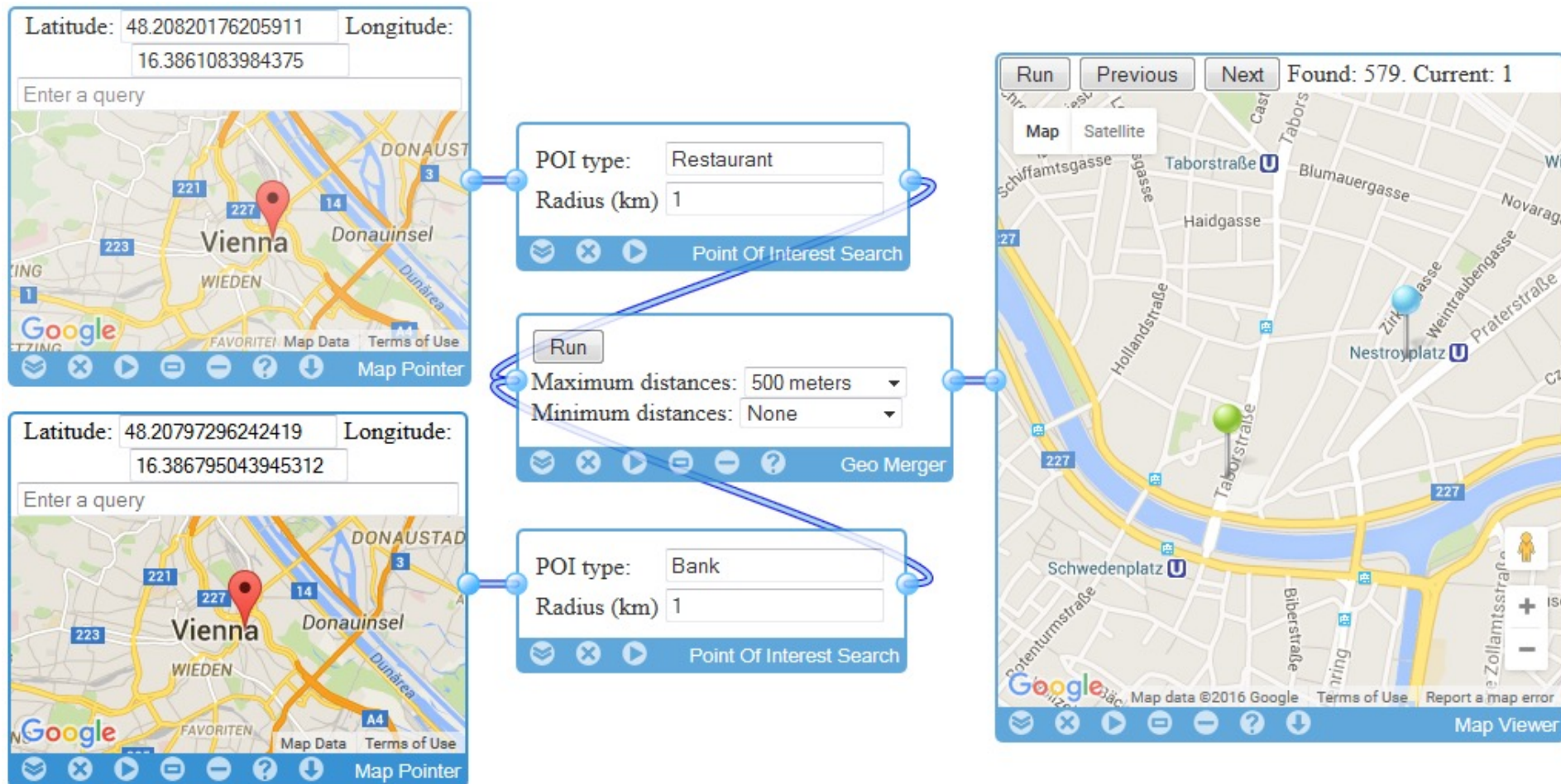
**dbpedia** <http://dbpedia.org/resource/>

**geo** <http://www.geonames.org/ontology#>



# Example

4 widgets: Map pointer, Geo merger, POI search, Map viewer

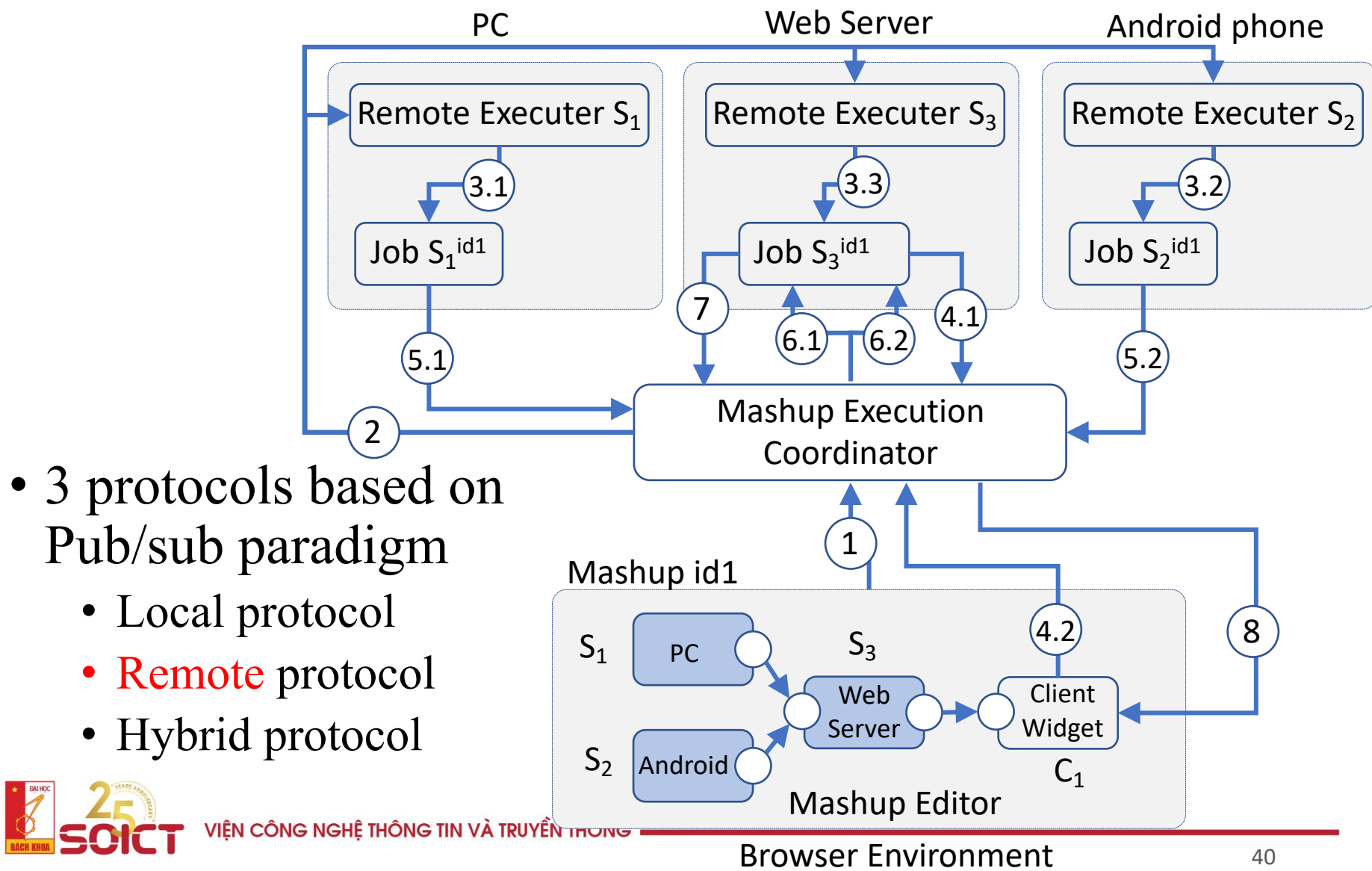


# 3. Mashup Models

# 3. Mashup models

- Model 1: collaborative mashups
- Model 2: persistent mashups
- Model 3: streaming mashups
- Model 4: distributed mashups

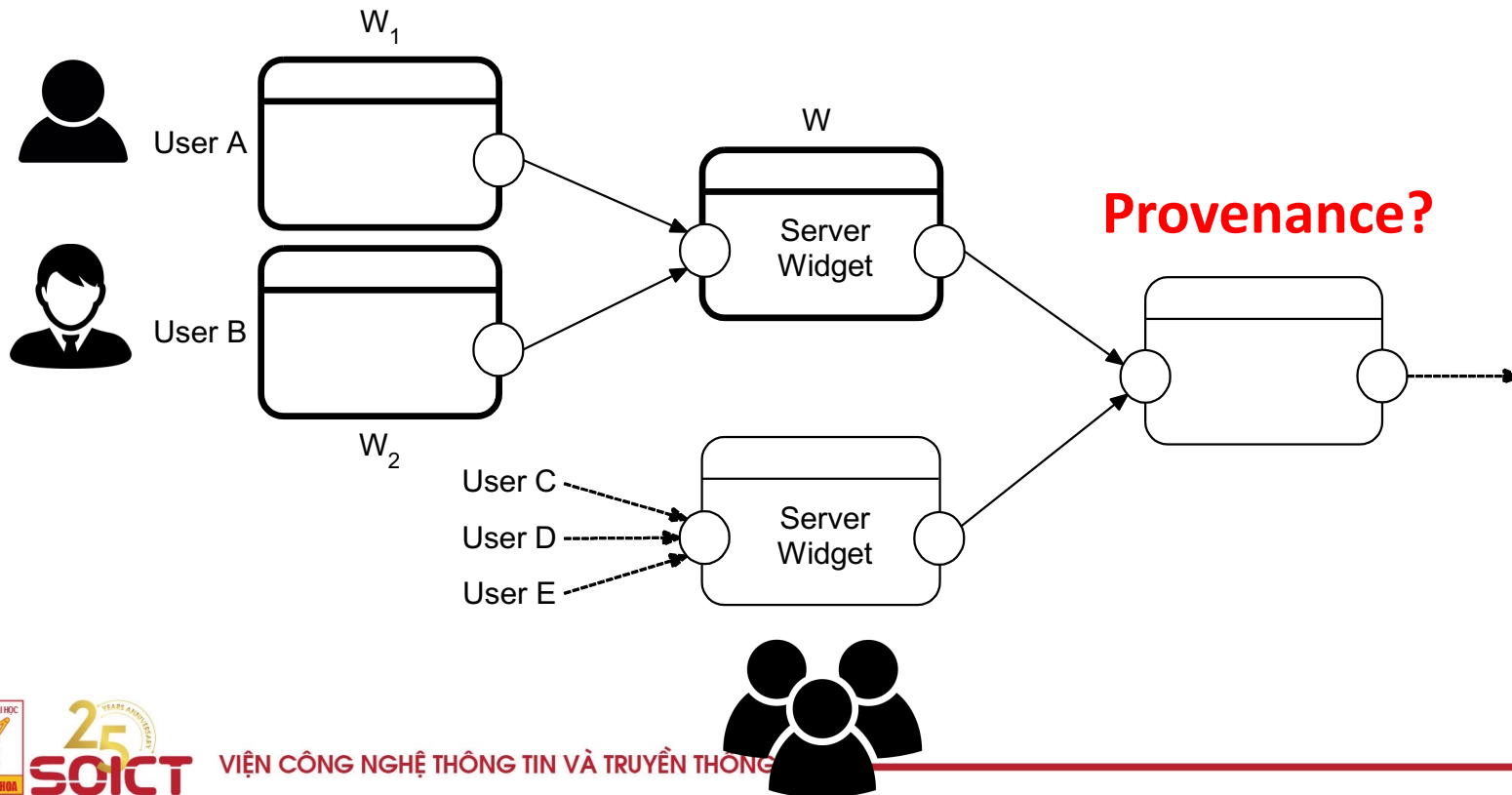
# Linked Widget communication protocols



- 3 protocols based on Pub/sub paradigm
  - Local protocol
  - **Remote** protocol
  - Hybrid protocol

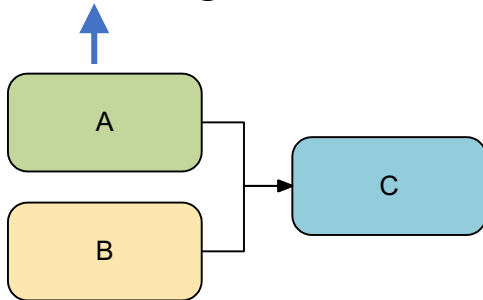
# Model 1 – Collaborative mashups

- Mashups are created and/or operated by multiple users at the same time



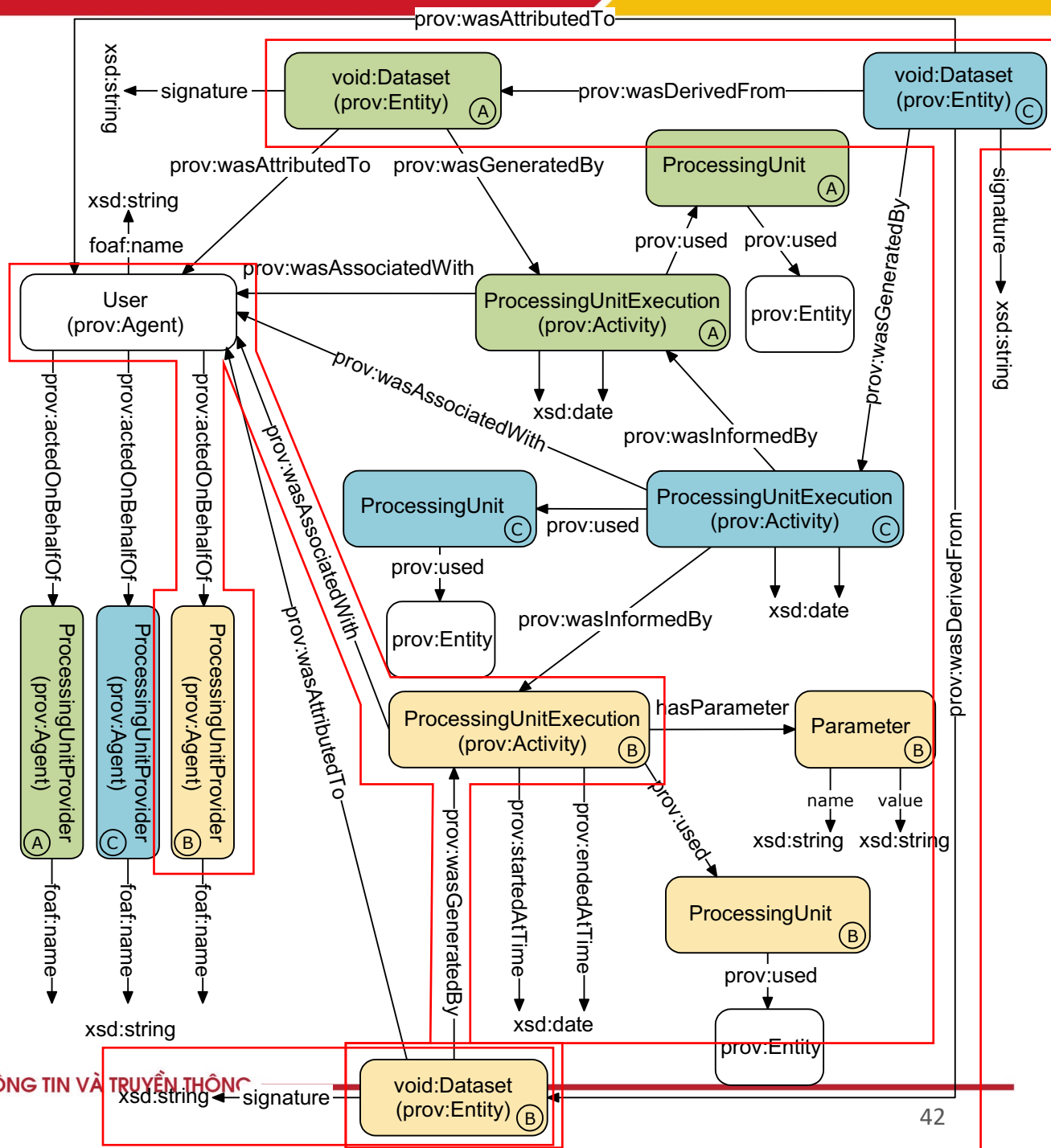
# Provenance Model

## Processing Unit



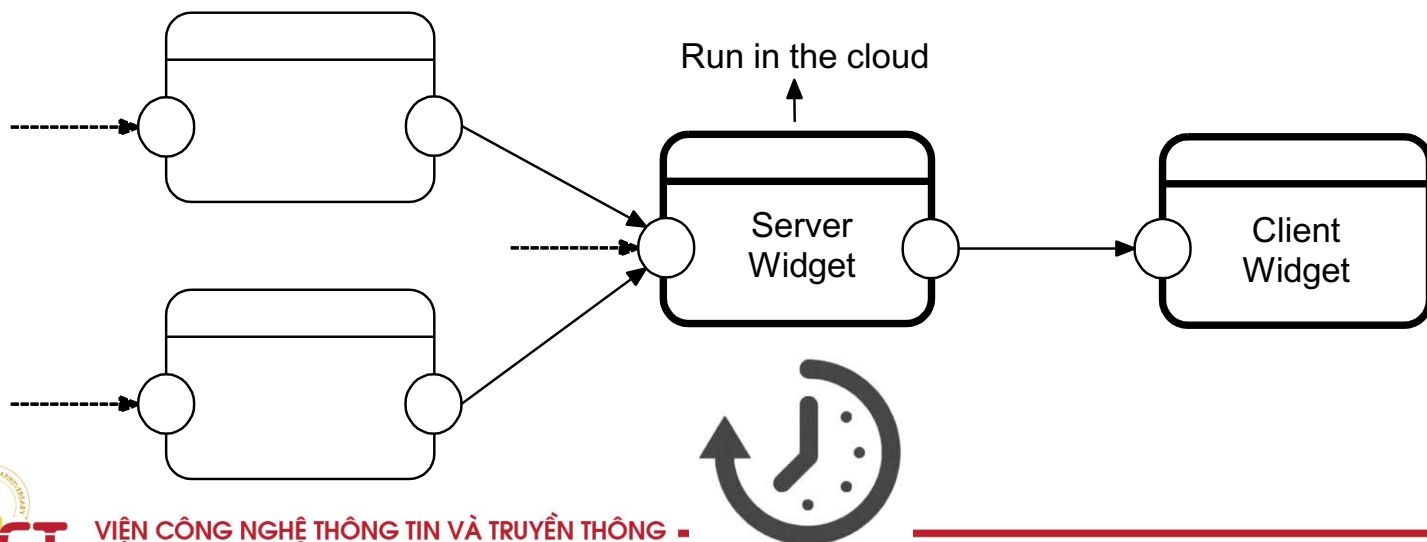
## Workflow

The model is based on  
**PROV-O ontology**



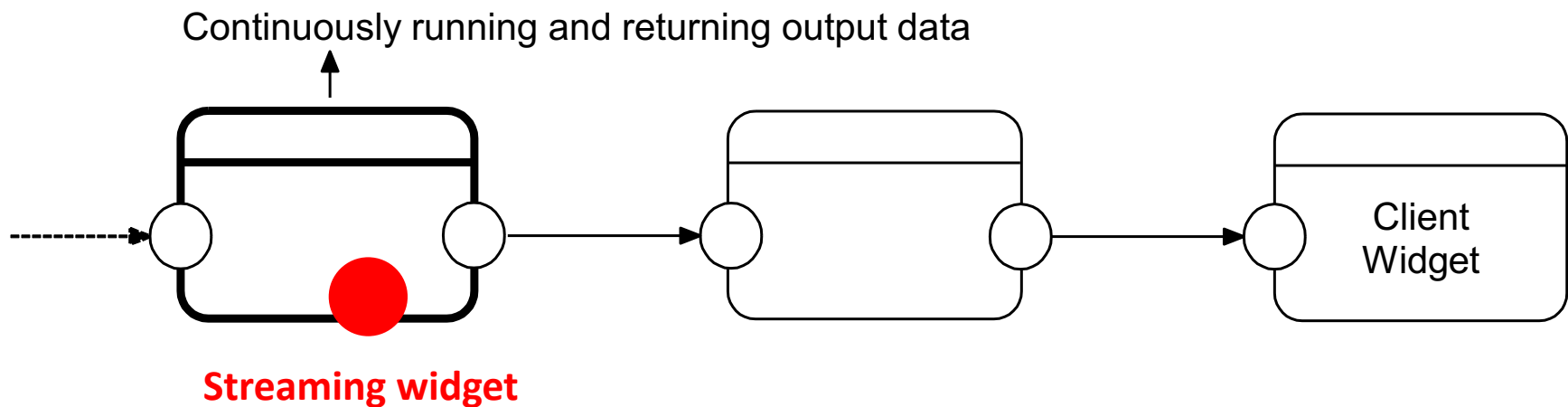
# Model 2 – Persistent mashups

- run continuously in the background
- are useful for **time-consuming** data integration tasks



# Model 3 – Streaming mashups

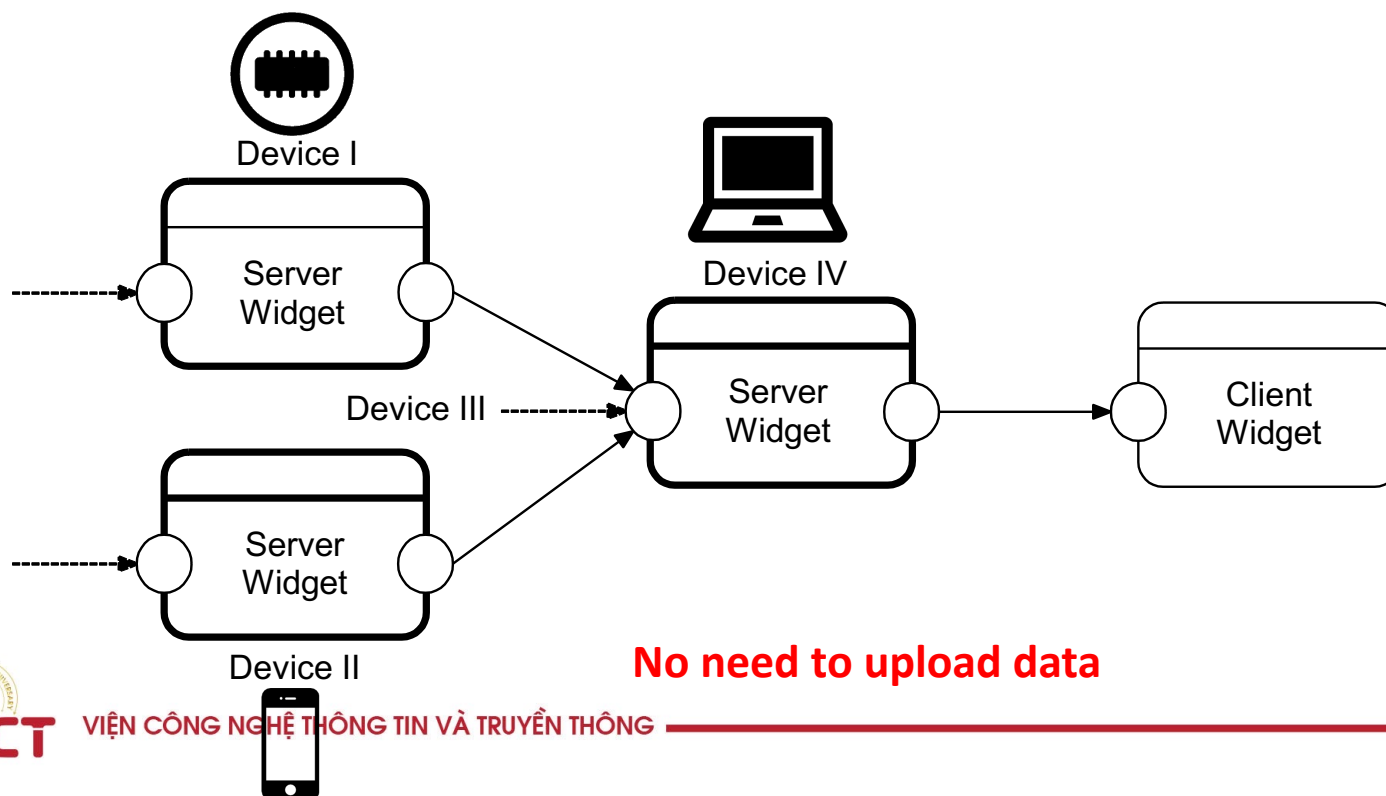
- Data is continuously flowing from a widget to others
- Streaming mashups are useful for data monitoring use cases.





# Model 4 – Distributed mashups

- Widgets are hosted by distributed nodes and devices
- Useful for integrating sensor data and data from embedded devices



# Distributed & collaborative mashup example: Combine and visualize sales data for a series of retail points of sale (POS)

POS Spreadsheet (pos, location, country, city)

Link to the Google sheet

```
https://docs.google.com/spreadsheets/d/1vcB8YooN1G7zaYebvI25k0hpULk0ossEUDkqJ3waIAk/edit#gid=0
```

Google Sheet

Merge multiple W3C cube datasets

Cube Merger

Filter

1. POS --
2. Date --
3. Category Fruit
4. Location --
5. Country --
6. City --
7. Shop --

Filter

Function: ☒ Sum ☐ Average

1. ☐ POS
2. ☐ Date
  - ☐ Monthly
  - ☐ Quarterly
  - ☐ Yearly
3. ☐ Category
4. ☐ Location
5. ☐ Country
6. ☒ City

Aggregation

Please remember to run your spreadsheet server widget first!

POS 1

Please remember to run your spreadsheet server widget first!

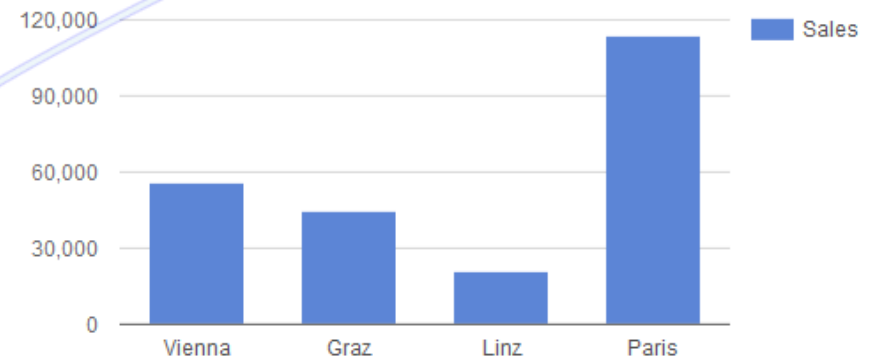
POS 2

Please remember to run your spreadsheet server widget first!

POS 3

Run

Type of chart Column chart



Sales Spreadsheets (pos, date, category, sale)

# 4. Conclusions and Future Work

# Conclusions

- Development of concepts combining semantic web and mashups
- Semantic model for widgets
- Facilitating collaborative work among **end-user**, **data publisher** and **developer** communities
- **Introducing a new model** of *semantic*, *distributed*, and *collaborative* mashups
- Running prototype: <http://linkedwidgets.org>