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.)

980 parks in Vienna

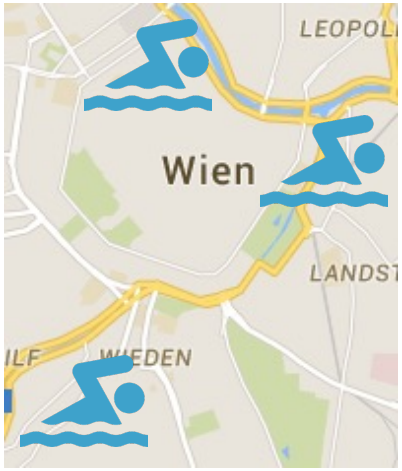
XML(lat, long, Image URL)

Flickr/Google Image Search

JSON(lat, long, CO, NO_2, etc.)

Air quality data

CSV(lat, long, name, etc.)



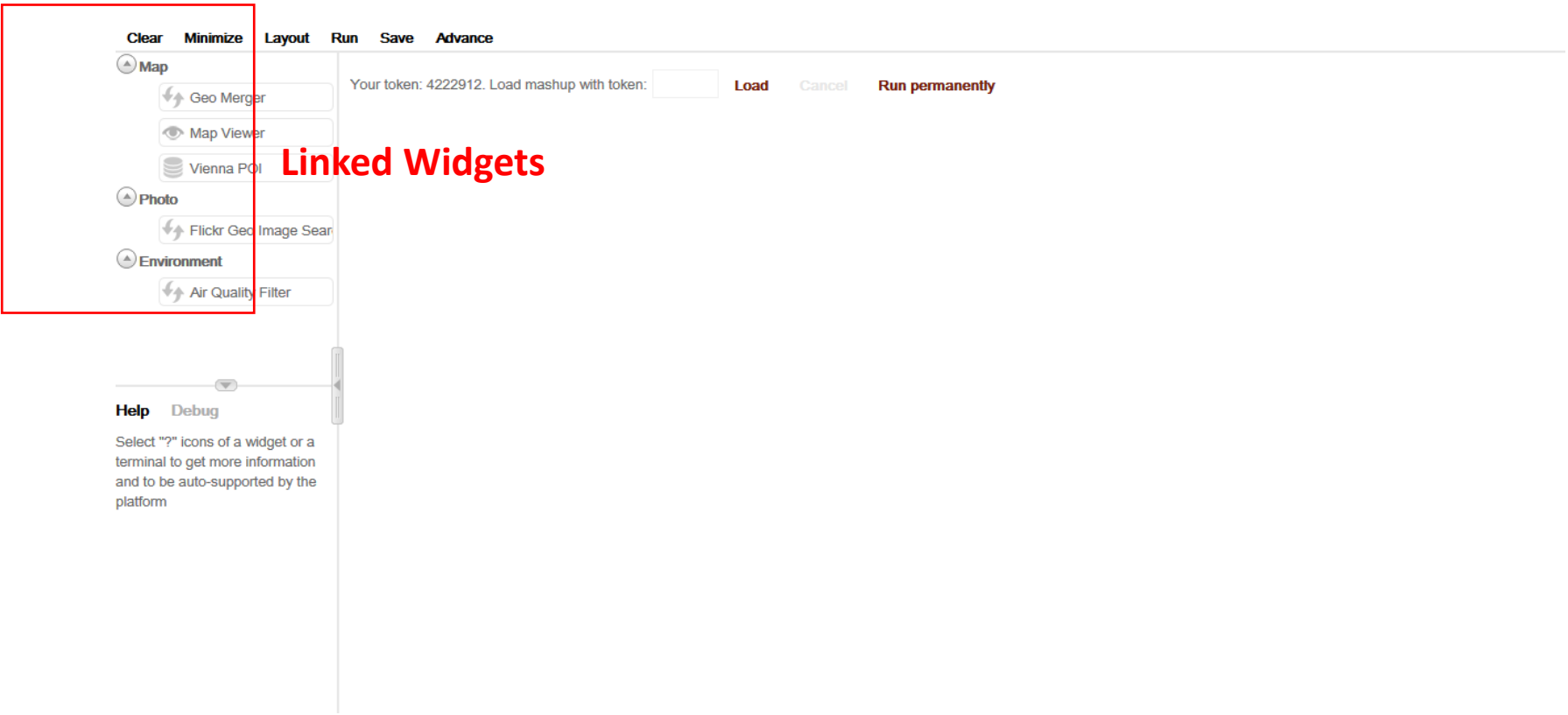
51 public swimming pools in Vienna

?

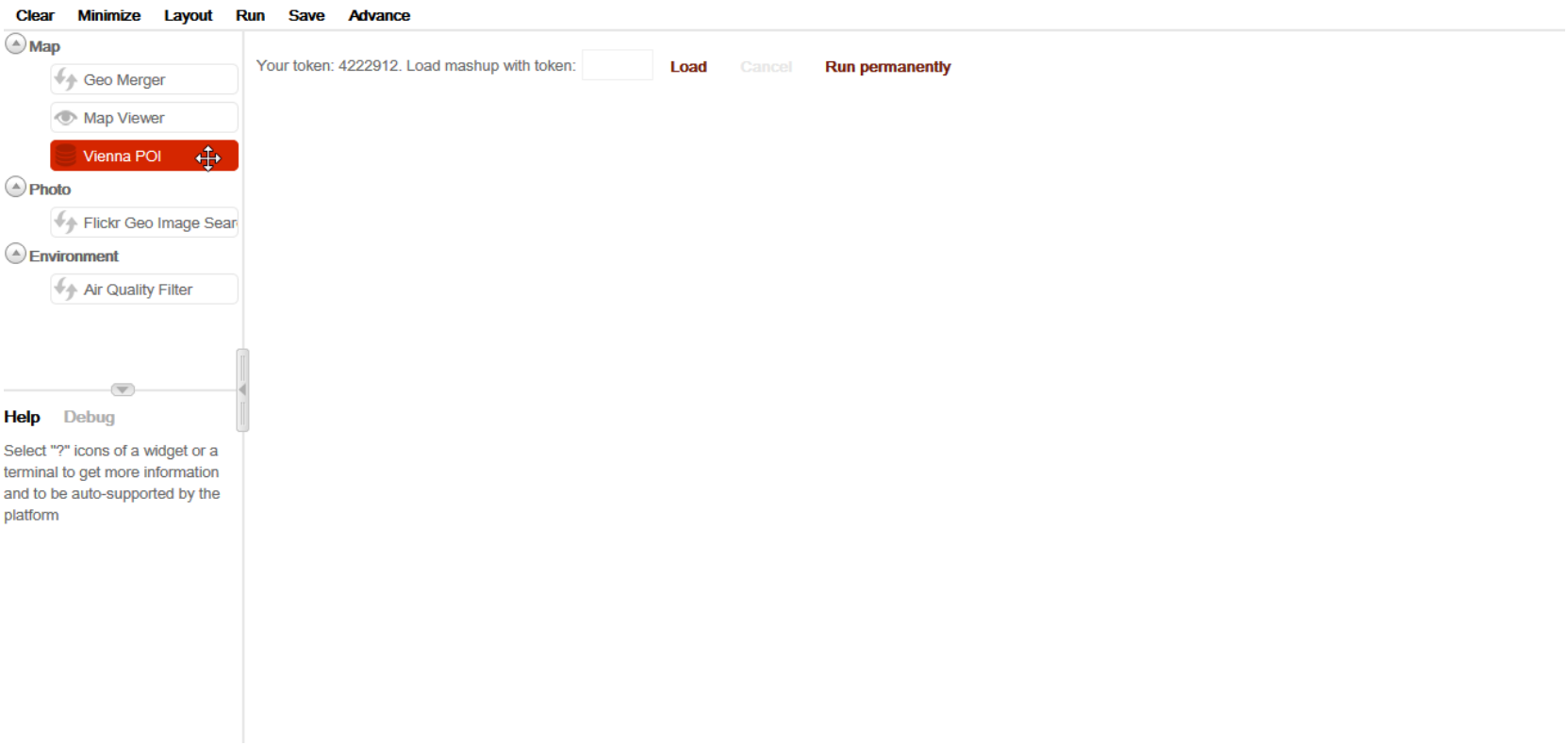
Motivation


Pötzleinsdorfer Schlosspark
Krottenbachstraße
BRIGITTENAU
Donaustraße
WÄHRING
Gymnasiumstraße
Dresdner Str.
Alszeile
Alte Donau
Industriestraße
Siebenbürgerstraße
Hardeggasse
OTTAKRING
CO
Schloss Belvedere
Schloß Schönbrunn
Tiergarten Schönbrunn
ERDBERG
FAVORITEN
Kopalgasse
Hafen Freu
Söng Danube

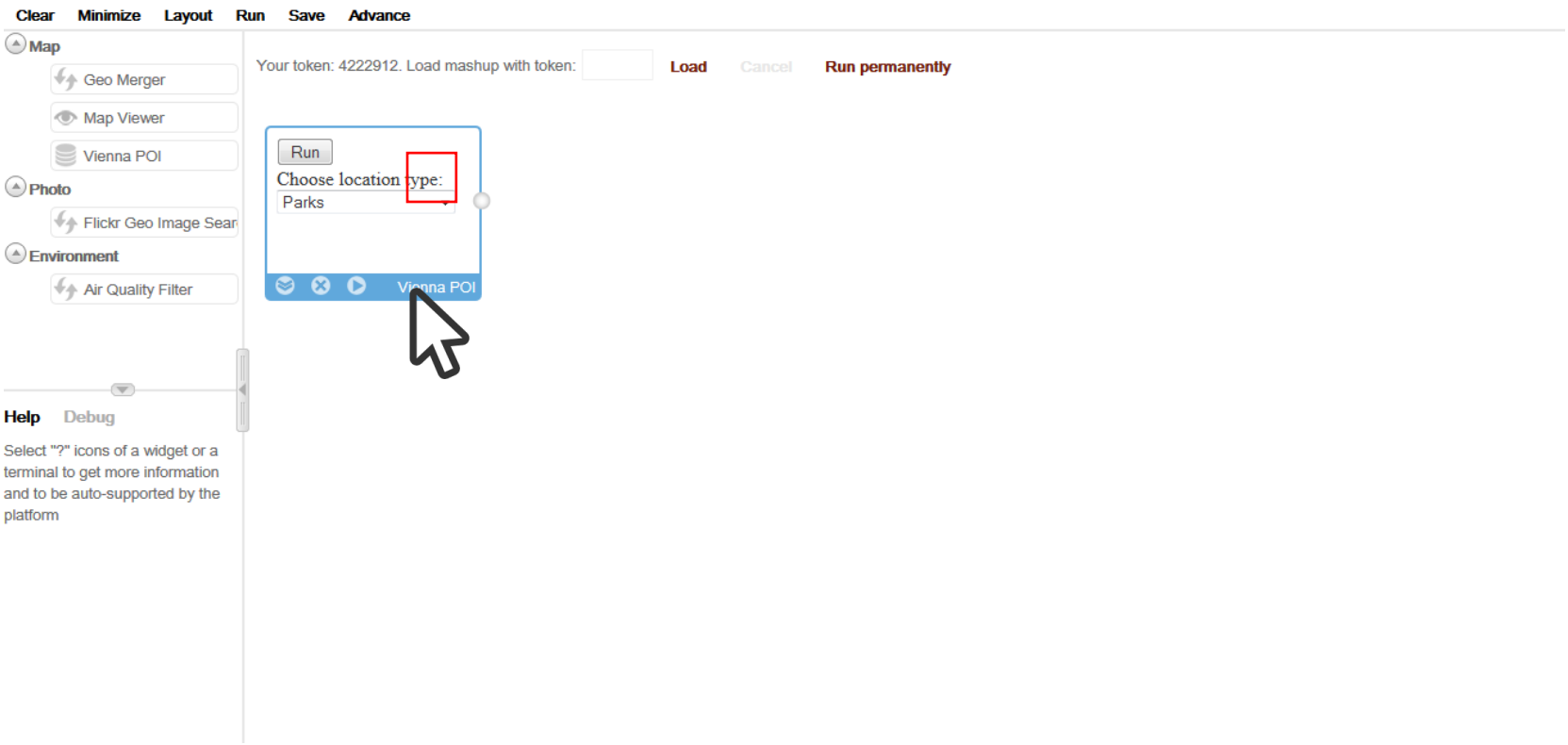
Modular approach for data integration



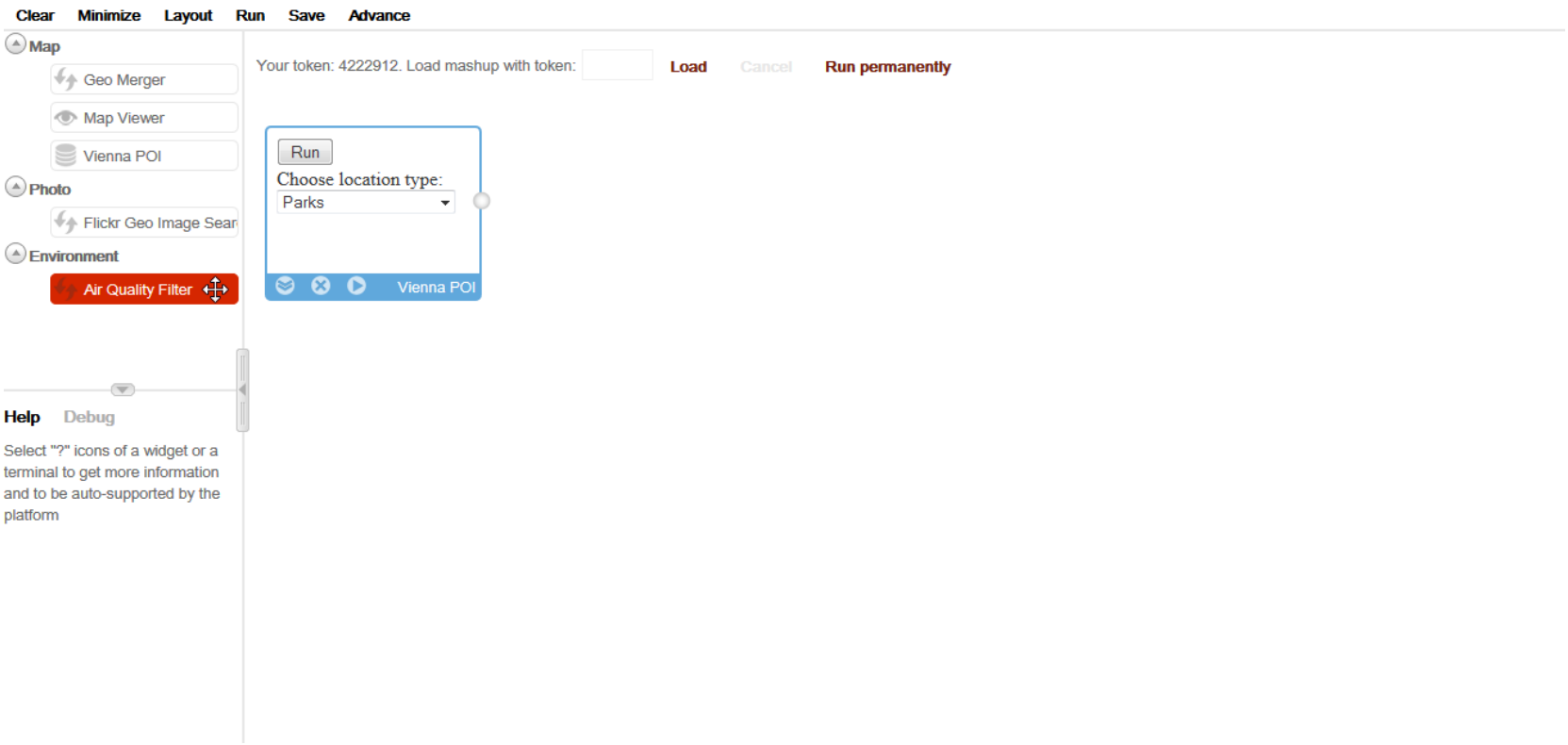
Modular approach for data integration



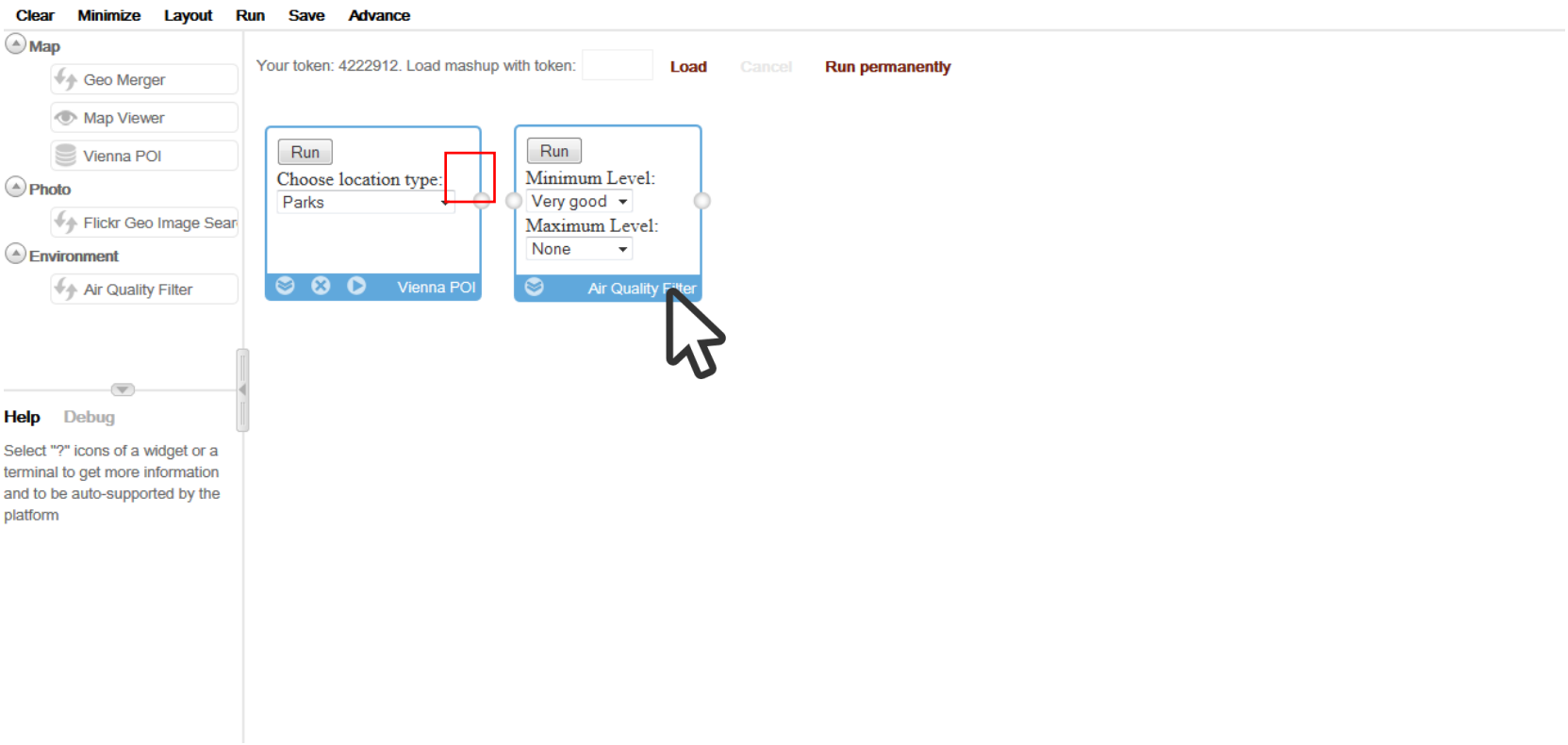
Modular approach for data integration



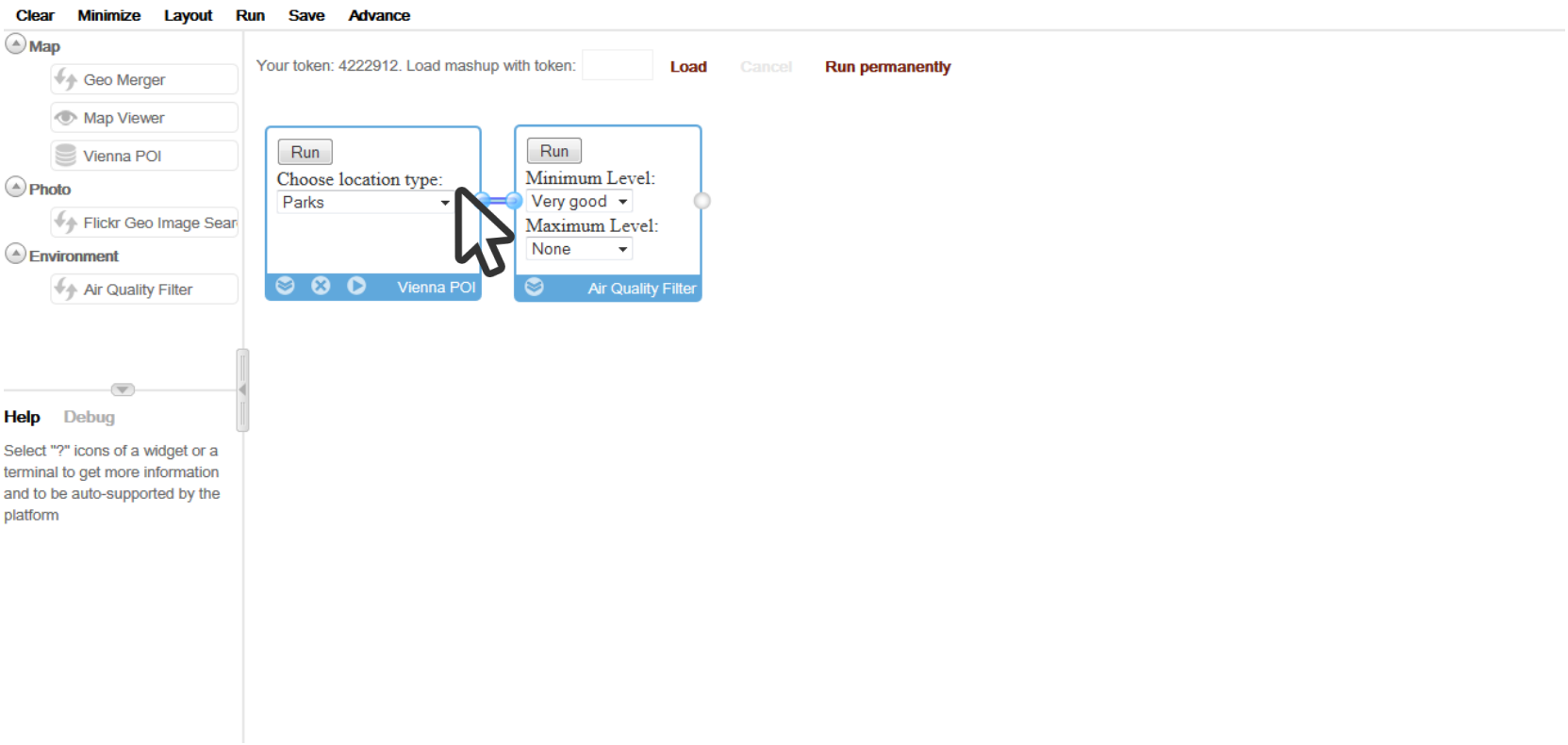
Modular approach for data integration



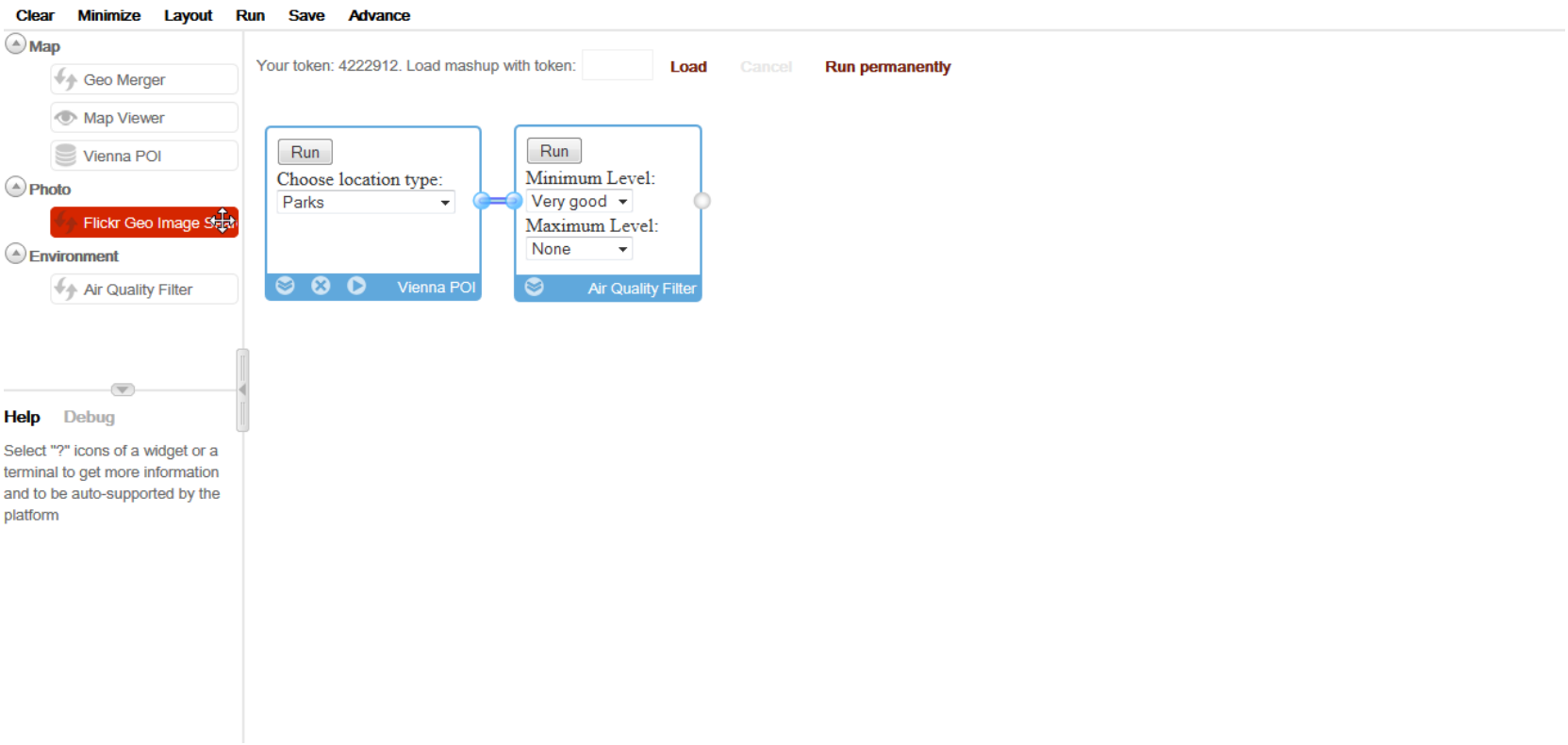
Modular approach for data integration



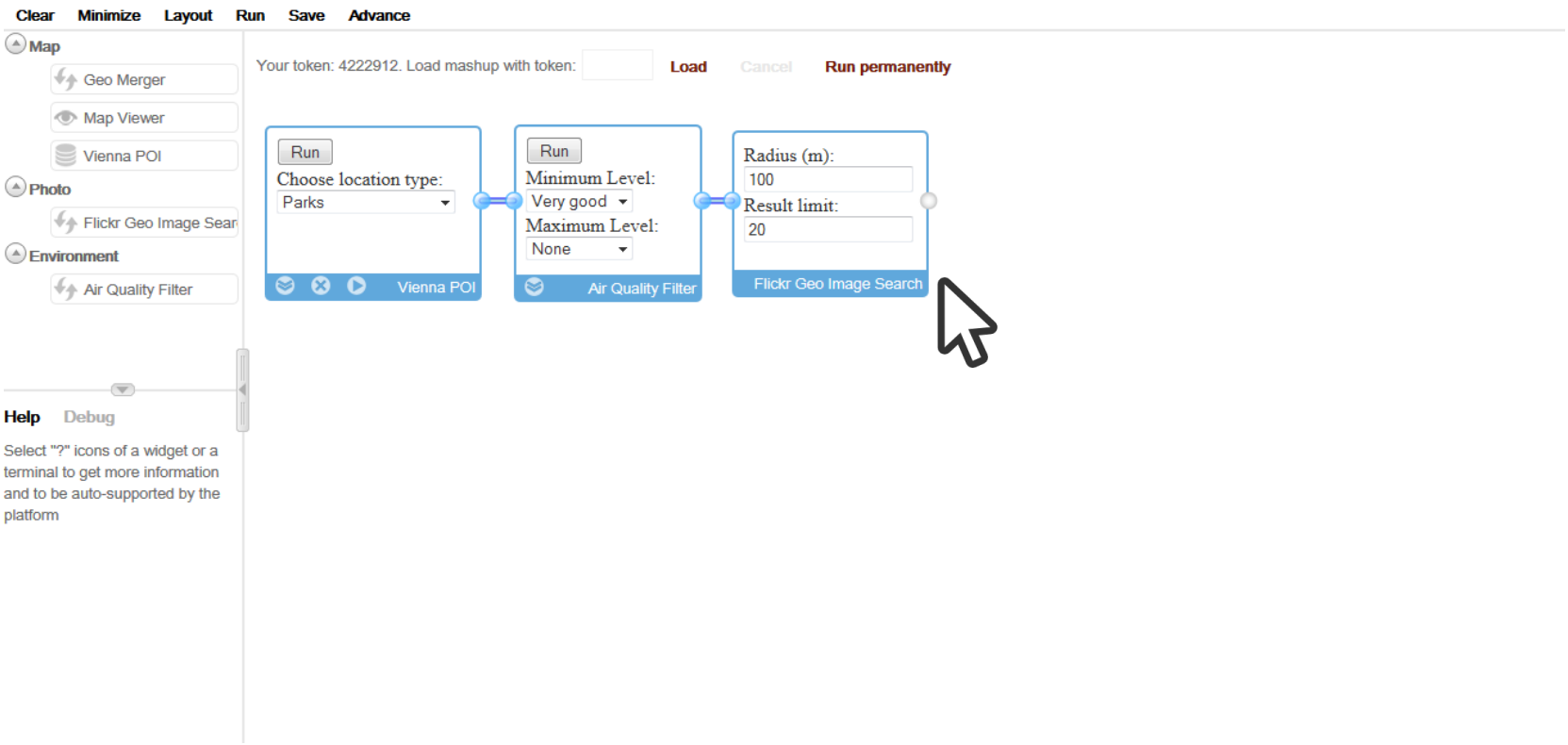
Modular approach for data integration



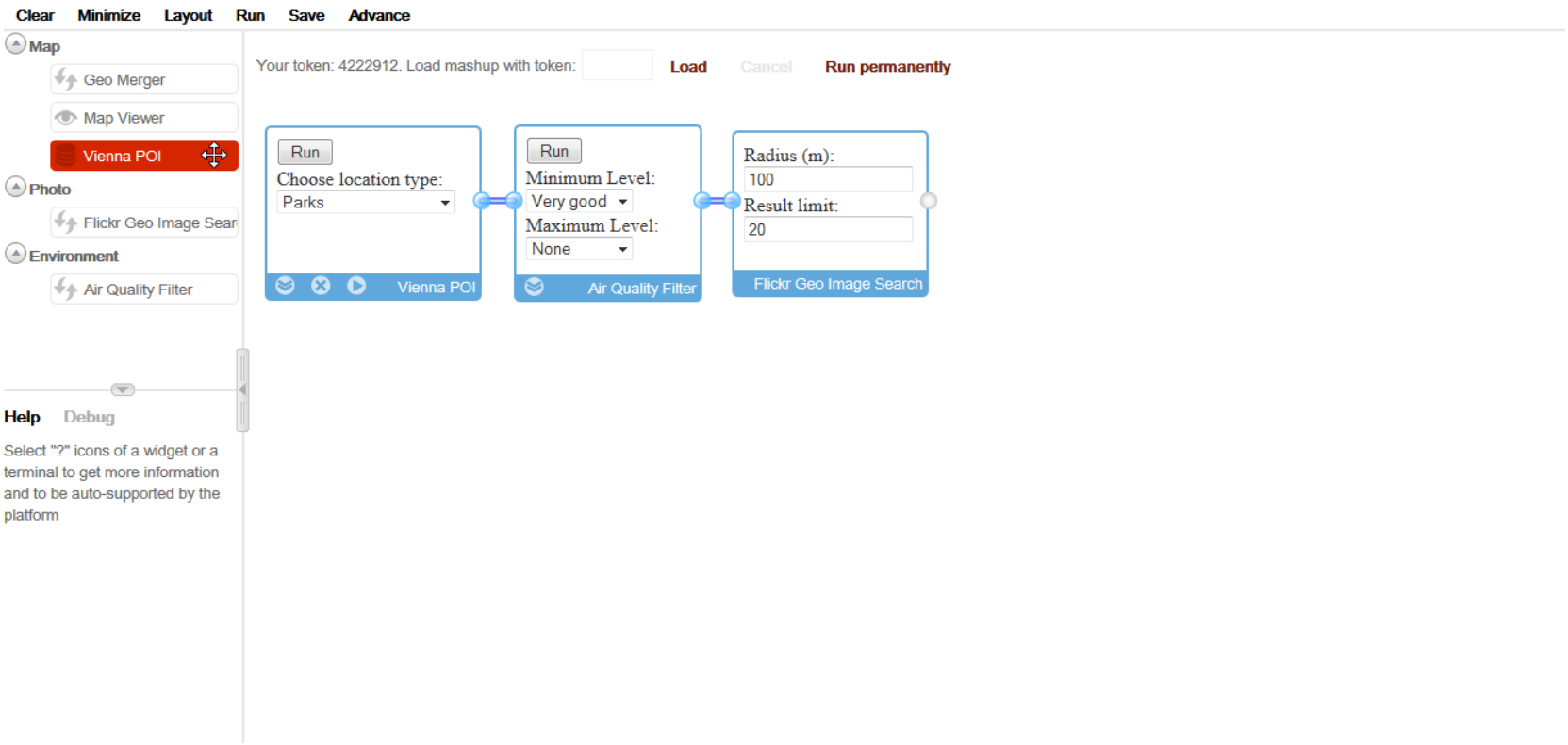
Modular approach for data integration



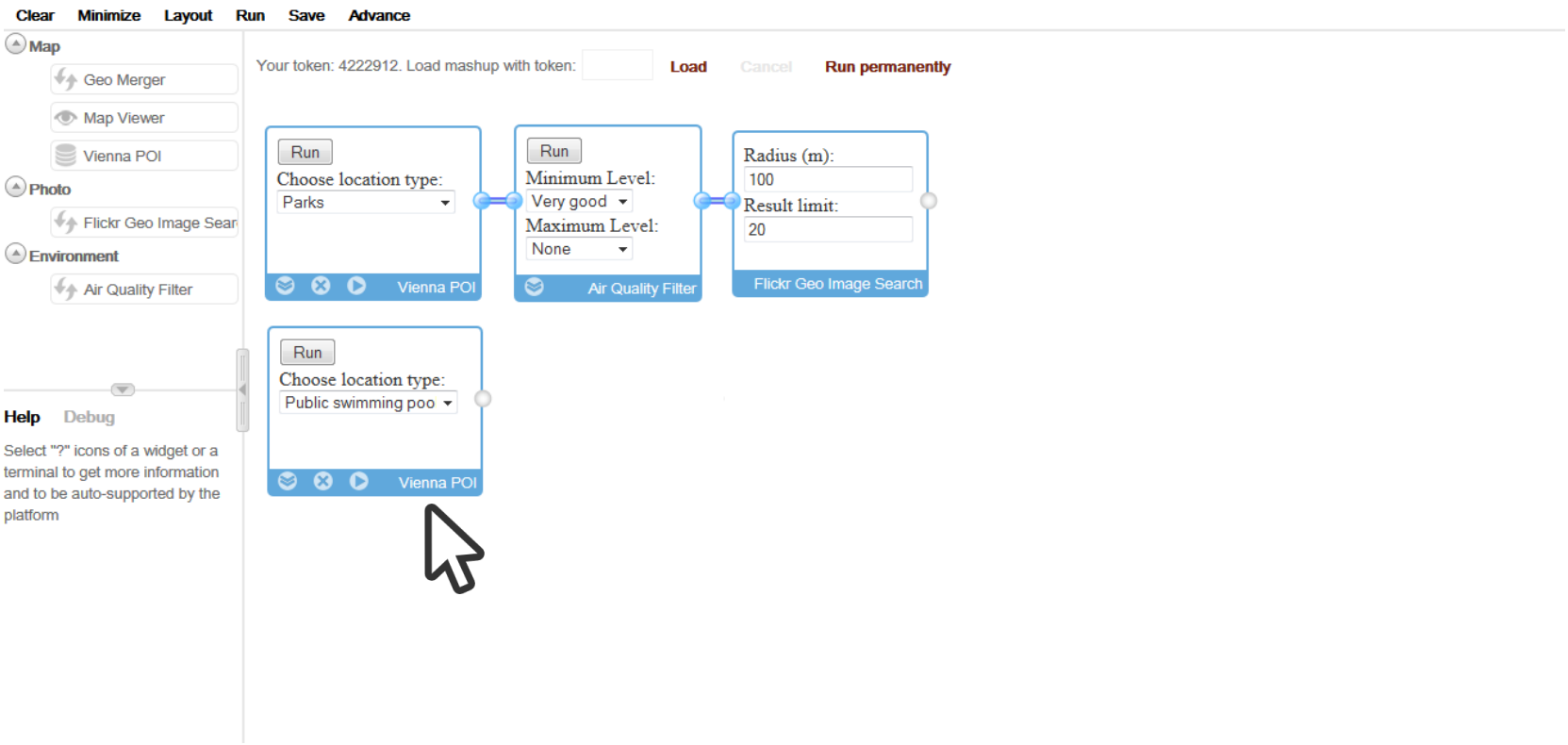
Modular approach for data integration



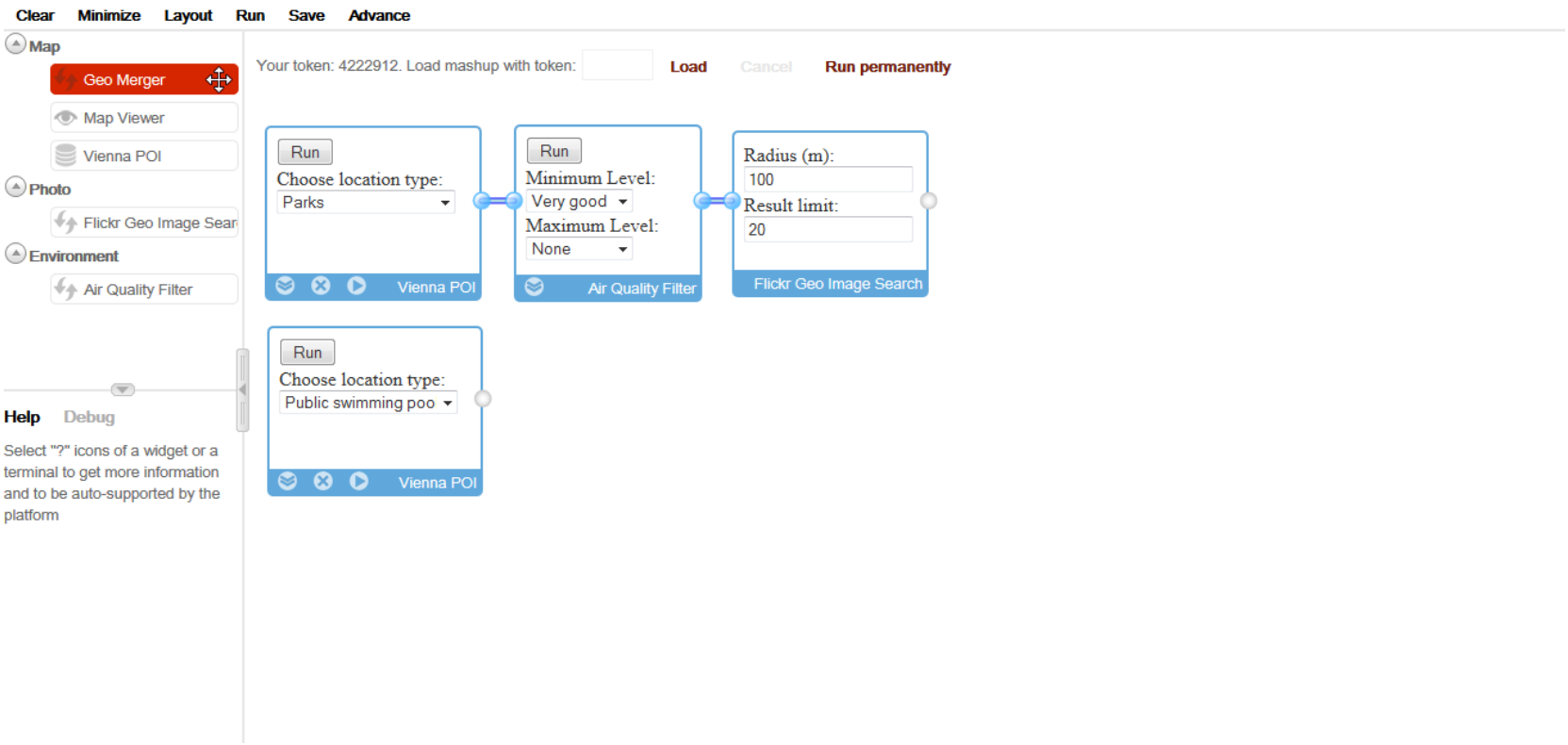
Modular approach for data integration



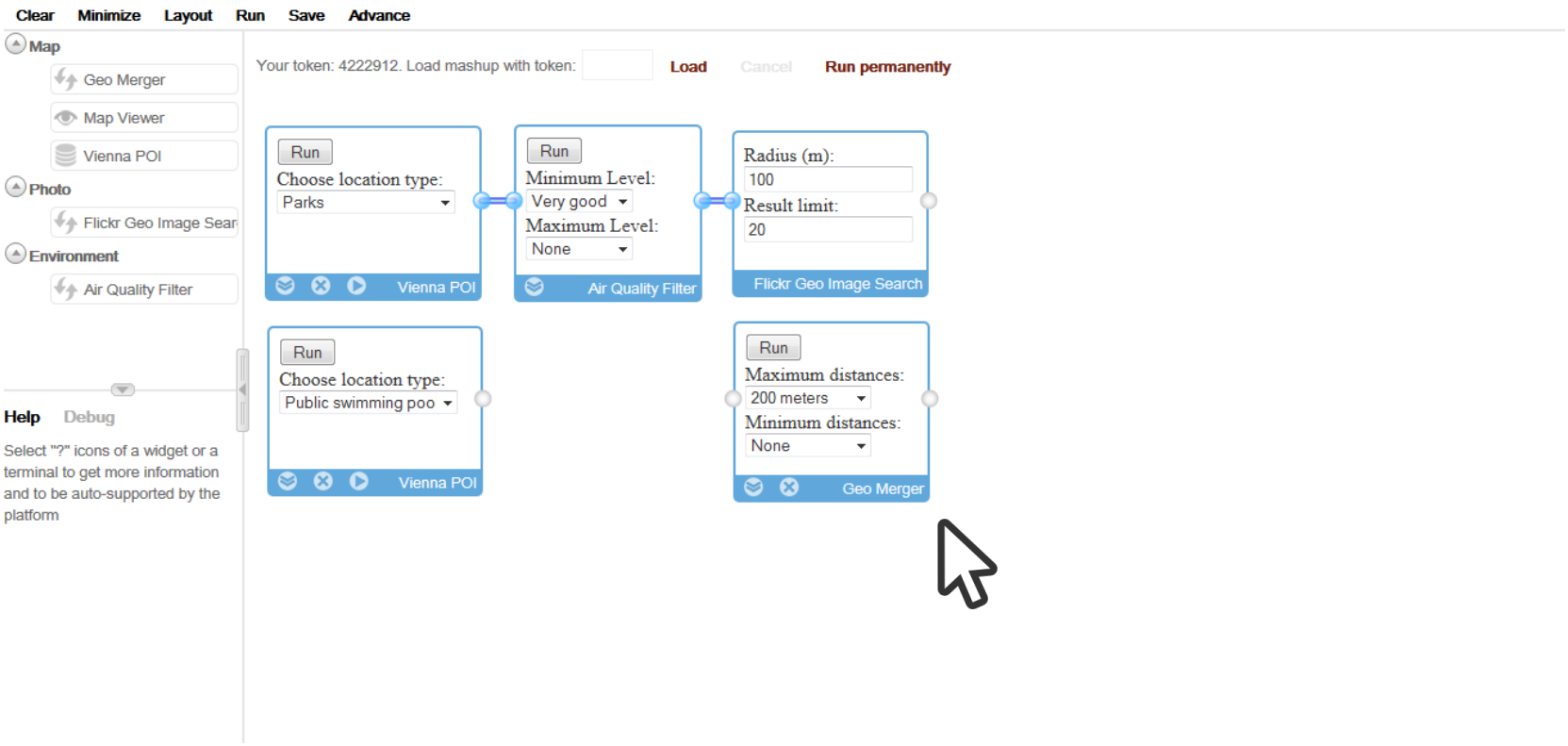
Modular approach for data integration



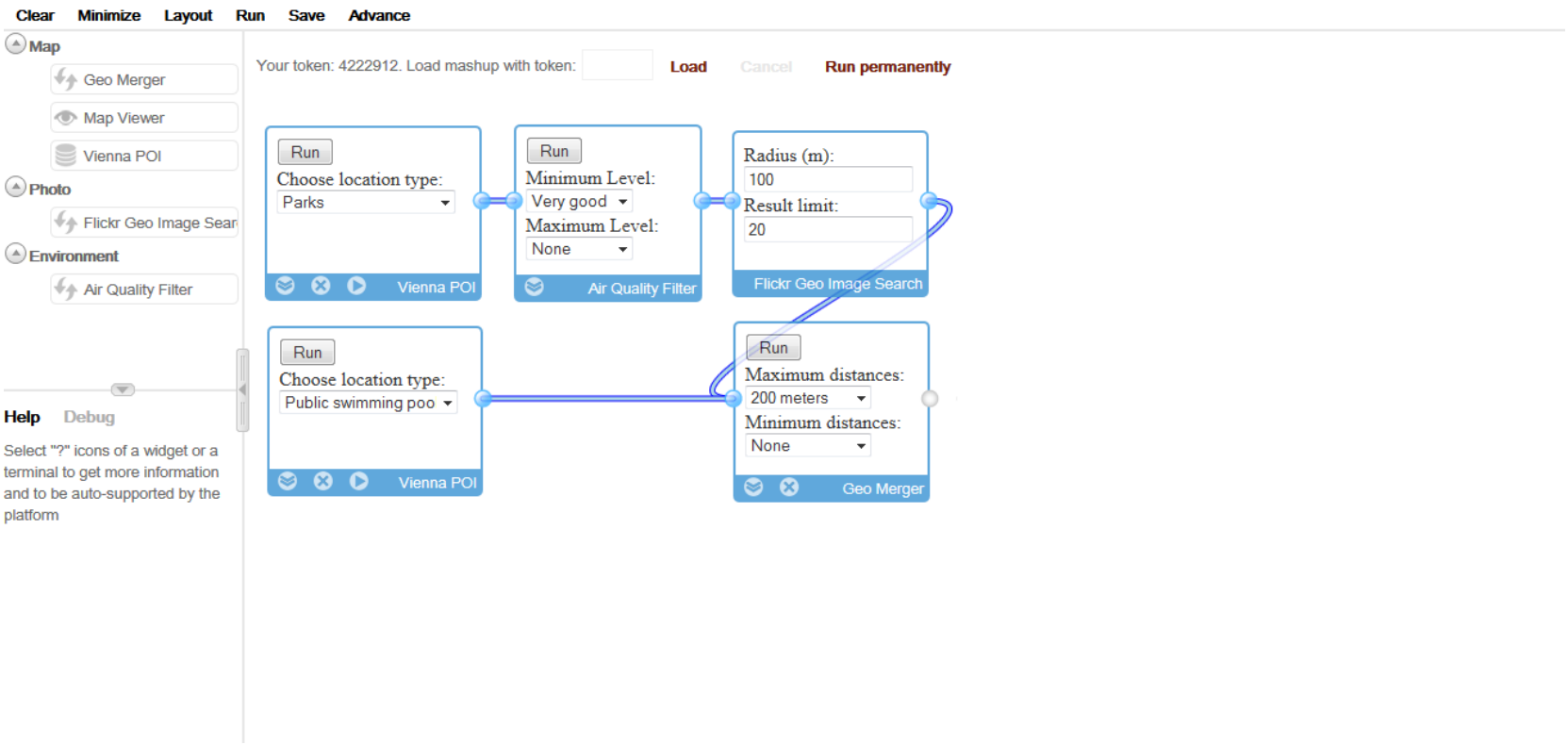
Modular approach for data integration



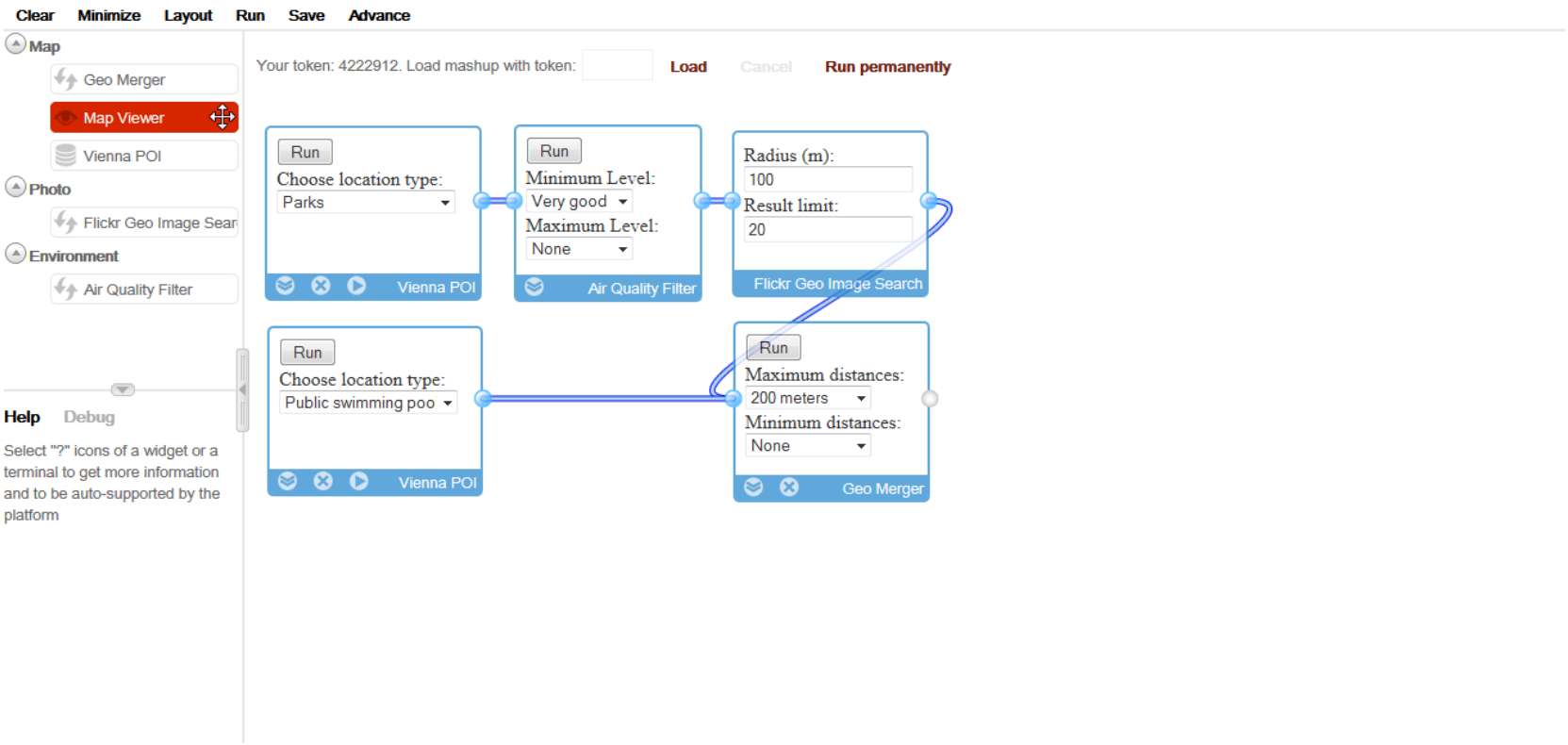
Modular approach for data integration



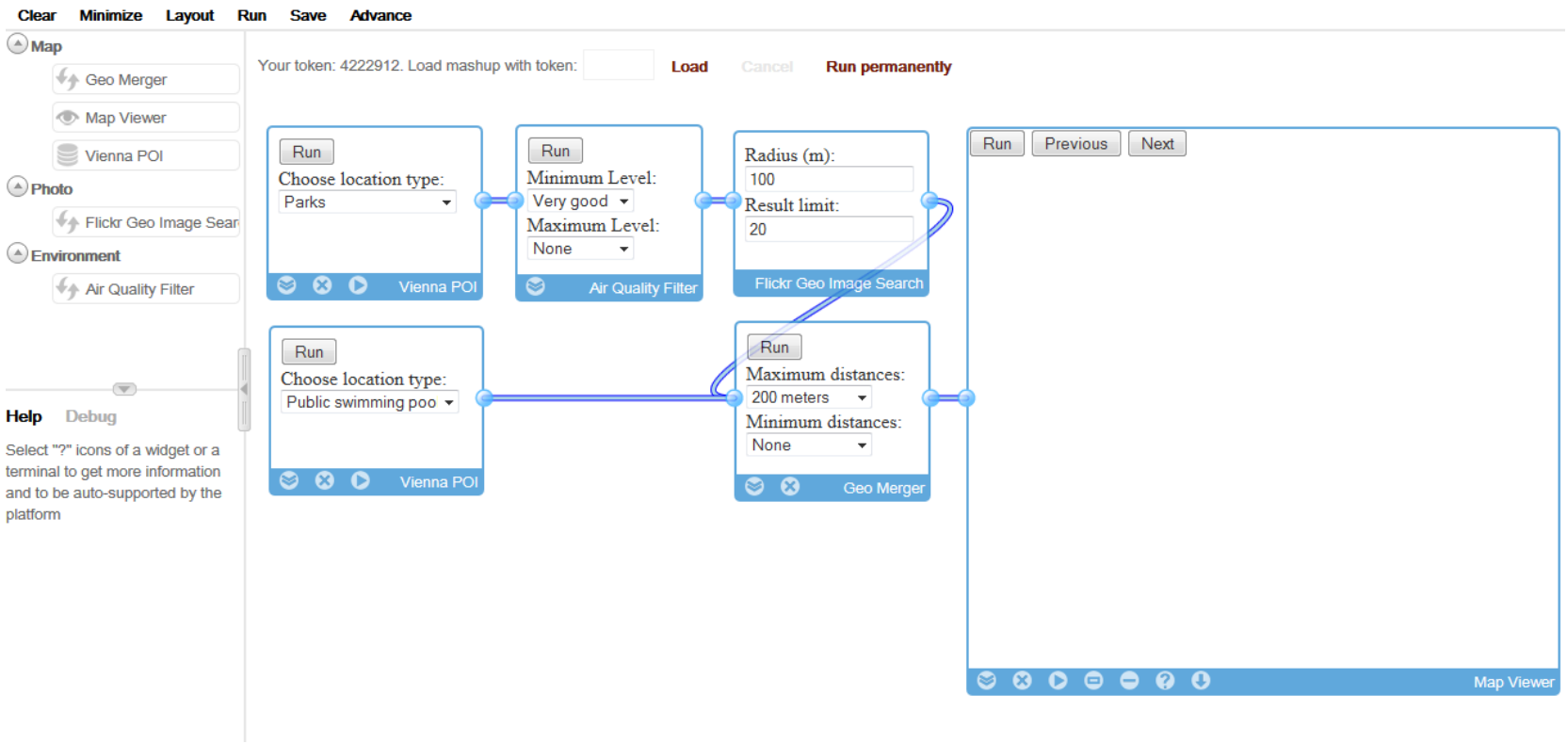
Modular approach for data integration



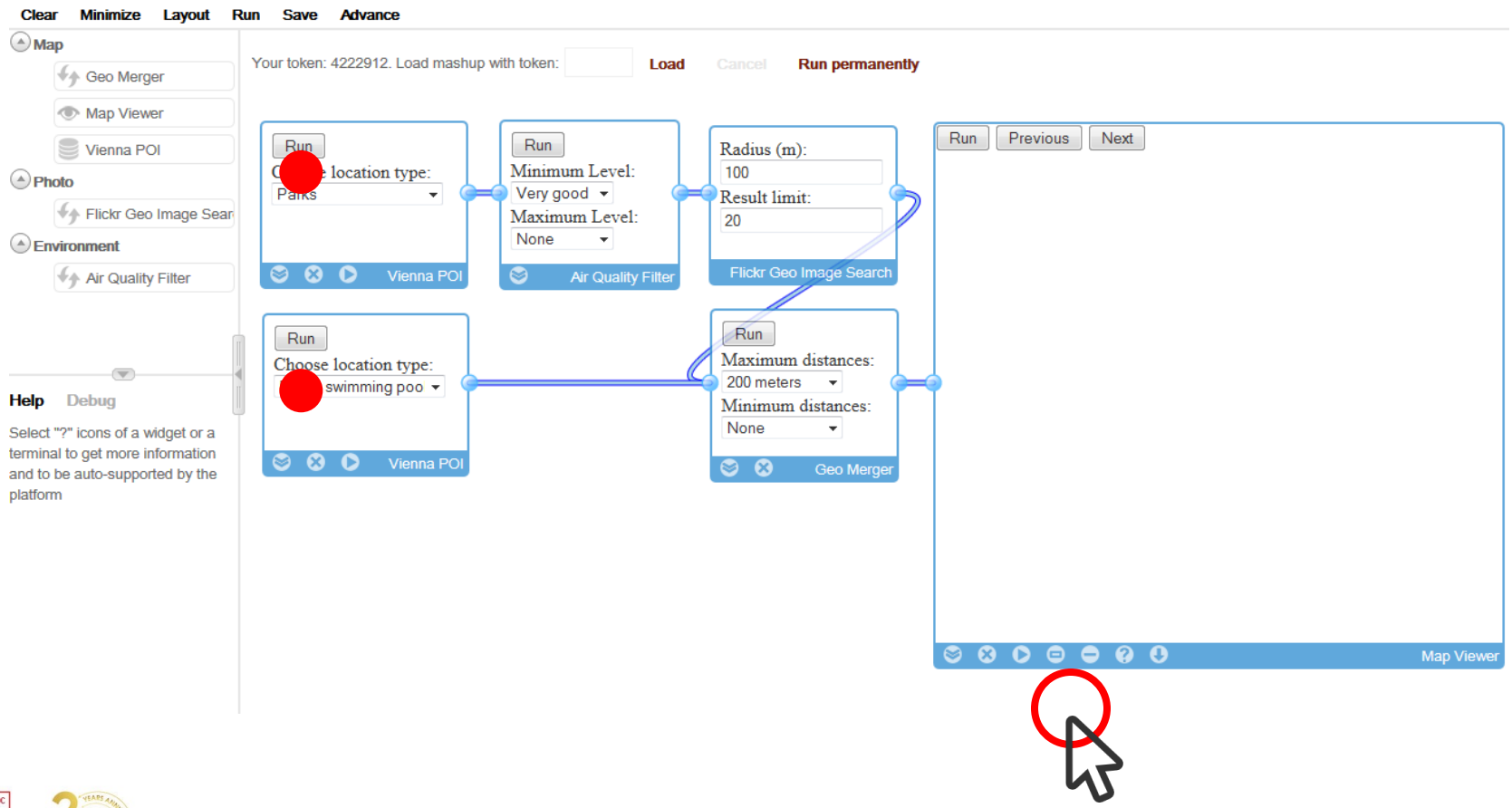
Modular approach for data integration



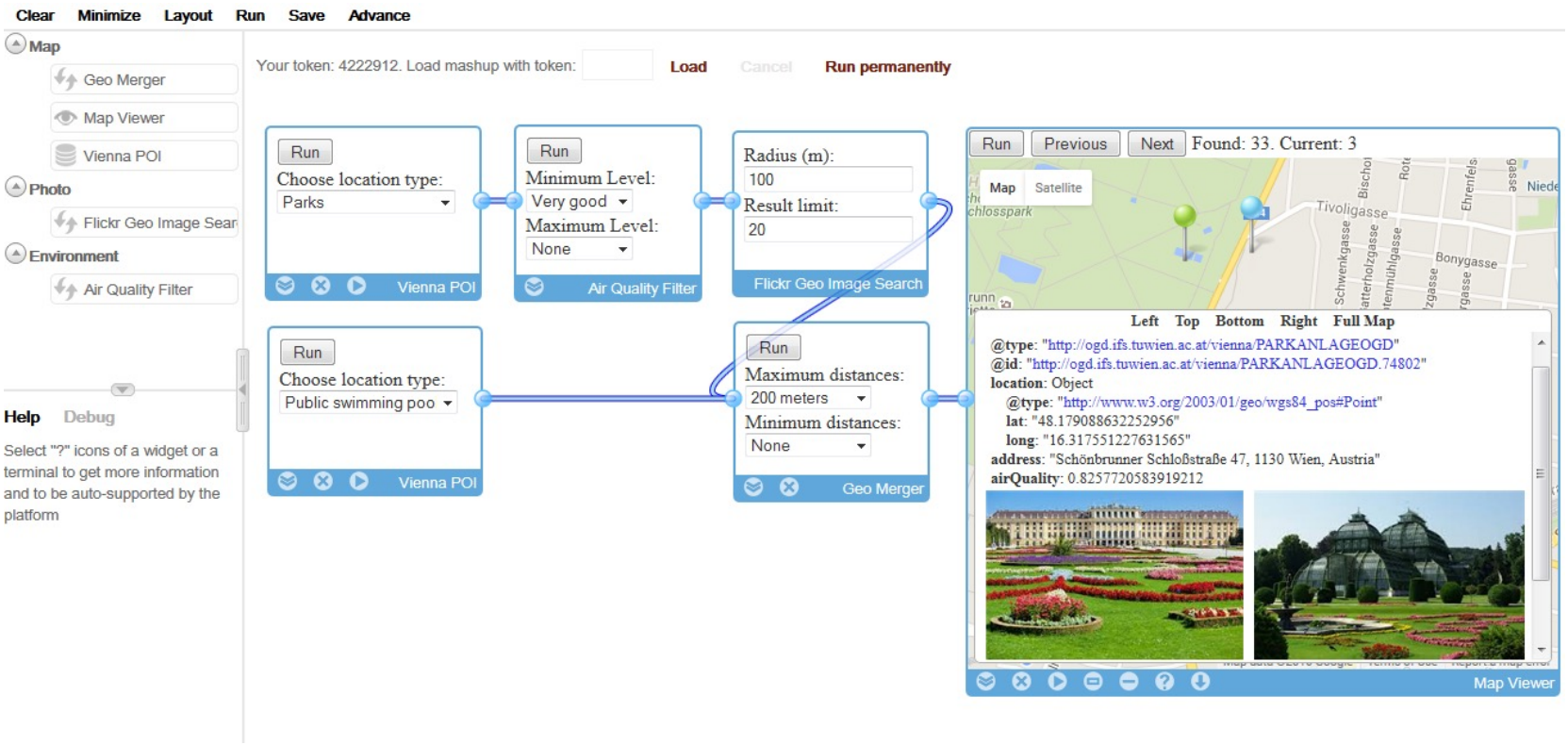
Modular approach for data integration



Modular approach for data integration



Modular approach for data integration



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

[1] Yu, J., Benatallah, B., Casati, F., Daniel, F.: *Understanding mashup development*. IEEE Internet Computing (2008)

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

[2] Endres-Niggemeyer, Brigitte. *Semantic mashups: Intelligent Reuse of Web Resources* (2013)

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?*

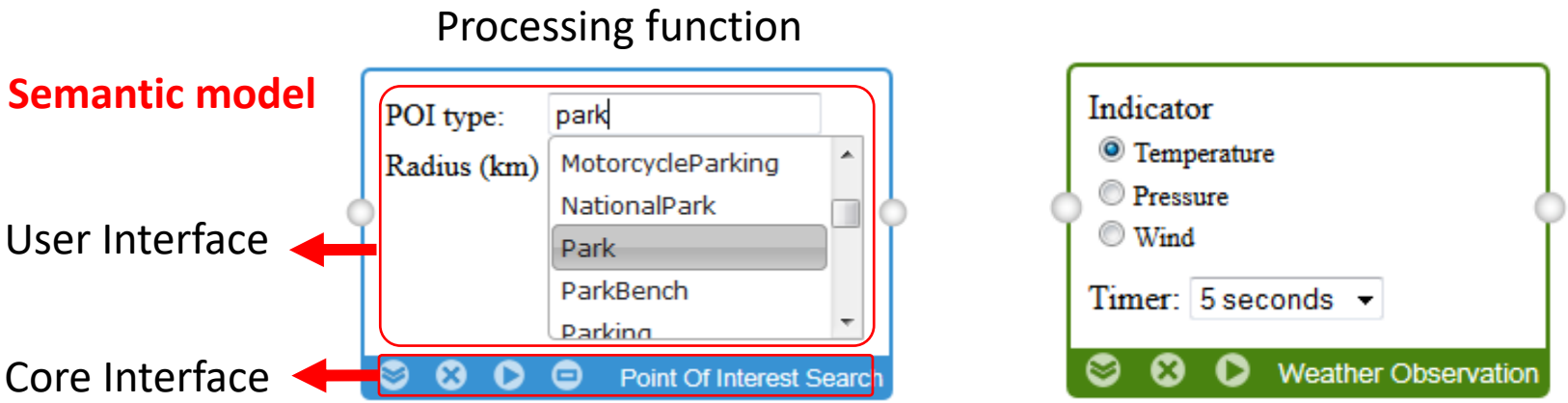
RQ2: *How can non-expert users be enabled to collaboratively integrate data?*

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

2. Linked Widgets Framework

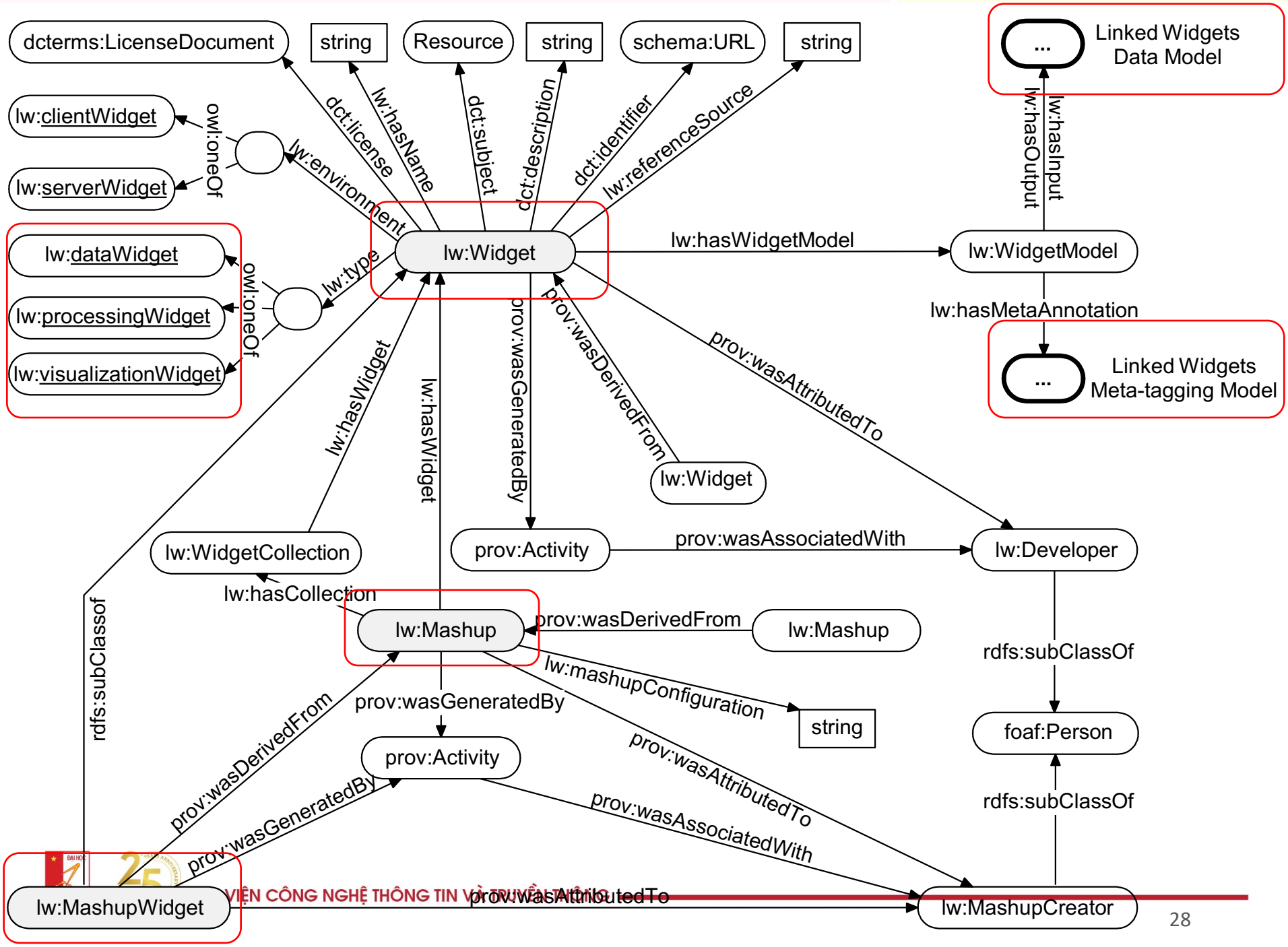
Linked Widgets

- 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

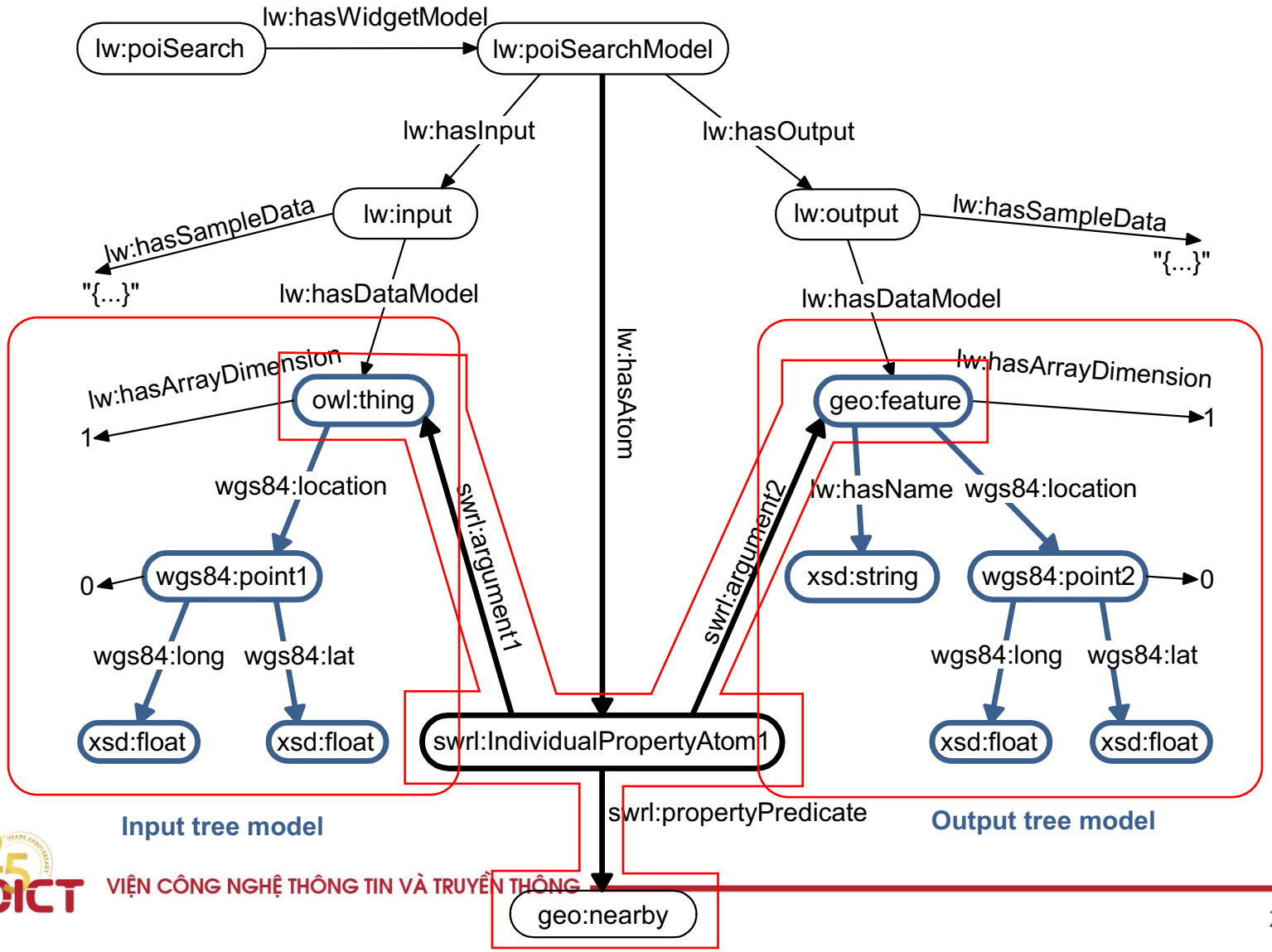


Client Linked Widgets run in a browser

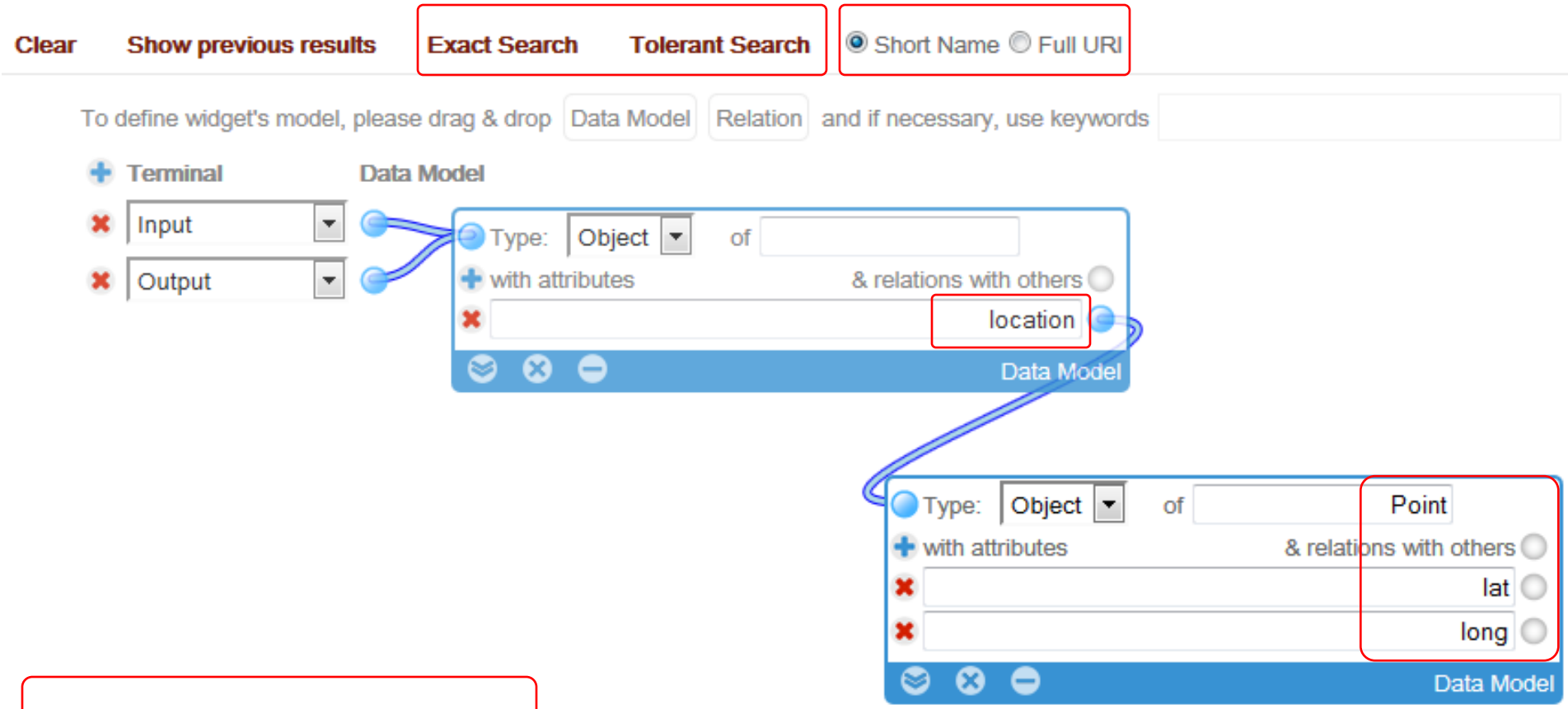
Server Linked Widgets run in a web server, PC, mobile phones, sensor, or embedded device


dcterms:LicenseDocument
string
Resource
string
schema:URL
string
dct:license
lw:hasName
dct:subject
dct:description
dct:identifier
lw:referenceSource
lw:clientWidget
lw:serverWidget
owl:oneOf
lw:environment
lw:dataWidget
lw:processingWidget
lw:visualizationWidget
owl:oneOf
lw:type
lw:Widget
lw:hasWidgetModel
lw:WidgetModel
...
Linked Widgets
Data Model
lw:hasOutput
lw:hasInput
lw:hasMetaAnnotation
...
Linked Widgets
Meta-tagging Model
lw:hasWidget
lw:hasWidget
prov:wasGeneratedBy
prov:wasDerivedFrom
lw:Widget
prov:wasAttributedTo
lw:WidgetCollection
prov:Activity
prov:wasAssociatedWith
lw:Developer
lw:hasCollection
rdfs:subClassof
lw:Mashup
prov:wasDerivedFrom
lw:Mashup
rdfs:subClassOf
lw:mashupConfiguration
string
foaf:Person
prov:wasGeneratedBy
prov:Activity
prov:wasDerivedFrom
prov:wasGeneratedBy
prov:wasAttributedTo
prov:wasAssociatedWith
rdfs:subClassOf
lw:MashupWidget
prov:wasAttributedTo
lw:MashupCreator

Example: Data model of the POI search widget

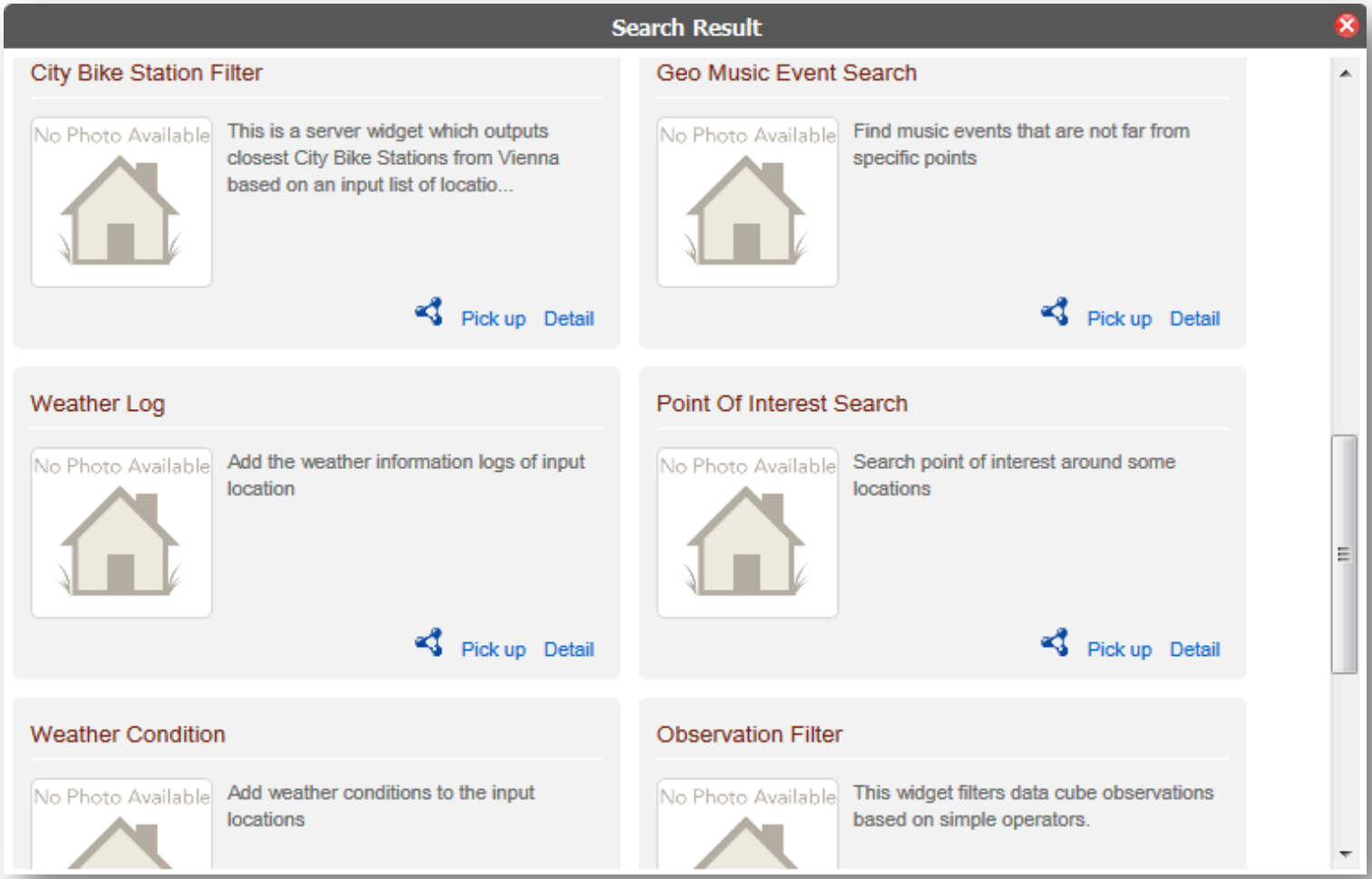


M1: Semantic Widget Discovery

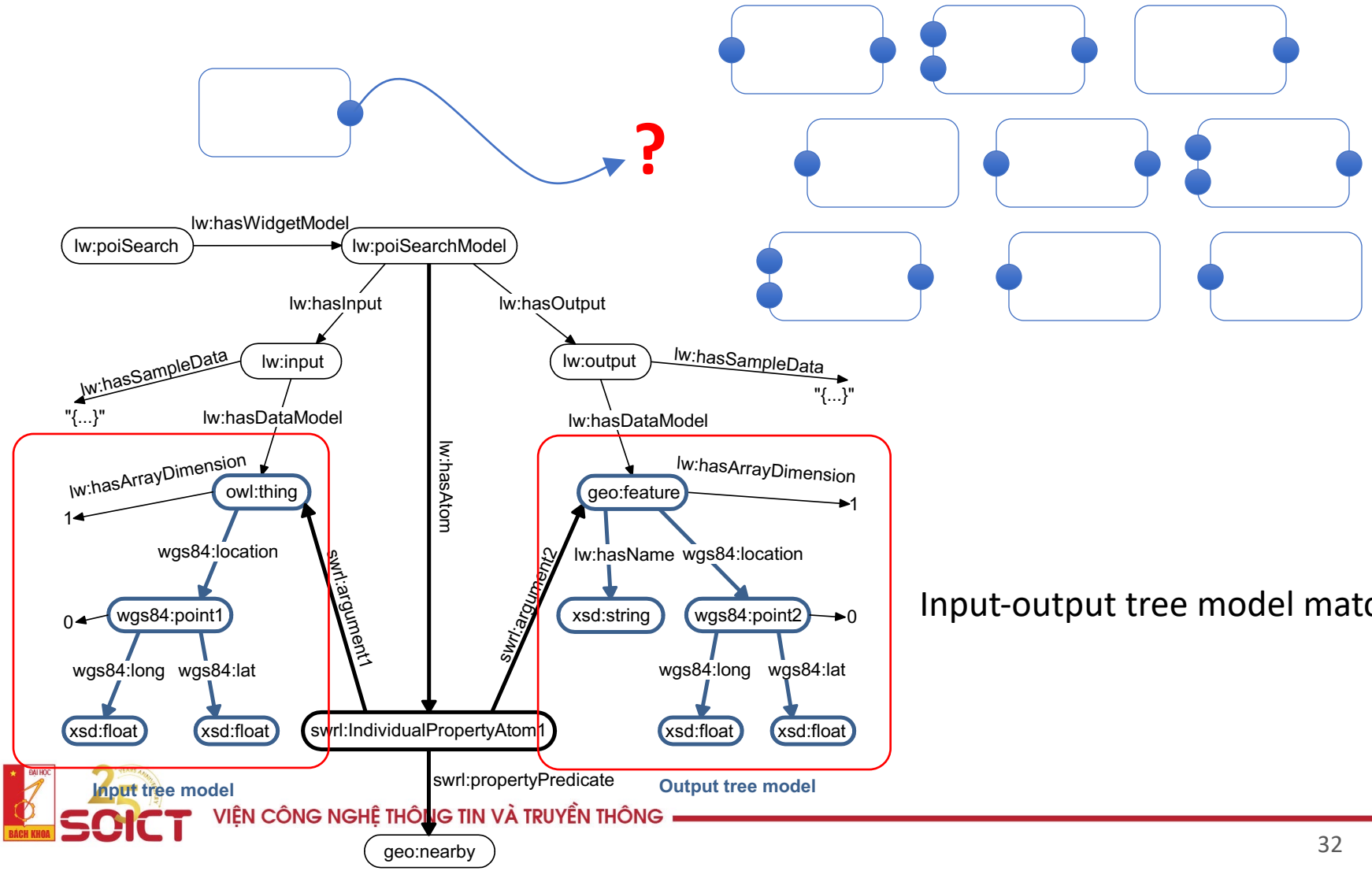


Ontology Alignment API:
http://alignapi.gforge.inria.fr/

M1: Semantic Widget Discovery



M2: Terminal matching



Input-output tree model matching

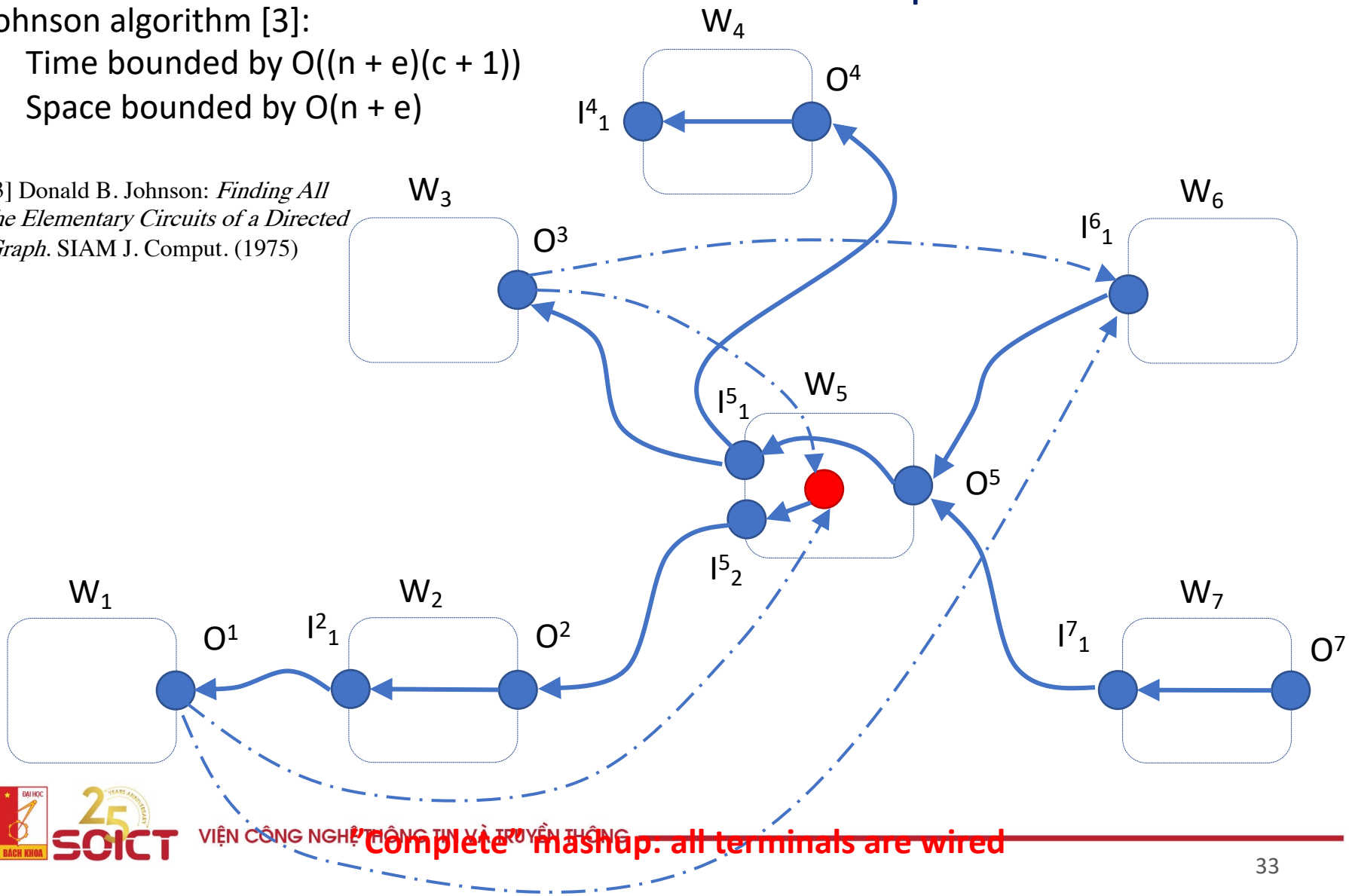
M3: Automatic mashup composition

Reduce "*automatic composition*" to "*list all cycles*" algorithm

Johnson algorithm [3]:

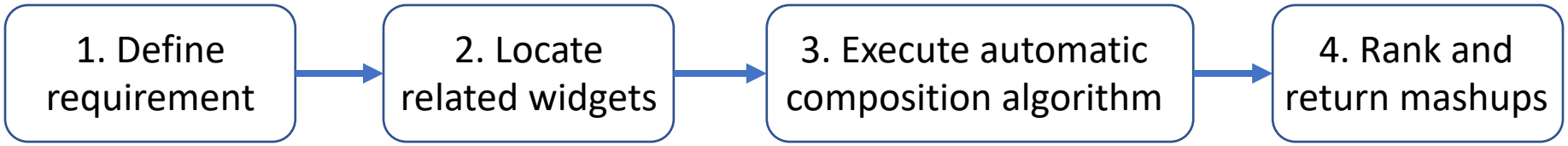
- Time bounded by O((n + e)(c + 1))
- Space bounded by O(n + e)

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



Complete mashup: all terminals are wired

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

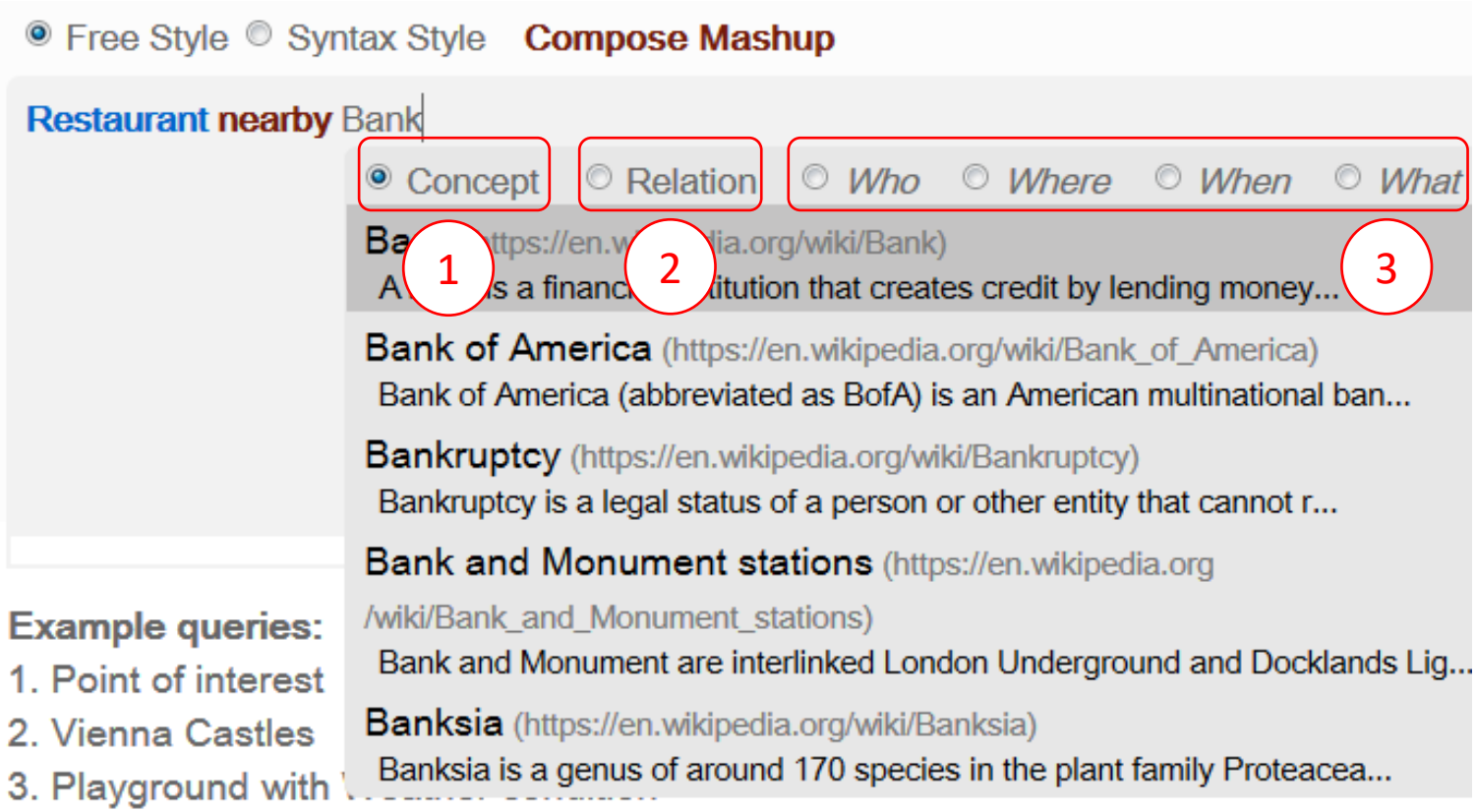
http://dbpedia.org/resource/Bank

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

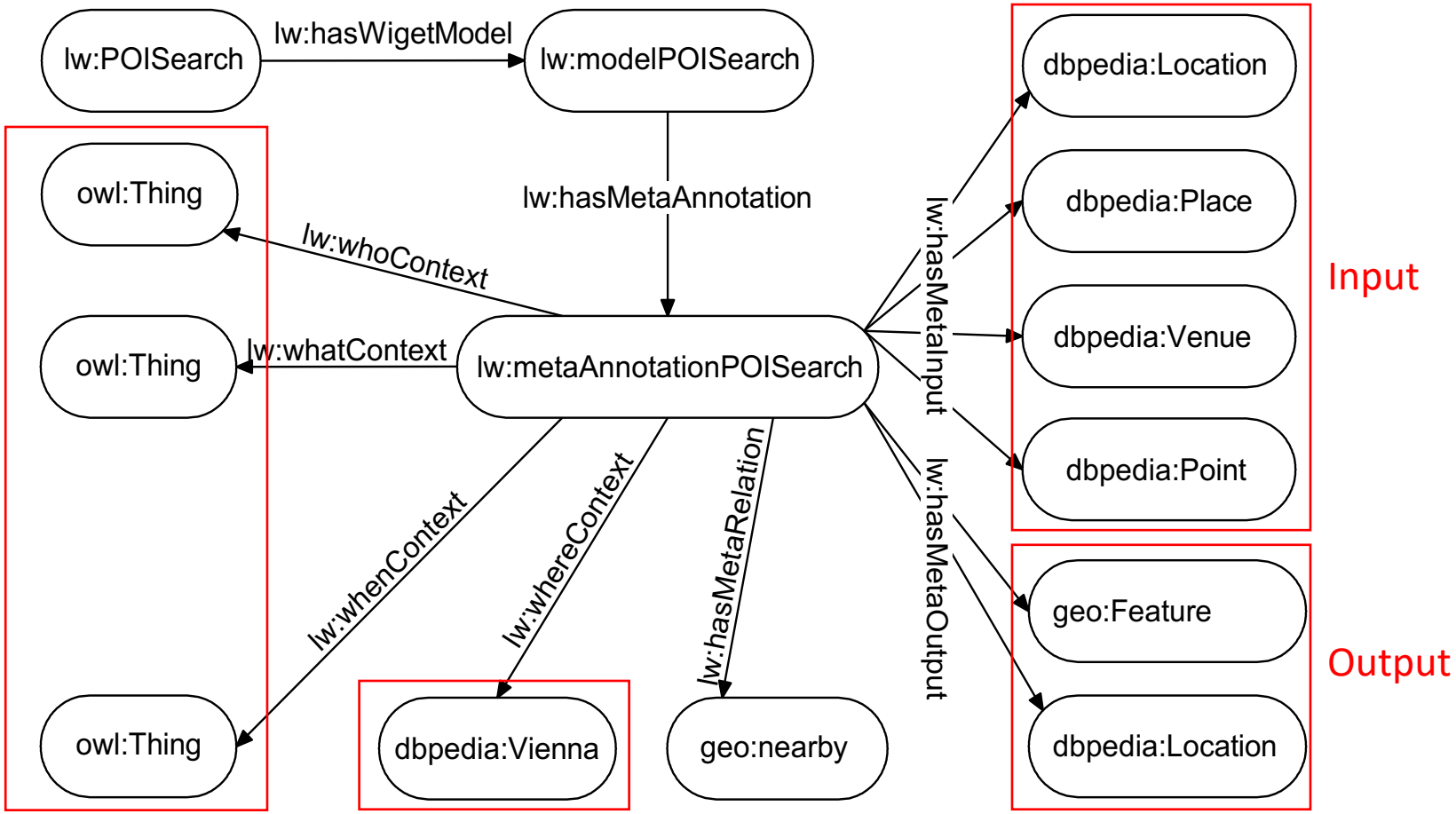


1. Concept tagging (DBpedia concepts)

2. Relation tagging (LOD properties)

3. Entity tagging DBpedia entities

Meta-tagging model

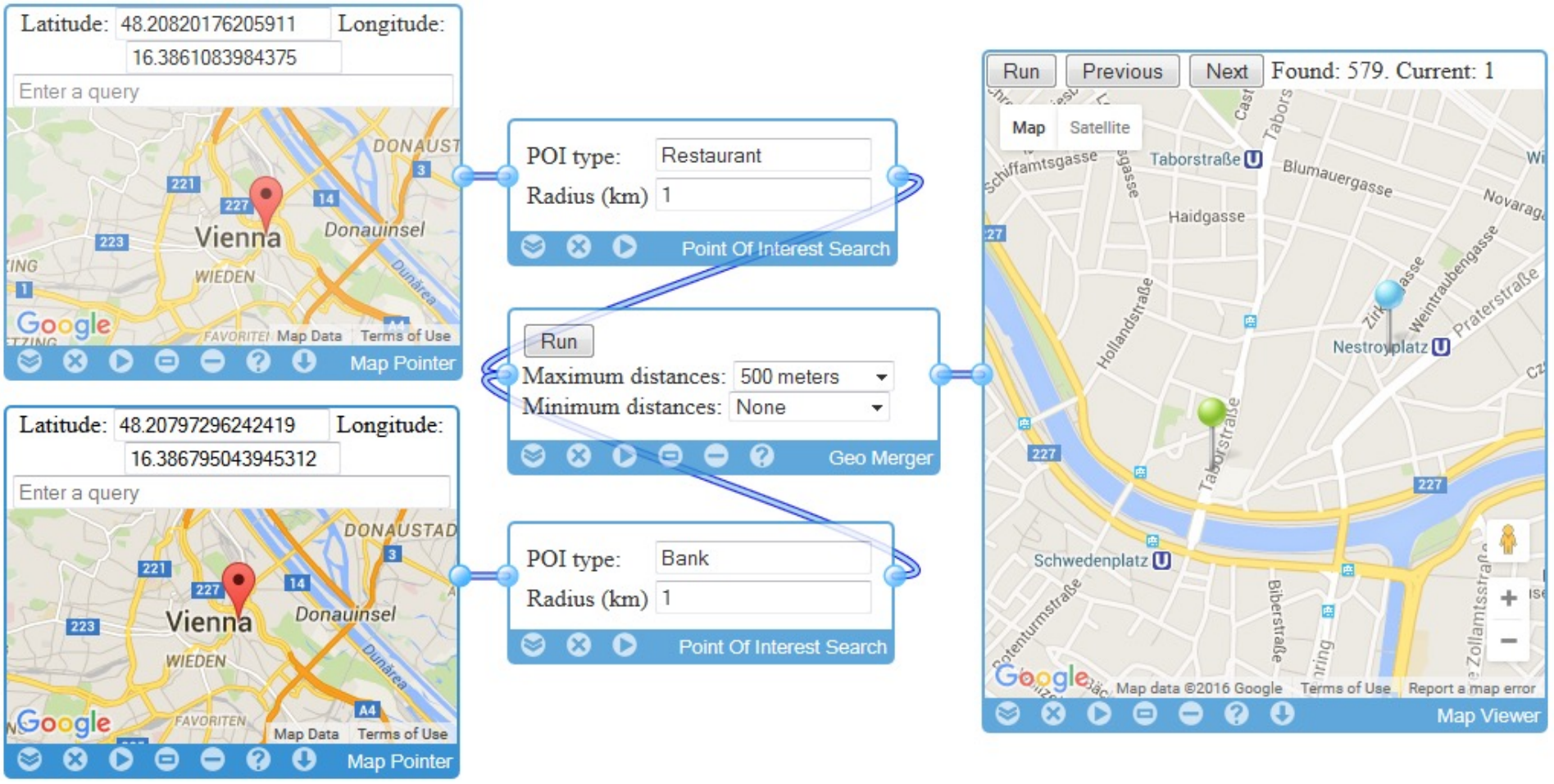


lw	http://linkedwidgets.org/ontologies/	**dbpedia**	http://dbpedia.org/resource/
owl	http://www.w3.org/2002/07/owl#	**geo**	http://www.geonames.org/ontology#

Example: Meta-tagging model of the Vienna POI Search widget

Example

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

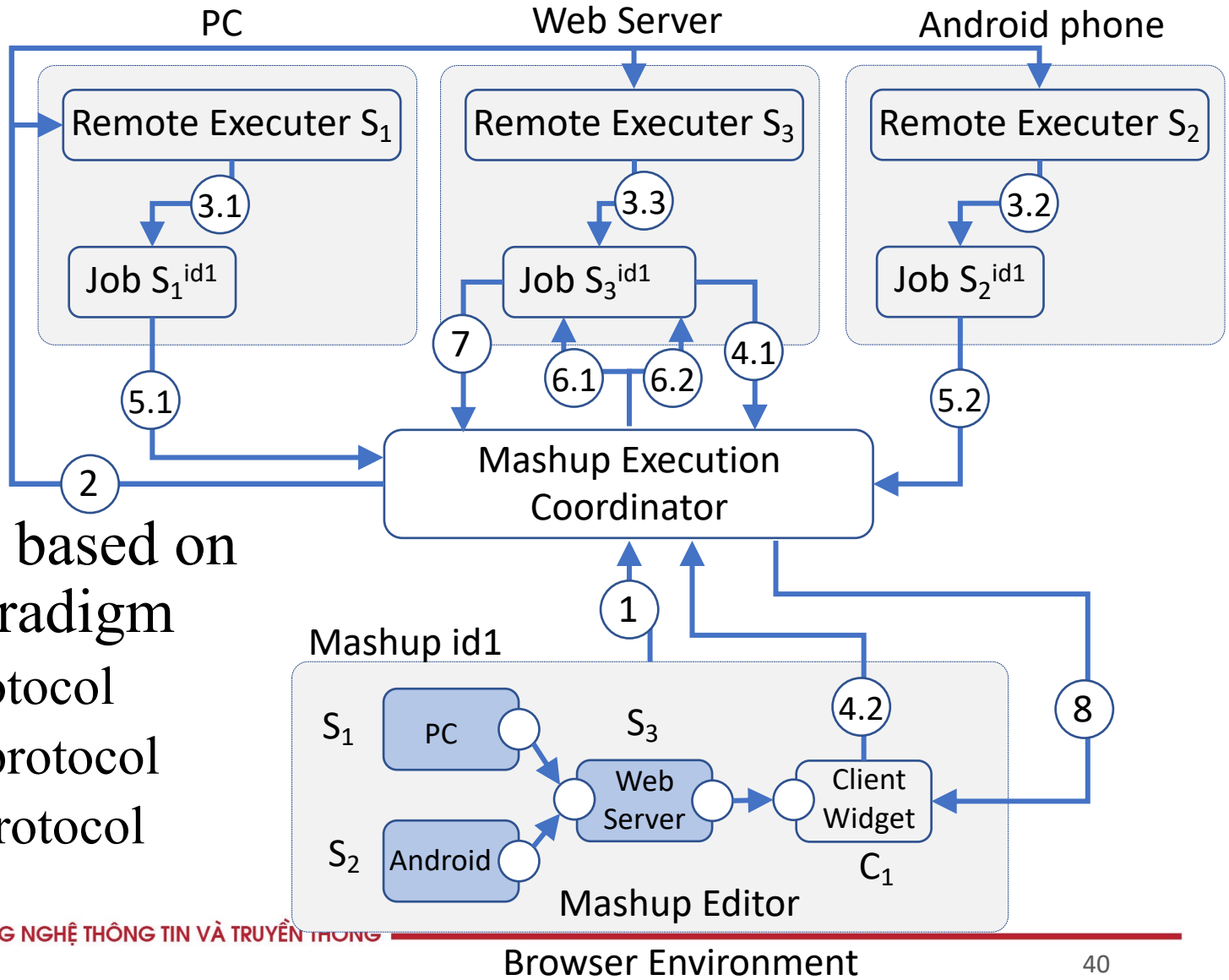


3. Mashup Models

3. Mashup models

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

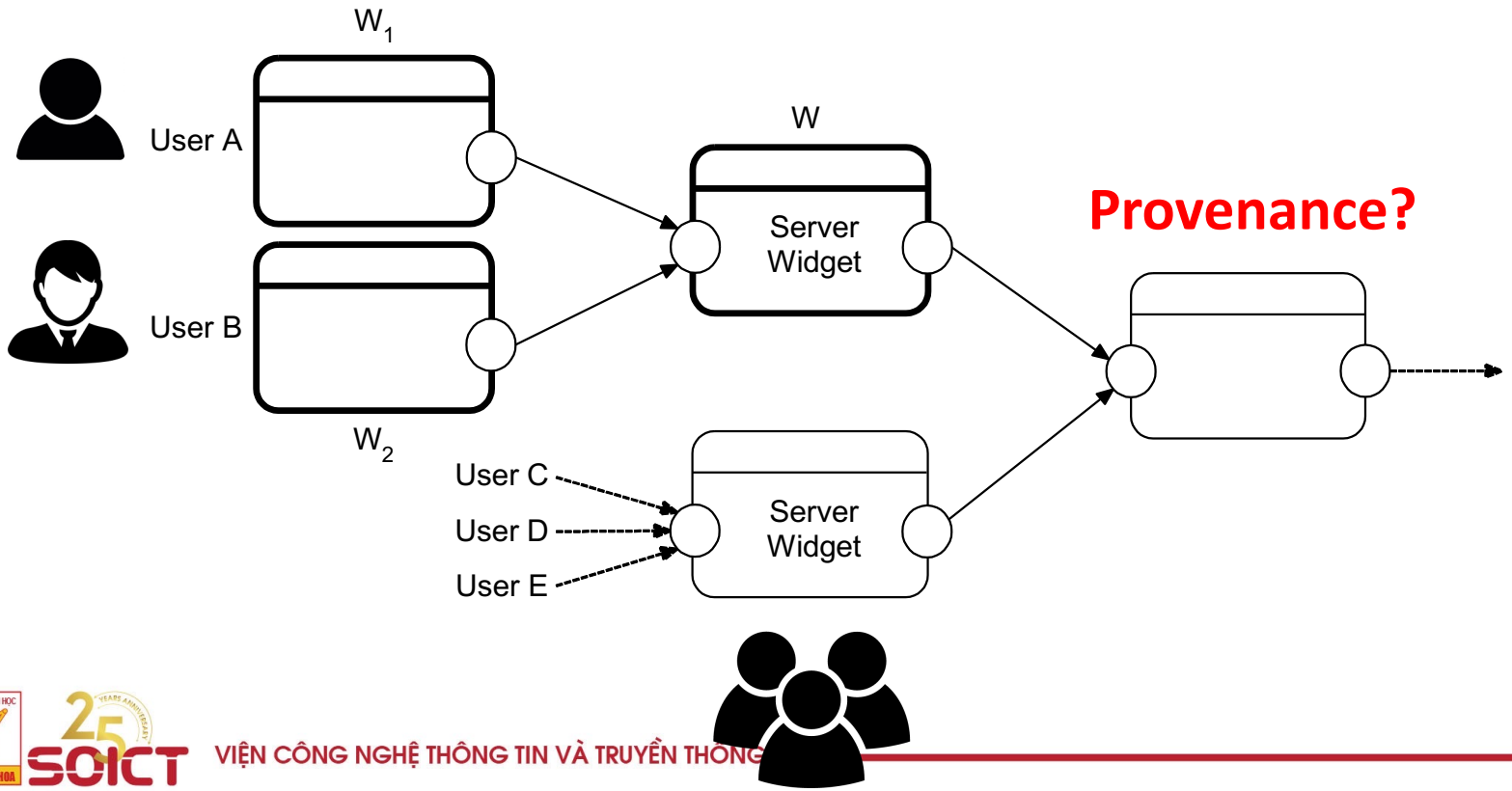
Linked Widget communication protocols



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

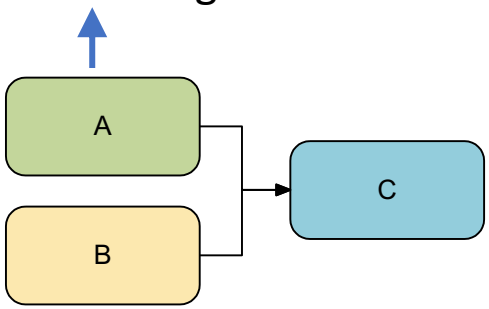
Model 1 – Collaborative mashups

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



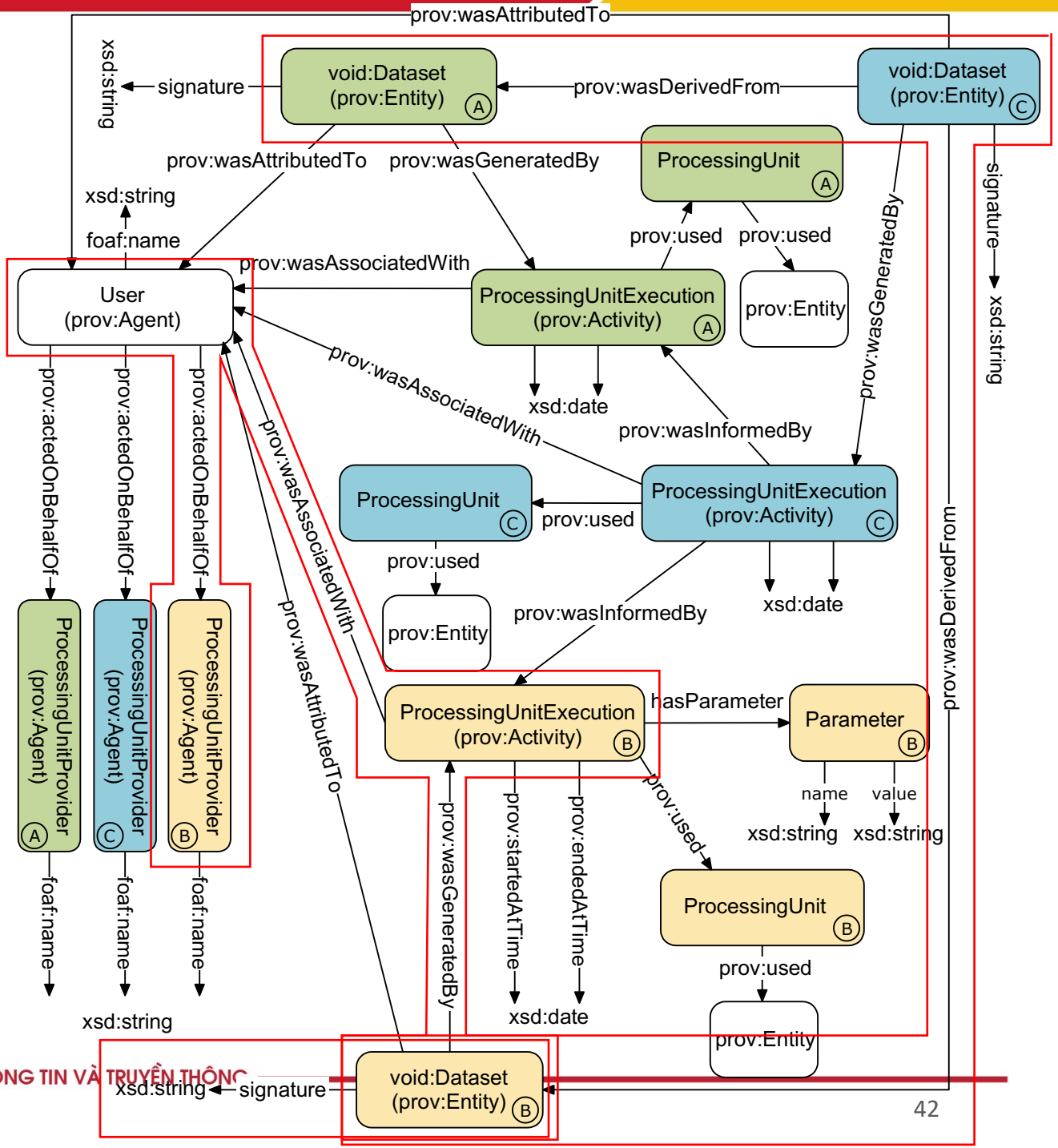
Provenance Model

Processing Unit



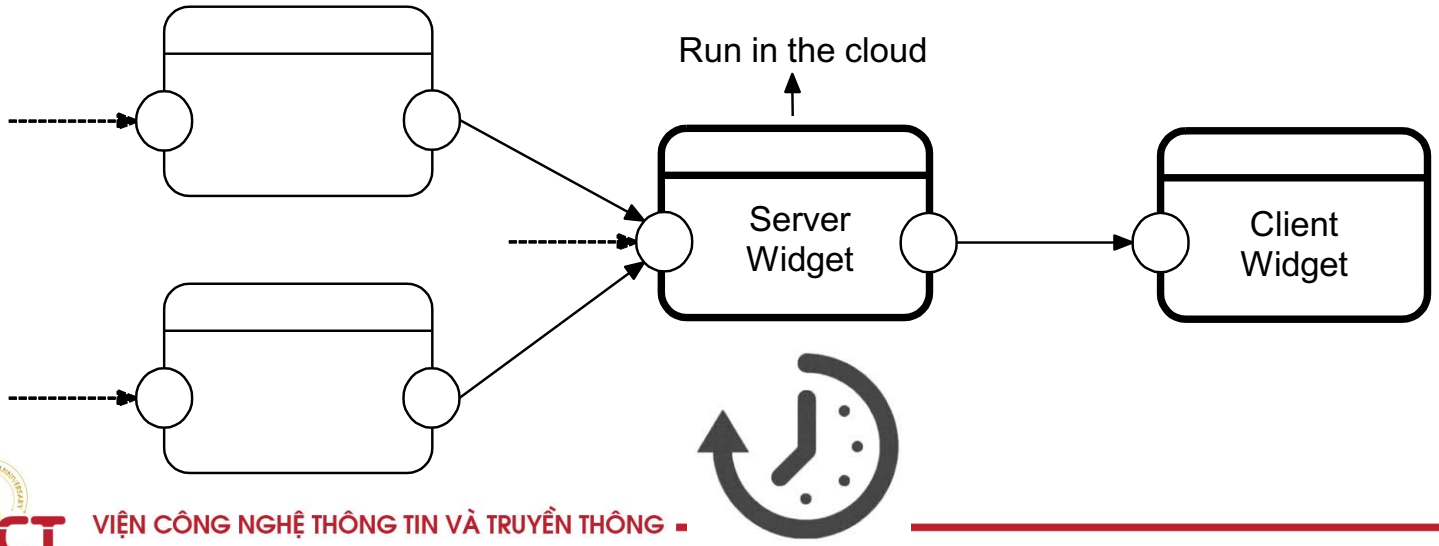
Workflow

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



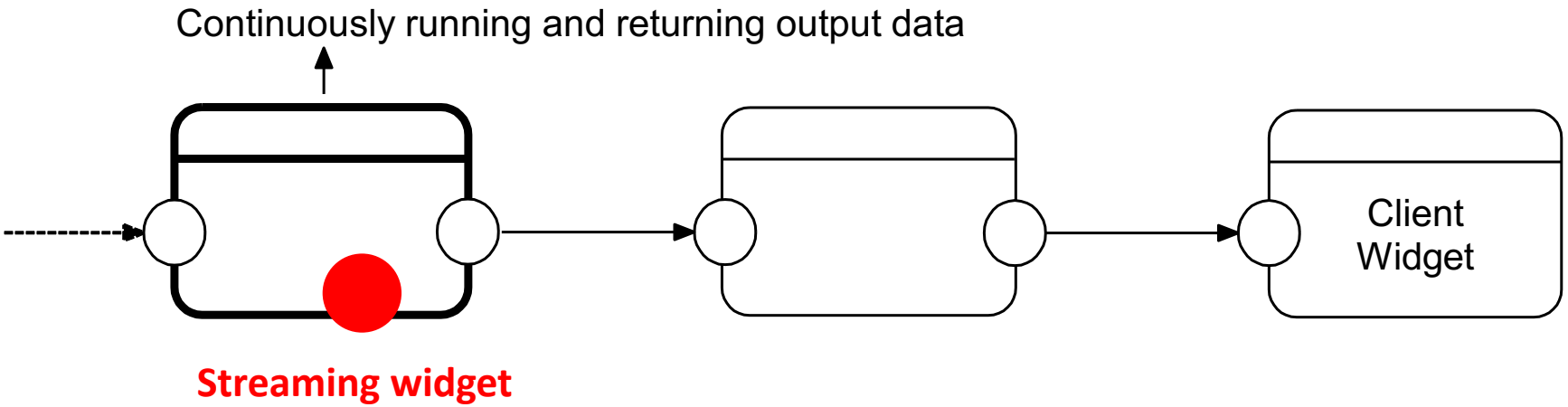
Model 2 – Persistent mashups

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



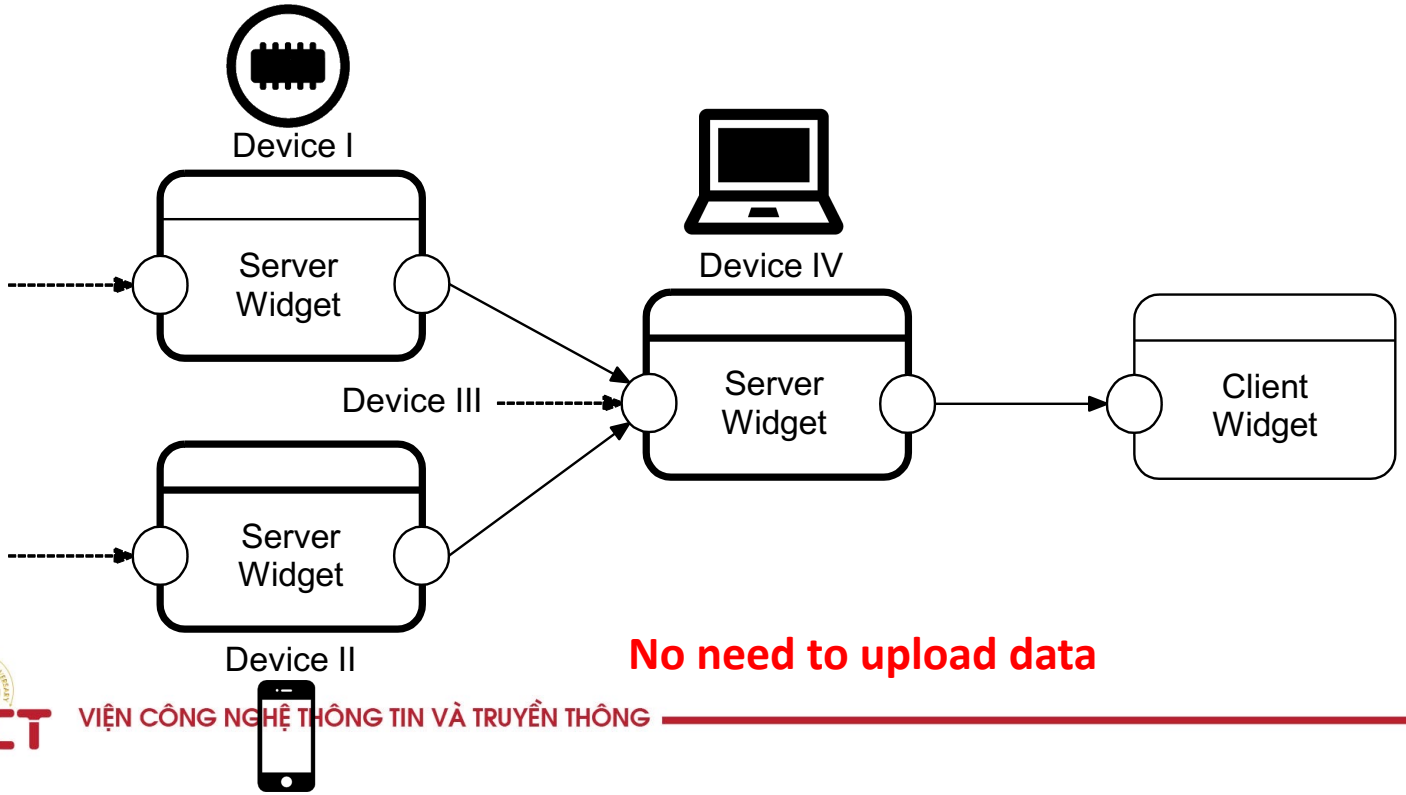
Model 3 – Streaming mashups

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



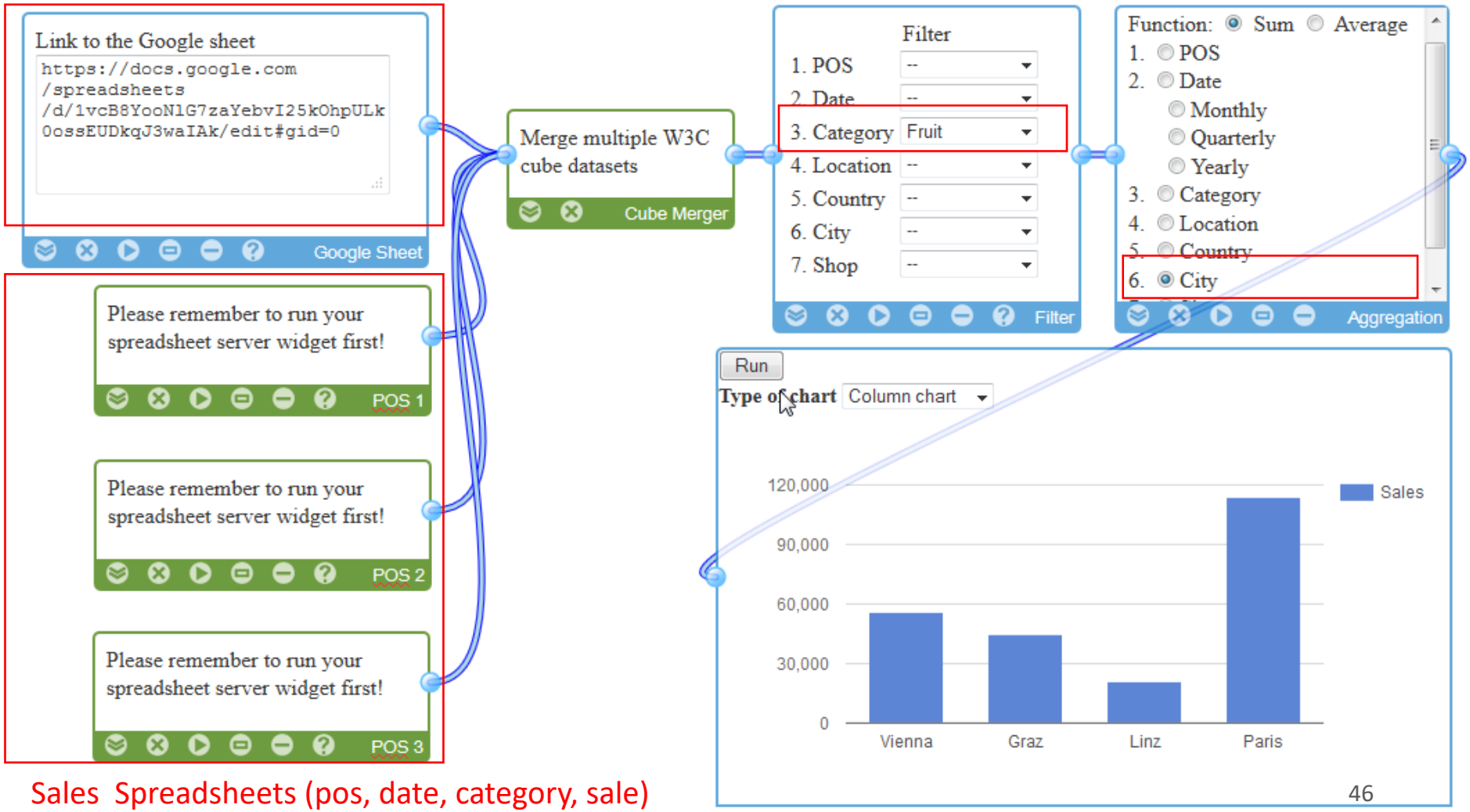
Model 4 – Distributed mashups

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



No need to upload data

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



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