



# Election Monitoring

- Parliamentary elections in Albania
- Citizen portal ZaLart.al (based on Ushahidi platform)
- Citizens can report issues via mobile phones etc.

## Reports by category

Violent Incidents	12
Damages to Electoral Materials or Offices	1
Unethical Electoral Speech	2
Problems with Electoral Commissions	9
Breach of the Right to Vote and the Anonymity of the Ballot	6
Tension near CEAZ/VCC	8
Problems delivering electoral materials	2
Misuse of the State Police	1
Unbalanced Media Coverage	3
Other problems	7



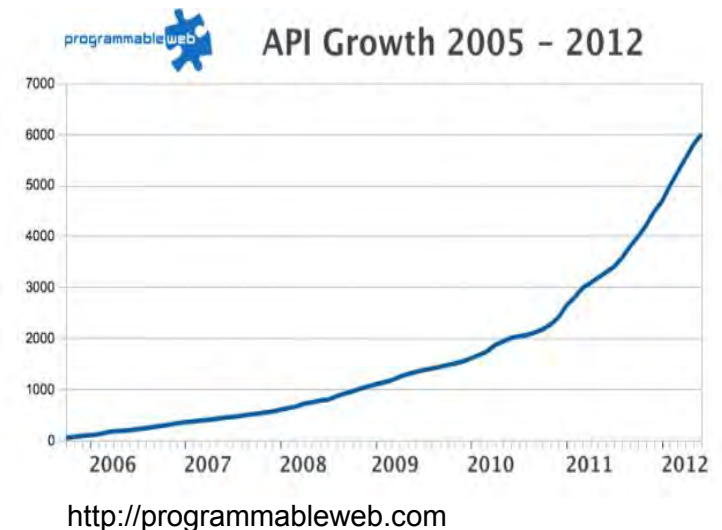
# OpenStreetMap Project Haiti



<http://vimeo.com/9182869>

# Issues in Complex Information Ecosystems

- The relevance of many types of data perishes or degrades over time (e.g., flood level, moving objects)
- A lot of websites provide access to their data via an API
- Data integration systems have to
  - Provide integrated access to live data
  - to be extendible for including new sources in an efficient manner



# Semantic Web Technologies

- Enable the transition from the “Web of Documents” to the “Web of Data”
- Technological basis: W3C Standards *Resource Description Framework (RDF)* and *Web Ontology Language (OWL)*
- Linked (Open) Data is “Semantic Web light”
  - Hitherto mostly focused on data publishing
  - Amount of available Linked Data is exploding
- Consuming and updating Linked Data is still to be improved
  - Semantics enable better means for data integration



# Dynamic Data to Decisions (D3)

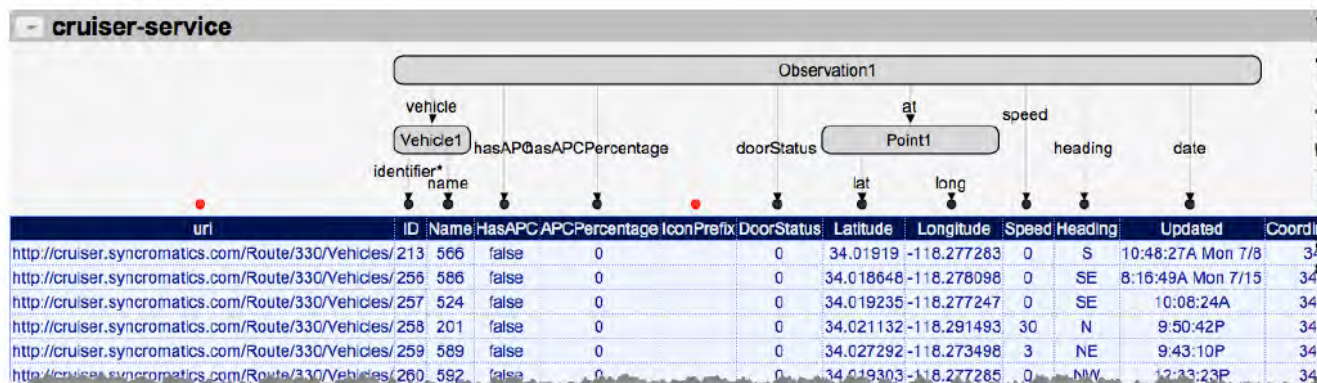
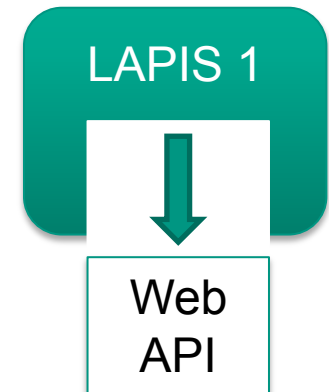
- Project in cooperation with USC/ISI
- Funded by ONR Global



- Objectives:
  - Real-time data integration for a variety of data sources in a range of data formats and access modalities
    - needed e.g. for election monitoring, crisis management and other quickly-evolving situations
  - Ability to quickly add sources (via modeling), see KARMA
  - Handling of geospatial data with temporal aspects

# Wrapping Sources with KARMA

- System acts on RDF sources
- However, we need to interact with all kinds of tabular and tree like data
- KARMA is a **visual data integration** tool from ISI/USC
  - Maps all kinds of sources onto an ontology to generate Linked Data

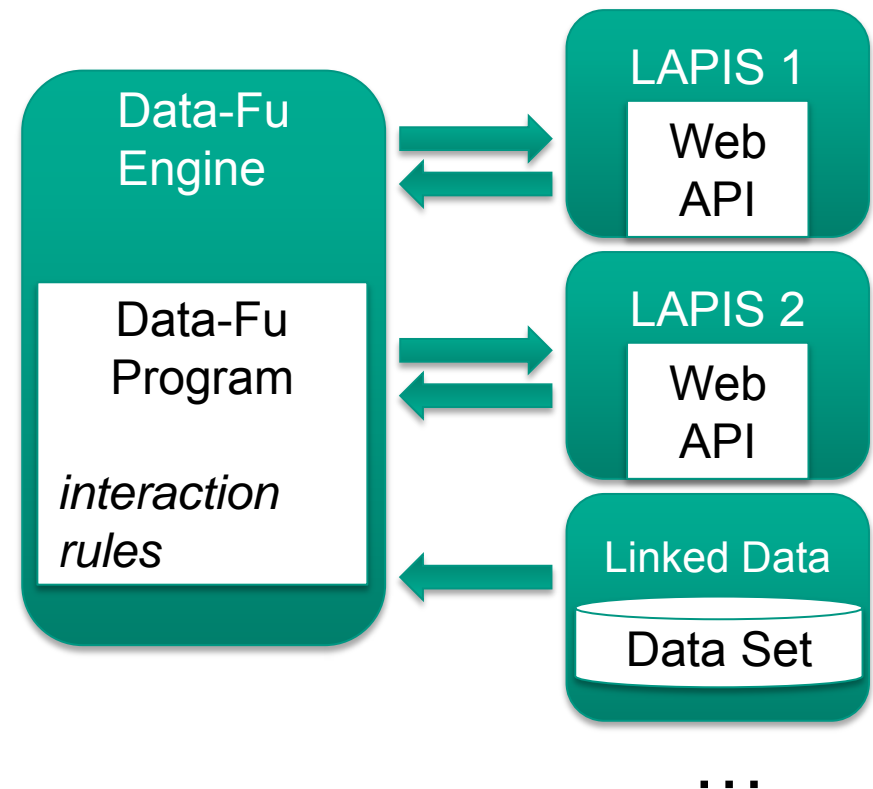


The screenshot shows the KARMA interface for a service named "cruiser-service". At the top, a visual ontology diagram shows a class "Observation1" with properties "vehicle", "at", "speed", "heading", and "date". "vehicle" is linked to a class "Vehicle1" which has properties "hasAPC", "asAPCPercentage", "doorStatus", and "name". "at" is linked to a class "Point1" which has properties "lat" and "long". Below the diagram is a table of data rows.

uri	ID	Name	HasAPC	APCPercentage	IconPrefix	DoorStatus	Latitude	Longitude	Speed	Heading	Updated	Coordi
http://cruiser.syncromatics.com/Route/330/Vehicles/213	566		false	0		0	34.01919	-118.277283	0	S	10:48:27A Mon 7/8	34
http://cruiser.syncromatics.com/Route/330/Vehicles/256	586		false	0		0	34.018648	-118.278098	0	SE	8:16:49A Mon 7/15	34
http://cruiser.syncromatics.com/Route/330/Vehicles/257	524		false	0		0	34.019235	-118.277247	0	SE	10:08:24A	34
http://cruiser.syncromatics.com/Route/330/Vehicles/258	201		false	0		0	34.021132	-118.291493	30	N	9:50:42P	34
http://cruiser.syncromatics.com/Route/330/Vehicles/259	589		false	0		0	34.027292	-118.273498	3	NE	9:43:10P	34
http://cruiser.syncromatics.com/Route/330/Vehicles/260	592		false	0		0	34.019303	-118.277285	0	NW	12:33:23P	34

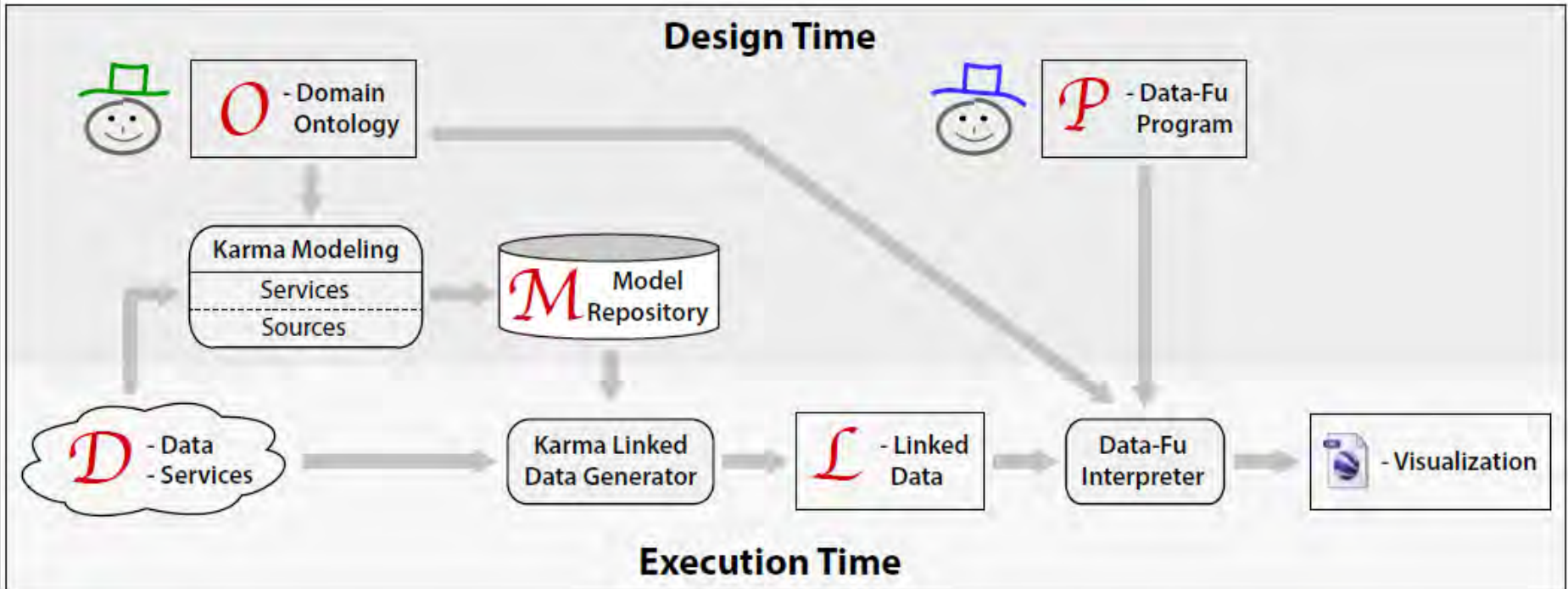
# Data-Fu - Overview

- Data-Fu is a **declarative rule language** to enable developers
  - to combine resources from several APIs
  - by flexibly defining the behaviour of the developed application (**dynamic reaction to states of resources at runtime**)
  - including the automated activation of links (i.e., that are discovered at runtime)
  - **scalable** execution

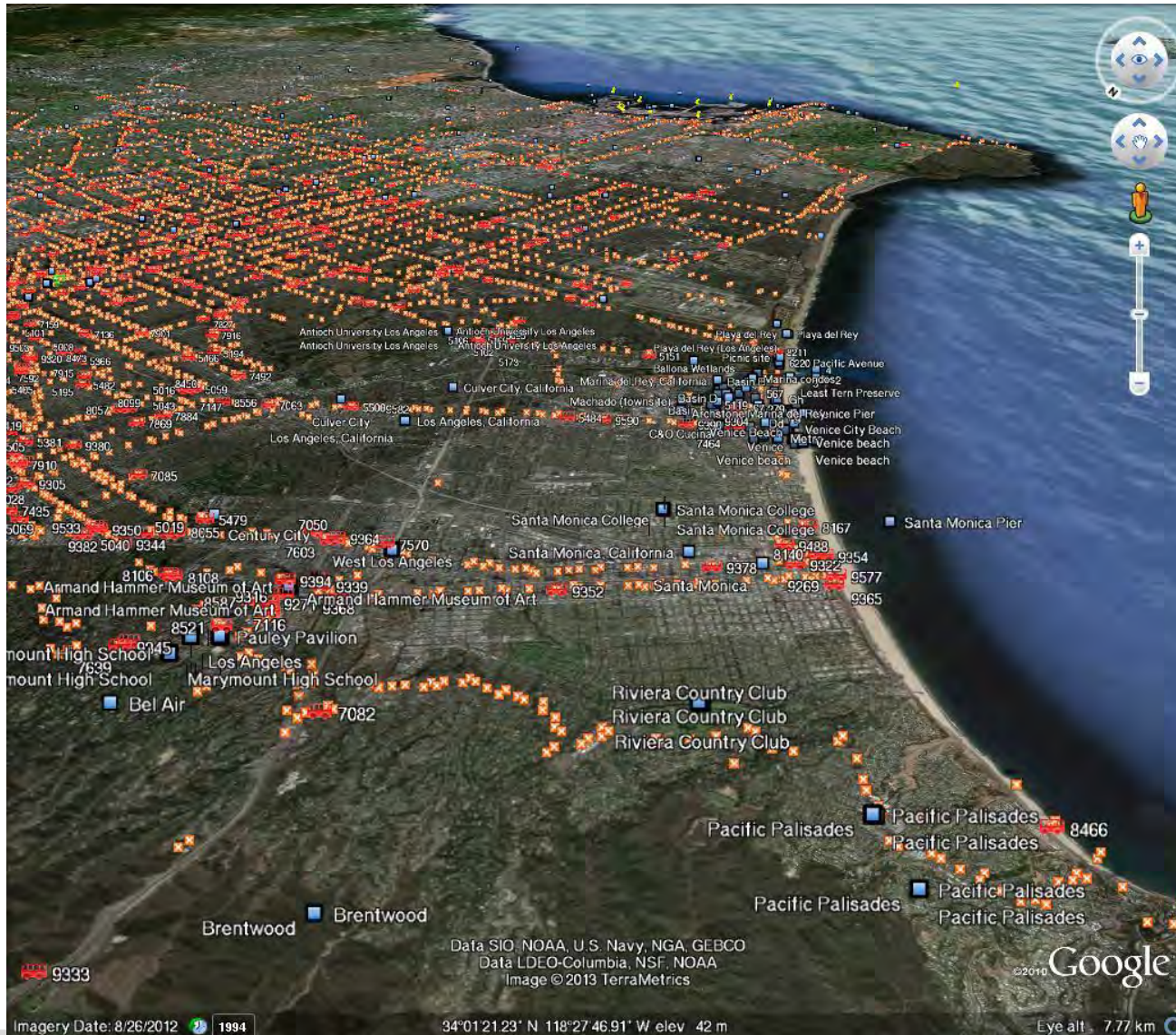




# System Architecture



# D3-Application: POIs, Events, Buses, ... in L.A.



## Enabled Actions

- Based on integrated data sources containing geospatial Regions, Points and georeferenced Events, Twitter Messages, Ushahidi, Epidemiological Records, Gnip data, ... you might realize application that ...
  - Notify me if tick-borne encephalitis cases in certain regions are over a given threshold
  - Notify me if there are online reports (in social media) of tick-borne encephalitis cases in high-risk areas
  - Notify me if the people density in an area rises above a certain threshold
  - Send me realtime reports about voter fraud from Ushahidi or from social media



# Crowdsourcing of Tasks: Data Quality Improvement

- Acting on data depends on quality of data
  - E.g. resolving entity identities
- **Example:** *Retrieve the ICAO identifier of METAR stations*

```
SELECT ?station ?icao WHERE {  
  ?station a metar:Station ;  
    owl:sameAs ?airport .  
  ?airport a dbp-owl:Airport;  
    dbp-ont:icaoLocationId ?icao .}
```

- Split consolidation task into automated and human computation (humans verify algorithm output)

## Identity Resolution



Are these the same airport?

A: Karlsruhe/Baden-Baden Airpark

B: BADEN AIRPARK

Yes

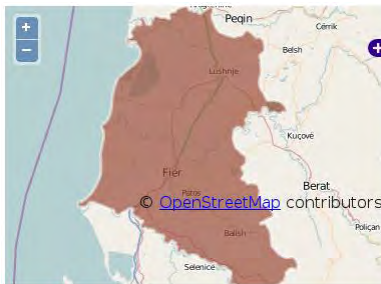
No

Source:

<http://www.slideshare.net/BarryNorton/crowdsourcing-tasks-in-linked-data-management>

# Linked Data for Data Access and Integration

- Identifiers for things
- E.g., Fier region is identified via [http://gadm.geovocab.org/id/1\\_71](http://gadm.geovocab.org/id/1_71)



- Access to [http://gadm.geovocab.org/id/1\\_71](http://gadm.geovocab.org/id/1_71) returns RDF description (name, description, geometry...)
- Other sources (e.g., DBpedia) can reuse the URI to reference unambiguously to Fier, facilitating querying over integrated data

## Albania Electoral Districts (Qarks)



<http://www.ndi.org/files/2013-Albania-Election-Watch-Report-May-22.pdf>

# Conclusion

- Goal: Interoperation in Complex Information Ecosystems
- Realtime access to data and reactive behaviour for actions
- Ability to rapidly integrate new sources via Karma models and Data-Fu programs
- Approach:
  - Simplicity based on basic web architecture (URI, HTTP, REST)
  - Simplicity based on basic data format (RDF)
  - Simplicity based on basic logic dialect (Datalog Rules)
  - Scalable interpreter suitable for parallelization
- Future work
  - Bandwidth issues: considering the “Disadvantaged Edge”
  - Integration of crowdsourcing, e.g. for data cleansing