Joshua Shinavier

Real-time interaction with the Web of Data

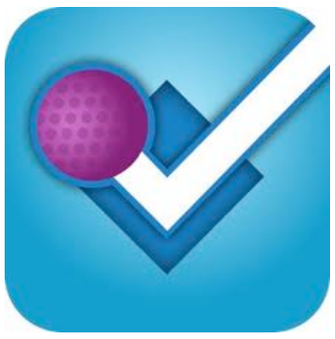
WIC Beijing invited talk

October 9th, 2010

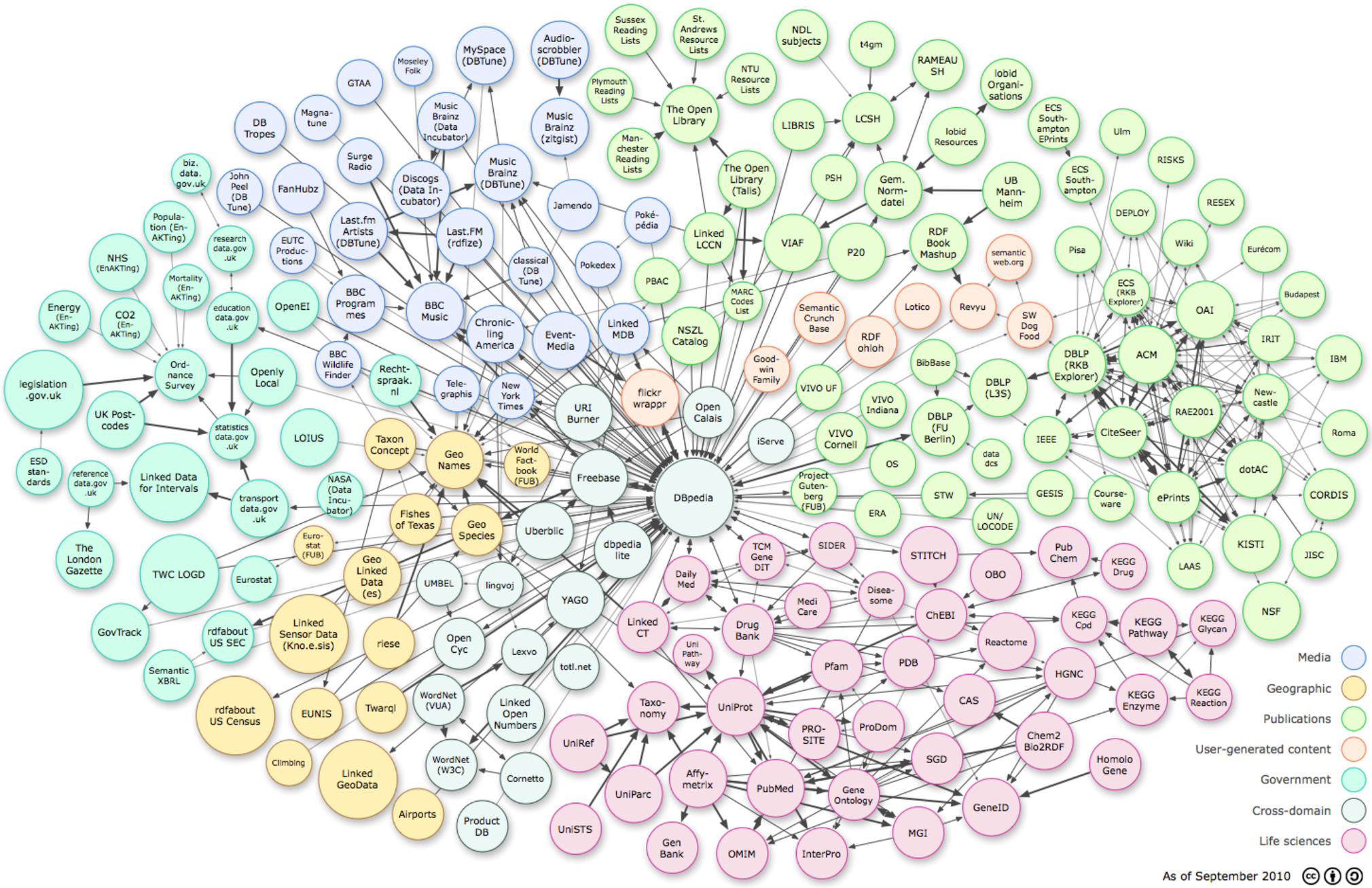
Overview

- real-time social data and the Semantic Web
- use cases
- RDF data streams
- real-time, real-place semantic search
- an application

The real-time Web

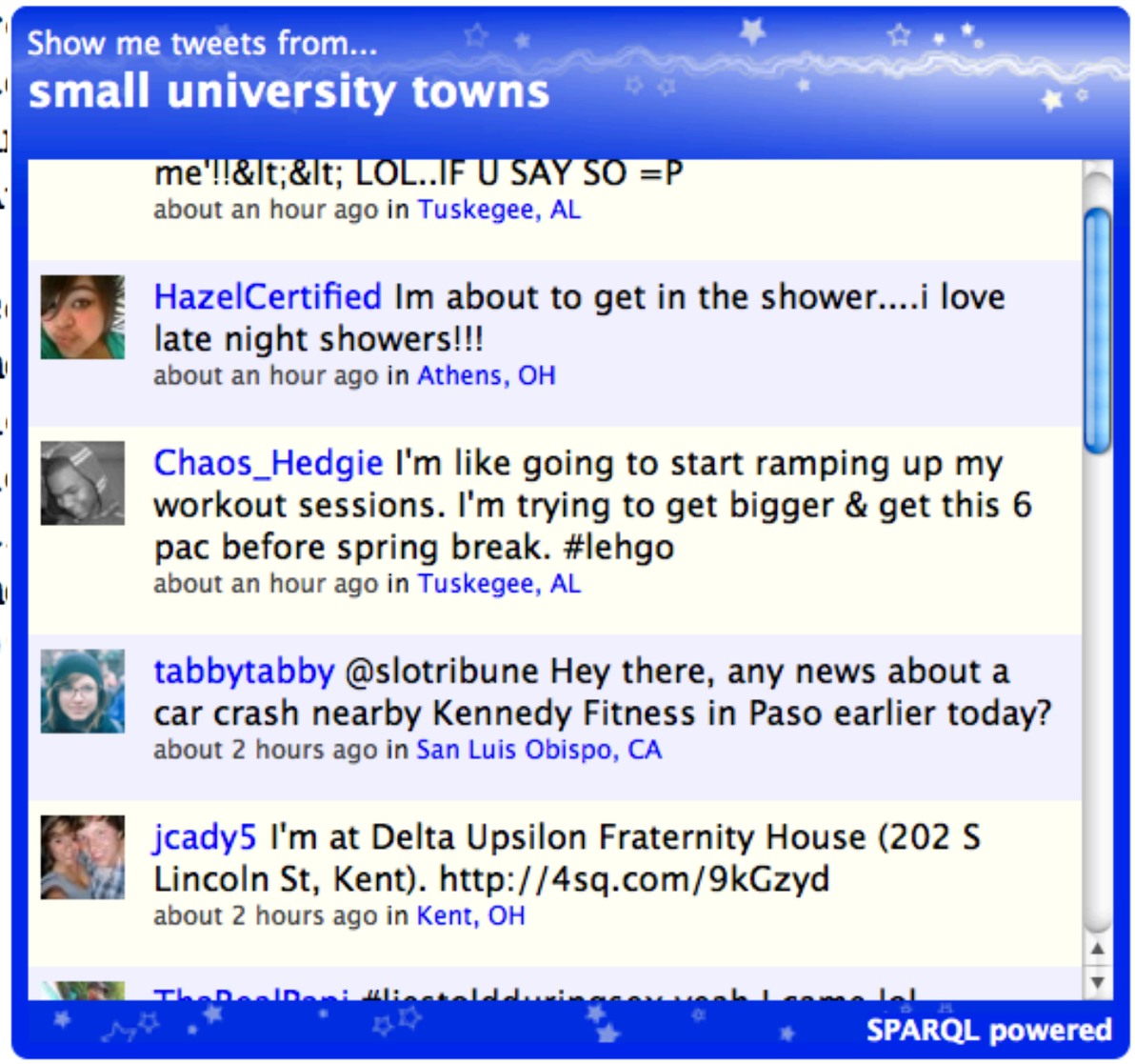


+ the Semantic Web



= the real-time Semantic Web

```
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX sioc: <http://rdfs.org/sioc/ns#>
PREFIX xsd: <http://www.w3.or
SELECT DISTINCT ?tweet ?creat
?profileImage ?place ?placeNa
  ?tweet dc:created ?createdA
  ?tweet sioc:content ?text .
  ?tweet sioc:has_creator ?ac
  ?account sioc:id ?screenNam
  ?account sioc:account_of ?a
  OPTIONAL { ?agent foaf:depi
  OPTIONAL { ?tweet geo:locat
    ?place dc:title ?placeNam
  FILTER(?createdAt > "2010-0
}
ORDER BY DESC(?createdAt)
LIMIT 15
```



Towards real-time Semantic Web

- let's mash up the "real-time" social Web with the Semantic Web
- requires rich, structured data
- services such as Facebook and Twitter are beginning to supply this data
  - e.g. Twitter Places/Annotations, Open Graph
- also: social sensor data (e.g. from mobile phones)
- the Semantic Web provides an ideal platform for data integration

Case in point: Twitter Annotations

- Annotations are structured data attached to a tweet
- data about movies, music, events, products, etc.
- expressed in JSON or XML
- constrained structure, open vocabulary
- coming soon!

Annotations example

```
...
"text":"Just watched Karate Kid (the new one, not the old one)",

"annotations":[
   {"movie":{
      "title":"The Karate Kid",
      "url":"http://www.rottentomatoes.com/m/karate_kid/",
      "text":"The Karate Kid (1984)"}},
   {"movie":{
      "title":"The Karate Kid",
      "url":"http://www.rottentomatoes.com/m/karate_kid_2010",
      "text":"The Karate Kid (2010)"}}],
...
```

Annotations + Semantic Web

- how do we interlink Annotations vocabularies?
- how do we interlink the data?
- how do we query over the data?
- these are Semantic Web problems
- why not:
  - map types into existing Semantic Web vocabularies
  - link Annotations resources into existing Linked Data datasets

Annotations types → ontologies

“webpage”	foaf:Document
“place”	geonames:Feature
“review”	rev:Review
“song”	mo:Track
“movie”	movie:film
“tvshow”	po:Series
“product”	gr:ActualProductOrServiceInstance
“offer”	gr:Offering
“event”	event:Event
...	...

Linking into the Semantic Web


```
...
"text":"Just watched Karate Kid (the new one, not the old one)",

"annotations":[
   {"movie":{
      "title":"The Karate Kid",
      "url":"http://www.rottentomatoes.com/m/karate_kid",
      "text":"The Karate Kid (1984)”,
      "sameas":"http://dbpedia.org/resource/The_Karate_Kid"}},
   {"movie":{
      "title":"The Karate Kid",
      "url":"http://www.rottentomatoes.com/m/karate_kid_2010",
      "text":"The Karate Kid (2010)",
      "sameas":"http://dbpedia.org/resource/The_Karate_Kid_
%282010_film%29"}}],
...
```

Without Semantic Web...

- “Show me tweets by my friends”
- “Show me tweets containing the term #china”
- “Show me tweets from Europe”
- etc.

“Show me tweets about…

…places in developing countries”


```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX sioc: <http://rdfs.org/sioc/ns#>
PREFIX geonames: <http://www.geonames.org/ontology#>
SELECT DISTINCT ?post WHERE {
    ?post sioc:embeds_knowledge ?graph .
    ?post sioc:topic ?place .
    GRAPH ?graph {
        ?place rdf:type geonames:Feature .
        ?place owl:sameAs ?geoplace .
    } .
    ?geoplace geonames:inCountry ?c .
    ?country geonames:inCountry ?c .
    ?country owl:sameAs ?dbc .
    ?dbc rdf:type <http://dbpedia.org/class/yago/LeastDevelopedCountries> .
}
```

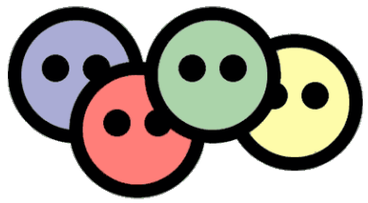
“Show me tweets about...

...English-language movies starring Chinese actors”


```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX sioc: <http://rdfs.org/sioc/ns#>
PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT DISTINCT ?post WHERE {
     ?post sioc:embeds_knowledge ?graph .
     ?post sioc:topic ?movie .
     GRAPH ?graph {
          ?movie rdf:type dbo:Film .
          ?movie owl:sameAs ?dbmovie .
     } .
     ?dbmovie dbo:language <http://dbpedia.org/resource/English_language> .
     ?dbmovie dbo:starring ?actor .
     ?actor rdf:type <http://dbpedia.org/class/yago/ChineseActors> .
}
```

“Show me tweets about...

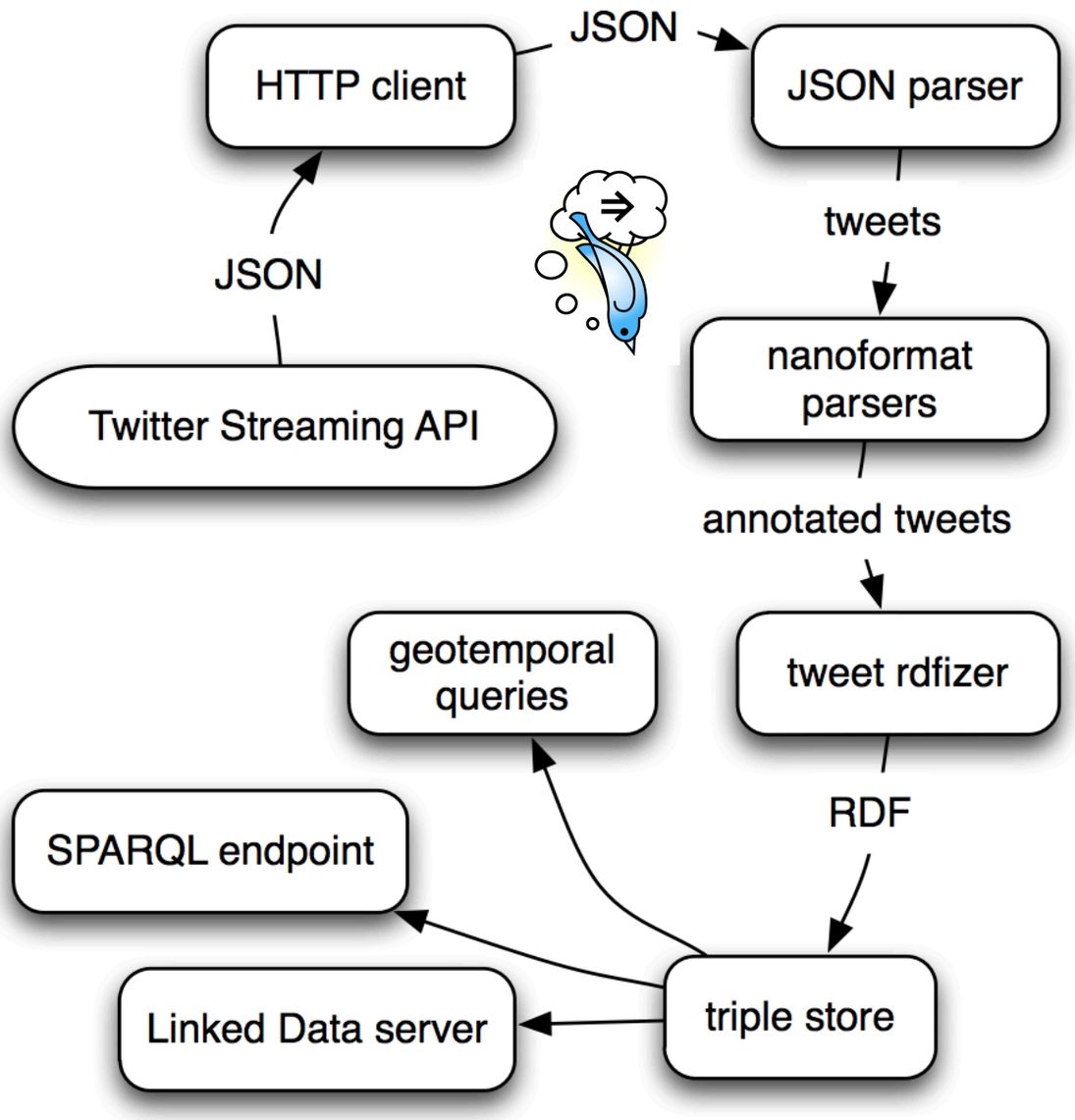
...songs by artists my friends like”



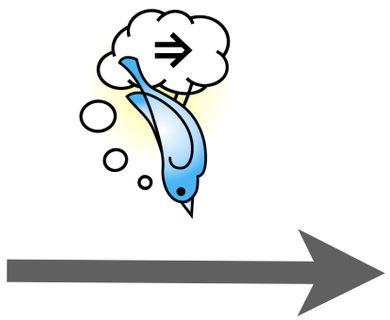
FOAF

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX sioc: <http://rdfs.org/sioc/ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?post WHERE {
     ?post sioc:embeds_knowledge ?graph .
     ?post sioc:topic ?song .
     GRAPH ?graph {
          ?song rdf:type mo:Track .
          ?song owl:sameAs ?mbtrack .
     } .
     ?album mo:track ?mbtrack .
     ?artist foaf:made ?album .
     ?friend foaf:interest ?artist .
     ?me foaf:knows ?friend .
}
```

Social data streams to RDF streams



- capture social data
- translate it into RDF
- produce an RDF stream
- publish it as Linked Data
  - preserve attribution
- provide real-time search



```
@prefix sioc: <http://rdfs.org/sioc/ns#> .
[...]

{
   <http://twitlogic.fortytwo.net/post/twitter/16810997455>
      a sioct:MicroblogPost ;
      dc:created "2010-06-23T00:43:35.000Z"^^xsd:dateTime ;
      geo:location <http://twitlogic.fortytwo.net/location/twitter/d0fc0f618c1eb790> ;
      sioc:content "the Karate Kids: karate vs. kung fu..." ;
      sioc:has_creator <http://twitlogic.fortytwo.net/user/twitter/7083182> ;
      sioc:topic
         <http://twitlogic.fortytwo.net/topic/2121544629>,
         <http://twitlogic.fortytwo.net/topic/1462827592> ;
      sioc:embeds_knowledge <http://twitlogic.fortytwo.net/graph/twitter/16810997455> .

   <http://twitlogic.fortytwo.net/graph/twitter/16810997455>
       a rdfg:Graph .

   <http://twitlogic.fortytwo.net/location/twitter/d0fc0f618c1eb790>
      a geonames:Feature , <http://dbpedia.org/resource/City> ;
      dc:title "Half Moon Bay, CA" ;
      geo:lat "32.528832"^^xsd:double ;
      geo:long "-124.482003"^^xsd:double ;
      rdfs:seeAlso <http://api.twitter.com/1/geo/id/d0fc0f618c1eb790.json> ;
      rdfs:label "Half Moon Bay" ;
      geonames:countryCode "US" .
}

<http://twitlogic.fortytwo.net/graph/twitter/16810997455> {
   <http://twitlogic.fortytwo.net/topic/2121544629>
       a dbo:Film ;
       dc:title "The Karate Kid" ;
       rdfs:label "The Karate Kid (1984)" ;
       foaf:page <http://www.rottentomatoes.com/m/karate_kid> ;
       owl:sameAs <http://dbpedia.org/resource/The_Karate_Kid> .
   <http://twitlogic.fortytwo.net/topic/1462827592>
       a dbo:Film ;
       dc:title "The Karate Kid" ;
       rdfs:label "The Karate Kid (2010)" ;
       foaf:page <http://www.rottentomatoes.com/m/karate_kid_2010> ;
       owl:sameAs <http://dbpedia.org/resource/The_Karate_Kid_%282010_film%29> .
}
```

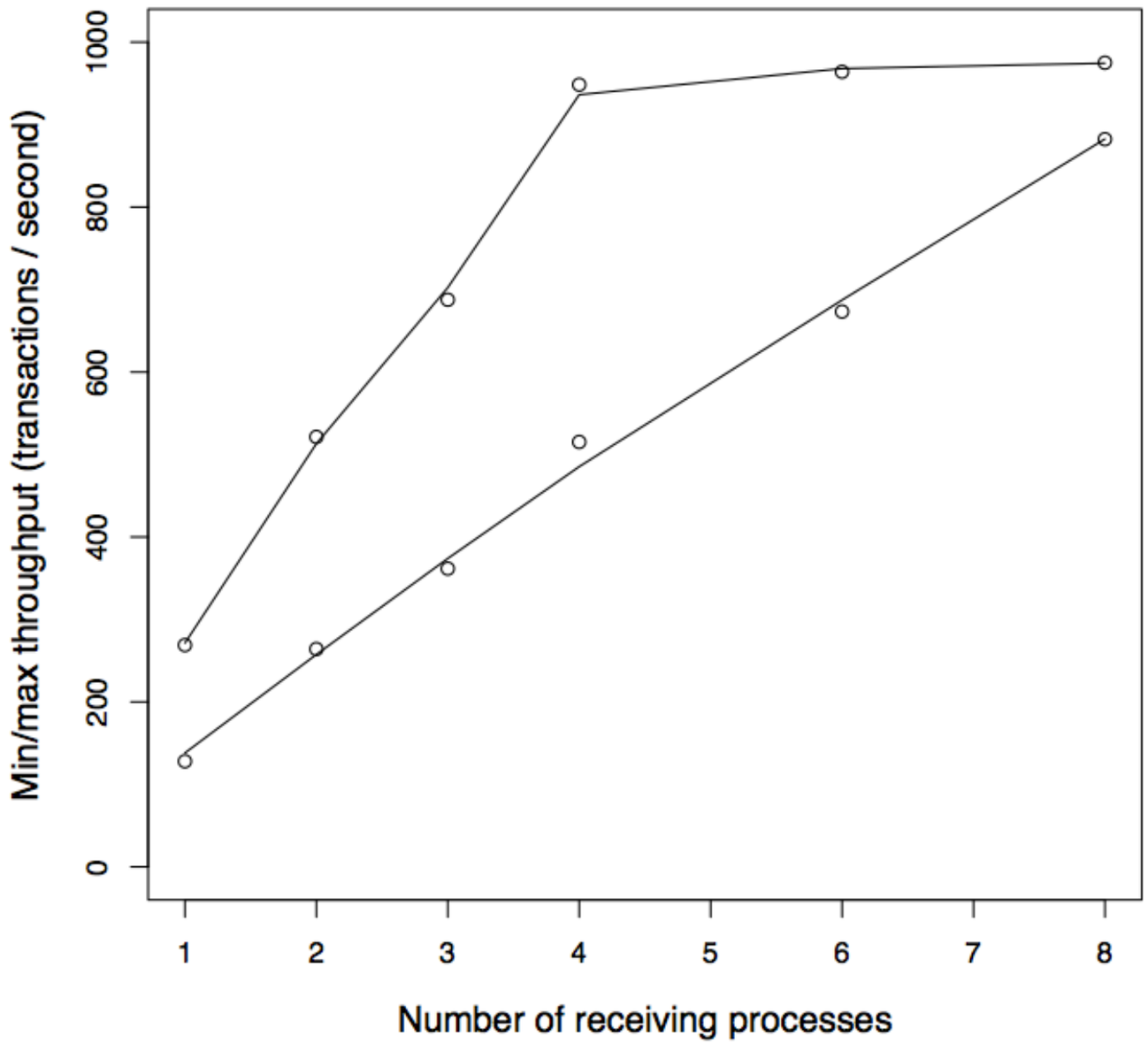
Keeping track of who said what

- in the social web, users own their content
- when mashing up data from different sources…
  - we need to preserve statement provenance
  - use Named Graphs with provenance vocabularies
- copyright information preserves authors' rights
  - attribution metadata for each statement
  - Open Data Commons license

Real-time RDF update streams

- transport dynamic RDF data as efficiently as possible
  - low latency
  - high throughput
- numerous RDF update formats are available
  - e.g. SPARQL Update
- HTTP protocol is most common
- UDP provides high performance but imposes some constraints

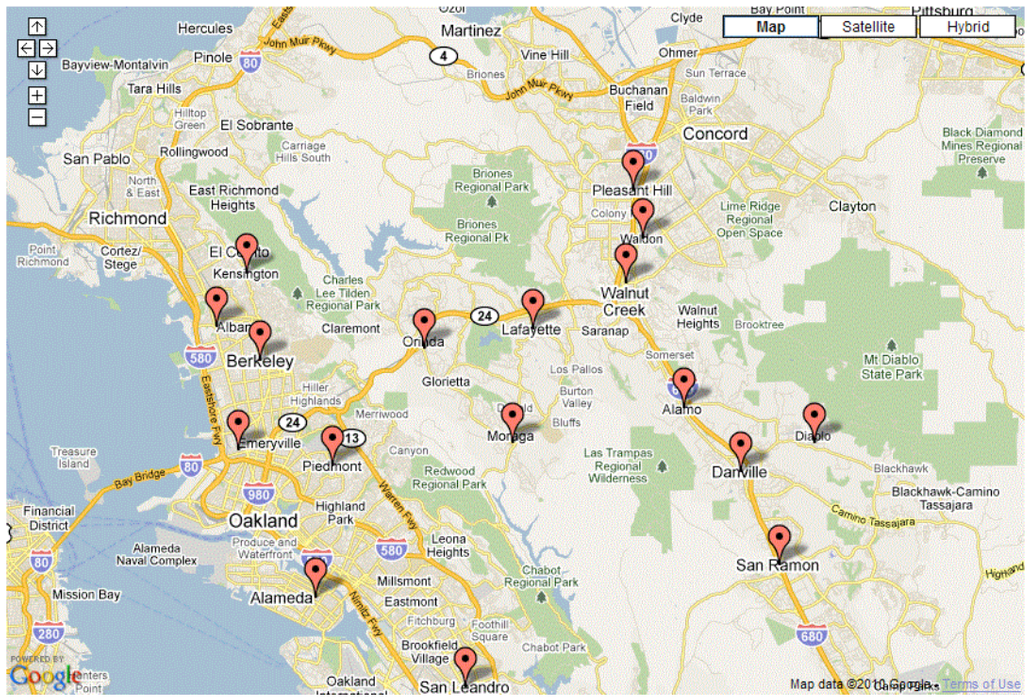
High-throughput graph databases: we're ready for the Twitter Firehose!



Real-time, real-place semantic search

- Geospatial SPARQL in Allegrograph:
  - order by distance
  - query for points within a geospatial circle, bounding box, or polygon
  - temporal reasoning
  - temporal reasoning

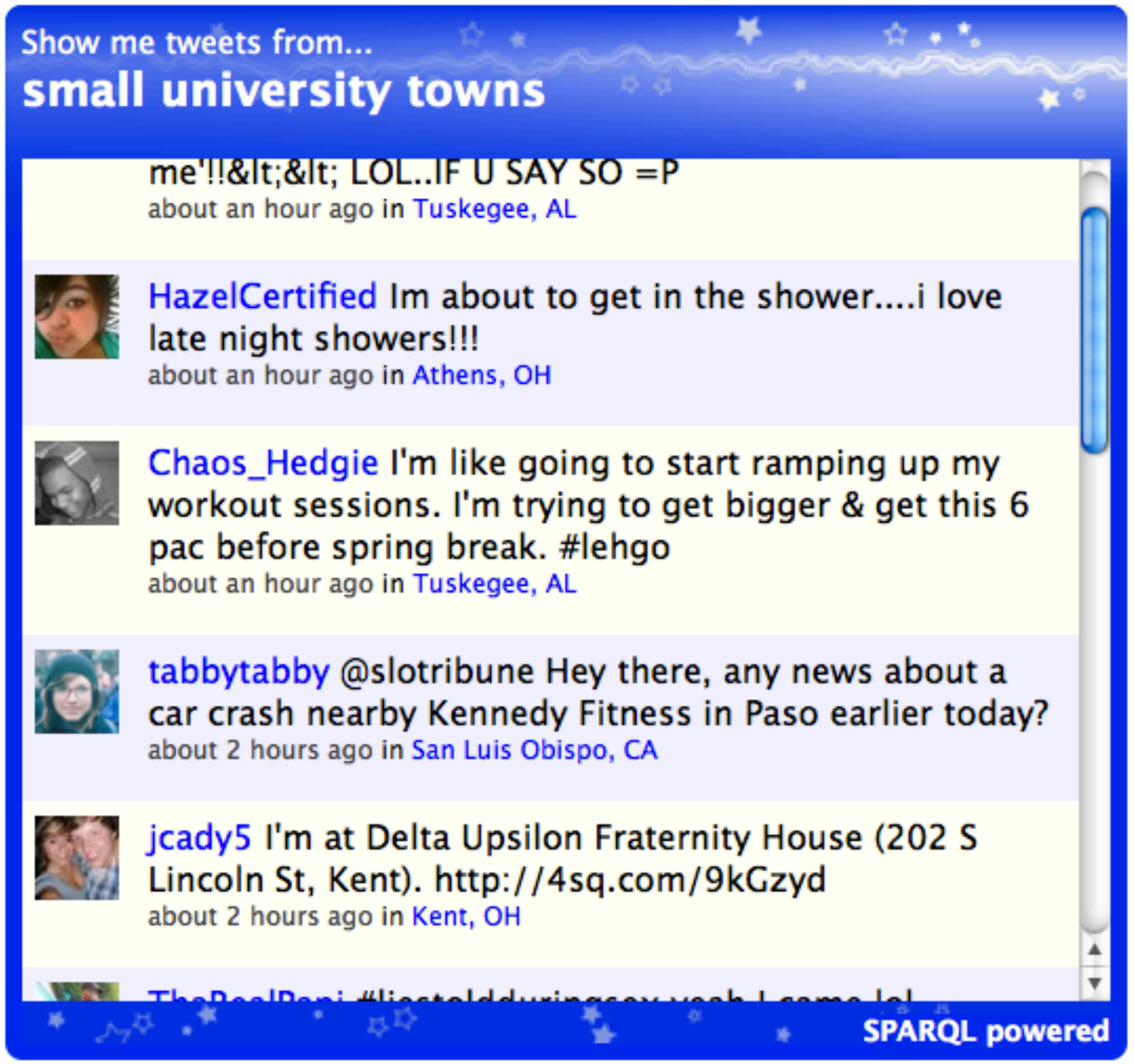
```
(google-map (select (?name ?lat ?lon)
   (q ?t !sioc:has_creator ?person)
   (q ?t !wgs84:location ?x)
   (q ?t !dc:created ?time)
   (temporal-within ?time "20100622T10:15:00"
                          "20100624T10:15:00")
   (geo-within-radius ?x ?y 10 miles)
   (q ?y !geo:asciiname ?name)
   (q ?y !geo:isAt5 ?pos)
   (pos->lon/lat ?pos ?lon ?lat)))
```



A demo: now we can use SPARQL queries...

```
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX sioc: <http://rdfs.org/sioc/ns#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT DISTINCT ?tweet ?createdAt ?text ?screenName
?profileImage ?place ?placeName WHERE {
  ?tweet dc:created ?createdAt .
  ?tweet sioc:content ?text .
  ?tweet sioc:has_creator ?account .
  ?account sioc:id ?screenName .
  ?account sioc:account_of ?agent .
  OPTIONAL { ?agent foaf:depiction ?profileImage . } .
  OPTIONAL { ?tweet geo:location ?place .
    ?place dc:title ?placeName . }.
  FILTER(?createdAt > "2010-09-29T13:51:56.000Z"^^xsd:dateTime) .
}
ORDER BY DESC(?createdAt)
LIMIT 15
```

...to support “smarter” content feeds



Conclusion

- much real-time data is compatible with the Semantic Web
- mashups with the Semantic Web benefit:
  - the end user (--> smarter applications)
  - the application developer (--> less case-by-case development)
  - the Semantic Web itself (--> “long tail” of real-time data)
- existing Semantic Web tools are ready for highly demanding applications

Thanks!

- Tetherless World Constellation
  - http://tw.rpi.edu
- Franz Inc. (makers of AllegroGraph)
  - http://www.franz.com
- These slides
  - http://fortytwo.net/2010/10/RealTimeSemWeb.pdf
- Contact
  - josh@fortytwo.net
  - http://fortytwo.net