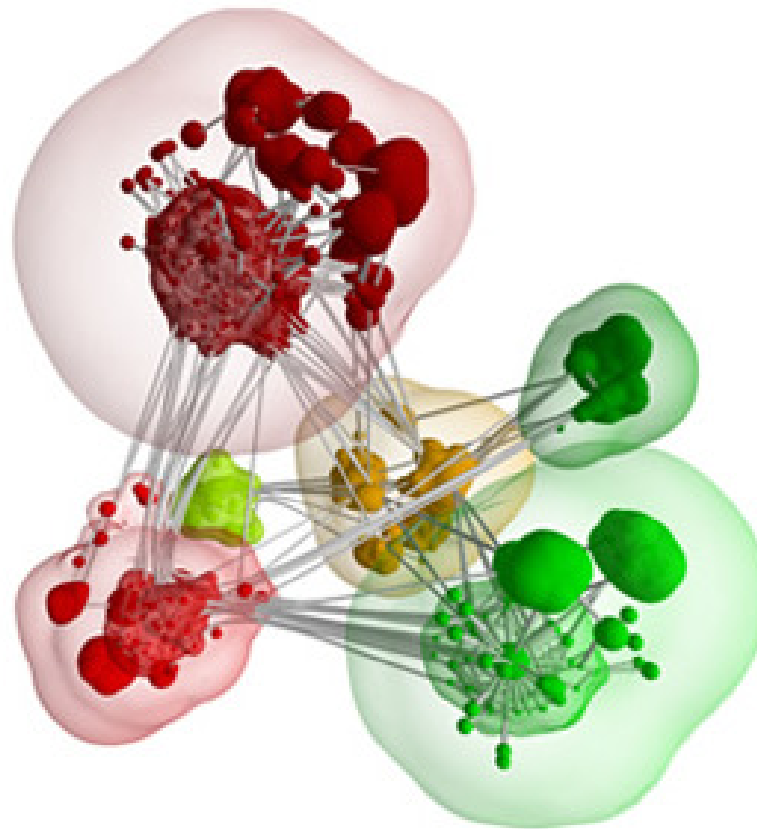




GeoSpatial capabilities and Moving Objects with RDF and AllegroGraph

Jans Aasman, Ph.D.
CEO Franz Inc
Ja@Franz.com





Trends in main stream IT

- Gartner group's 2008 list of the Top 10 Disruptive Technologies that will effect IT in the next five years
 - Multi core and hybrid processors
 - Virtualisation and fabric computing
 - Social networks and social software
 - Cloud computing and cloud/Web platforms
 - Web mashups
 - User Interface
 - Ubiquitous computing
 - Contextual computing
 - Augmented reality
 - Semantics



W3C Semantic Web Activity

The **Semantic Web** provides a common framework that allows **data** to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework ([RDF](#)). See also the separate [FAQ](#) for further information.

Introduction

The Semantic Web is a web of data. There is lots of data we all use every day, and it is not part of the web. I can see my bank statements on the web, and my photographs, and I can see my appointments in a calendar. But can I see my photos in a calendar to see what I was doing when I took them? Can I see bank statement lines in a calendar?

Why not? Because we don't have a web of data. Because data is controlled by applications, and each application keeps it to itself.

The Semantic Web is about two things. It is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing.

Further links

On this page: [Specifications](#) | [Publications](#) | [Presentations](#) | [Current Groups](#) | [Past Groups](#)

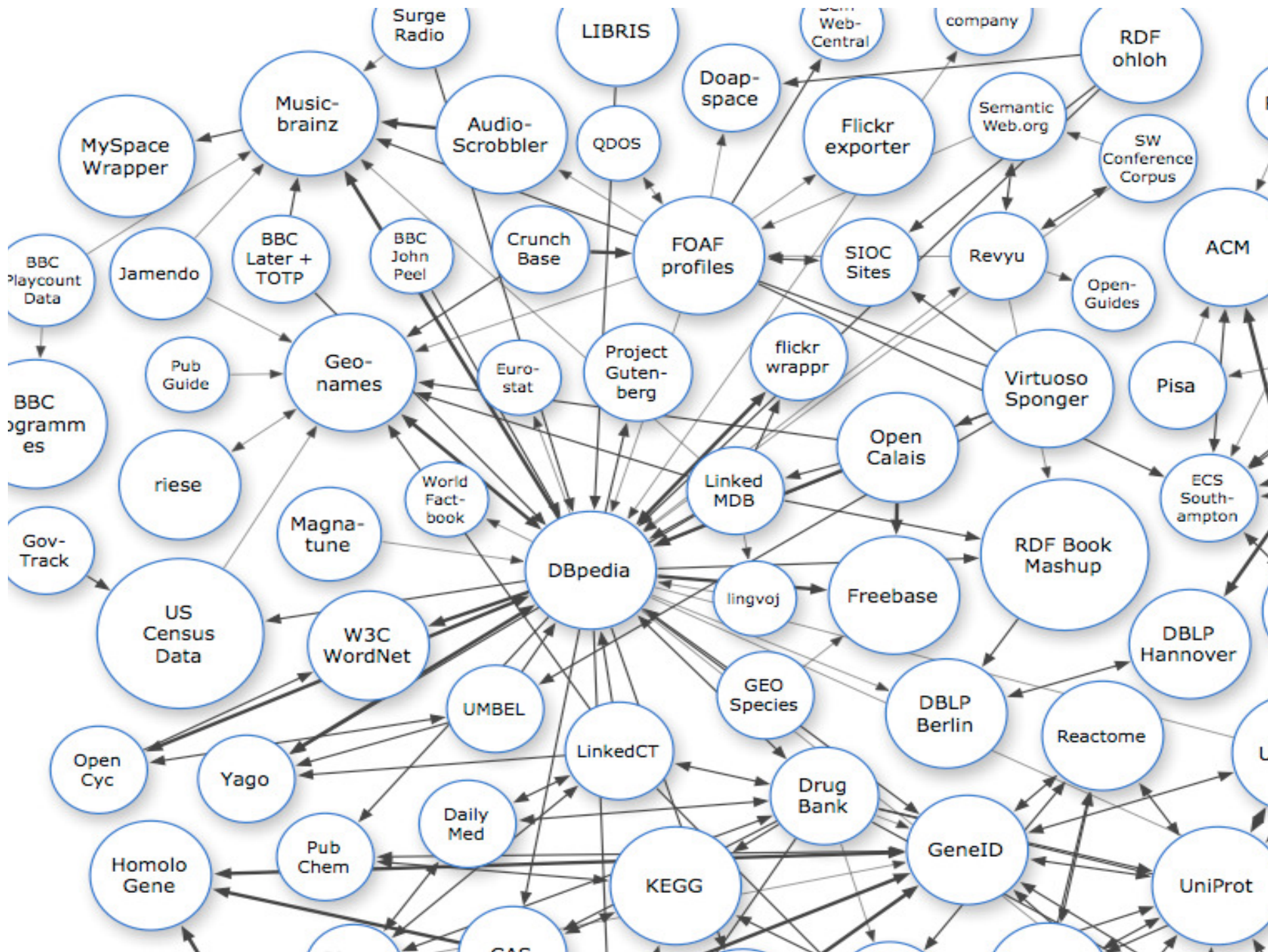
Latest news: [See the activity weblog](#)

Latest "layercake" diagram: in [SVG](#), [PNG](#), and [small sized \(300x315\) PNG](#) formats.

Active Groups: [Semantic Web Coordination Group](#) | [Rules Interchange Format Working Group](#) | [OWL Working Group](#) | [SPARQL Working Group](#) | [Semantic Web Deployment Working Group](#) | [POWDER Working Group](#) | [Health Care and Life Sciences Interest Group](#) | [Semantic Web Interest Group](#)

Past Groups: [Best Practices and Deployment](#)

- **The Semantic Web is about two things. It is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing.**
-





This talk

- Why RDF needs Geospatial capabilities, why the Geospatial community needs RDF.
- The basics of an RDF triple store
- Events everywhere
- One query language to combine
 - Geospatial, Temporal, Social Network Analysis
- Example: geospatial in AllegroGraph
- Example: a river network in GeoBC
- A future direction: moving objects
- Steve Haflich:
 - Research Topics in 3D indexing,
 - Query types on moving objects



Why the RDF community needs GeoTemporal reasoning capabilities

- Many of the Semantic Web projects are about
 - People and their relationships
 - Events that happen in space and time
- Reasoning about time and space needs to be an integral part of Semantic Web.
- The basic thing a geo interested web developer wants is
 - Find the distance between two places
 - Find everything with a particular description within x yards of place p

O'REILLY
where2.0[™]
 CONFERENCE

09 | MAY 19-21
 SAN JOSE, CALIFORNIA



Andrew Turner
 Mapufacture Inc



Dan Gillmor
 Arizona State University's Cronkite School...



Eric Horvitz
 Microsoft Research



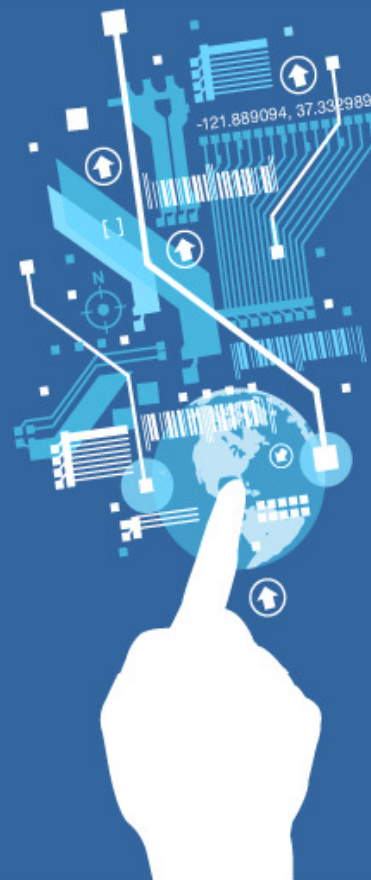
Alex Oliver
 Igloo Studios

NEWS & COVERAGE

The 2009 Where 2.0 Conference Has Concluded

NEWS & COVERAGE [See all >](#)

- News Release: O'Reilly Where 2.0 Conference Maps Geo's Borders: New Ideas and Practical Solutions Push Boundaries of Geolocation into New Territories
- cyrusfarivar.com: June 3: Cyrus on CBC's Spark
- duncandavidson.com: Jack Dangermond at **Where 2.0**



Where 2.0 Media

Videos, presentation files, photos, and more from Where 2.0, updated frequently. Participate in the **Attendee Directory** - connect with other Where 2.0 Conference attendees

TWITTER

(TAG: #WHERE20)

- Top Tweeps for [conference]: @mathewi @loic @loiclemeur @where20

Platinum Sponsor



Gold Sponsors



Silver Sponsors



Lanyard Sponsor





Why the geospatial community needs RDF

- RDF is about MetaData
- MetaData is usually very unstructured and sparse
- Objects that you describe are usually part of an object hierarchy
- You need ontologies to describe geospatial objects..
 - See how the national map project uses ontologies to solve the metadata problem

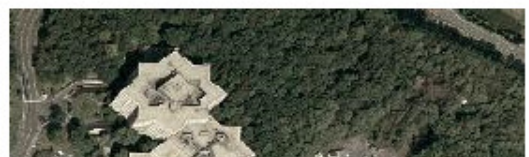
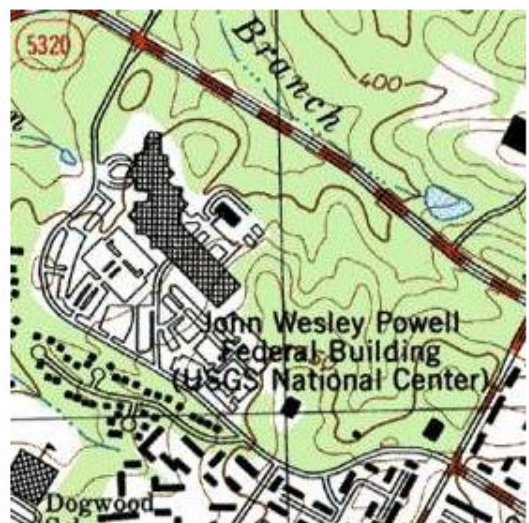


The National Map

- The National Map Home
- Products and Services
- Partnerships
- Reading Room
- News & Events
- The National Map Corps
- National Geospatial Program

The National Map

As one of the cornerstones of the U.S. Geological Survey's (USGS) [National Geospatial Program](#), *The National Map* is a collaborative effort among the USGS and other Federal, State, and local partners to improve and deliver topographic information for the Nation. It has many uses ranging from recreation to scientific analysis to emergency response. *The National Map* is easily accessible for display on the Web, as products and services, and as downloadable data. The geographic information available from *The National Map* includes orthoimagery (aerial photographs), elevation, geographic names, hydrography, boundaries, transportation, structures, and land cover. Other types of geographic information can be added within the viewer or brought in with *The National Map* data into a Geographic Information System to create specific types of maps or map views. *The National Map* is a significant contribution to the [National Spatial Data Infrastructure](#) (NSDI) and currently is being transformed to better serve the geospatial community by providing high quality, integrated geospatial data and improved products and services including new generation digital topographic maps.



In addition to being an important contribution to the NSDI, *The National Map* is foundational to implementation of the Department of the Interior (DOI) [Geospatial Modernization Blueprint](#) and meeting the DOI mission to protect America's treasures for future generations, provide access to our nation's natural and cultural heritage, offer recreation opportunities,



Graphs, triples, triple-store?

```
createTripleStore("seminar.db" )
```

```
addTriple (Person1 first-name Steve)
```

```
addTriple (Person1 isa Organizer)
```

```
addTriple (Person1 age 52)
```

```
addTriple (Person2 first-name Jans)
```

```
addTriple (Person2 isa Psychologist)
```

```
addTriple (Person2 age 50)
```

```
addTriple (Person3 first-name Craig)
```

```
addTriple (Person3 isa SalesPerson)
```

```
addTriple (Person3 age 32)
```

```
addTriple (Person1 colleague-of Person2)
```

```
addTriple (Person1 colleague-of Person3)
```

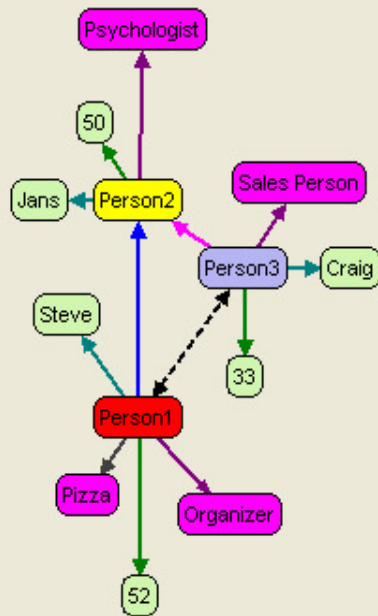
```
addTriple (Person1 likes Pizza)
```

```
addTriple (Person3 neighbour-of Person1)
```

```
addTriple (Person3 neighbour-of Person2)
```


- Age
- Colleague-of
- First-name
- Likes
- Neighbour-of
- Type
- Multiple Predicates

- Organizer
- Psychologist
- Sales Person
- Literal
- No Type





And deal with events

```
addTriple (Event100 type Meeting)
```

```
addTriple (Event100 actor Person3)
```

```
addTriple (Event100 actor Person2)
```

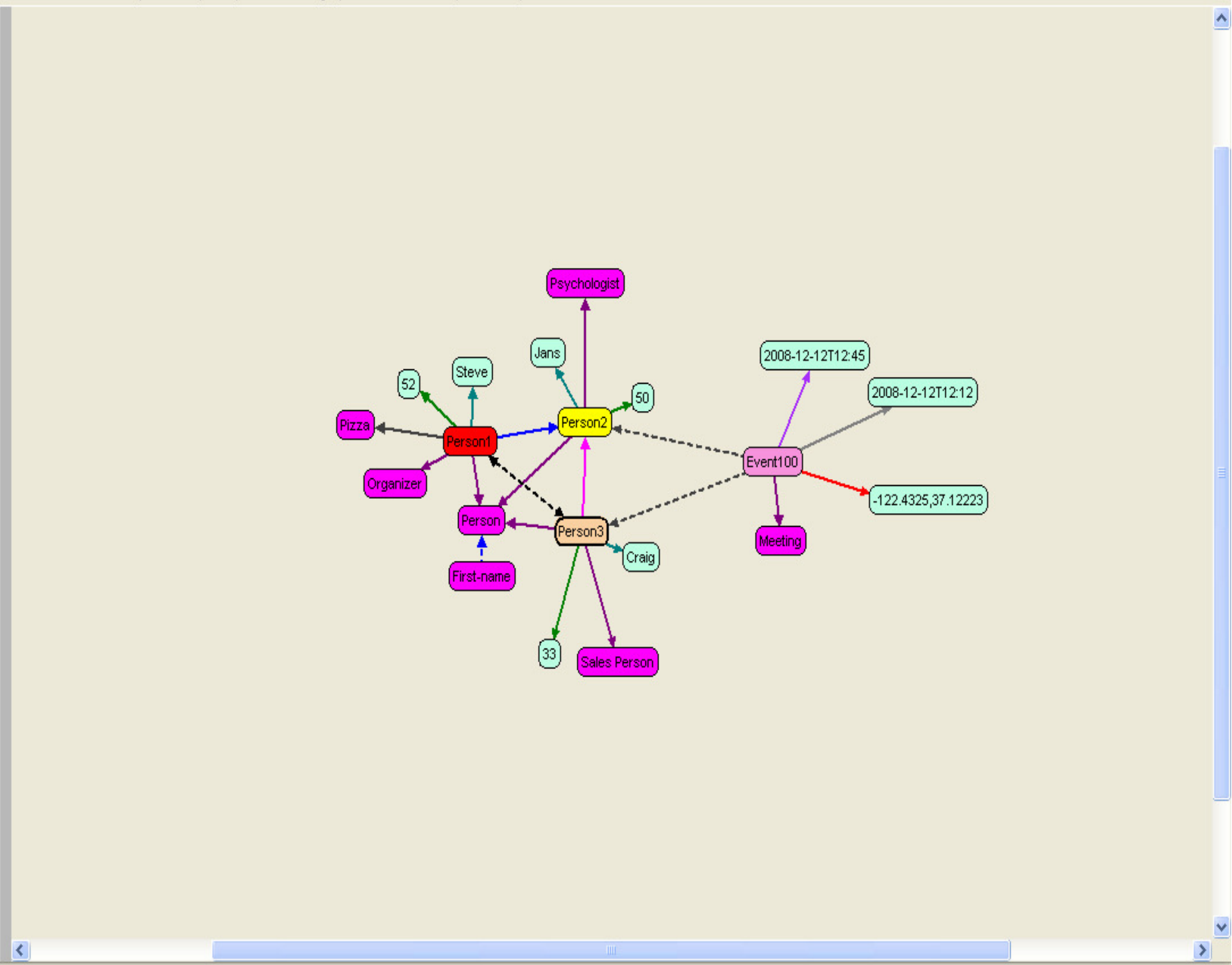
```
addTriple (Event100 start 2008-12-12T12:12)
```

```
addTriple (Event100 end 2008-12-12T12:45)
```

```
addTriple (Event100 is-at -122.4325,37.12223)
```

- Actor
- Age
- Colleague-of
- Domain
- End
- First-name
- Is-at
- Likes
- Neighbour-of
- Start
- Type
- Multiple Predicates

- Meeting
- Organizer
- Psychologist
- Sales Person
- Literal
- No Type





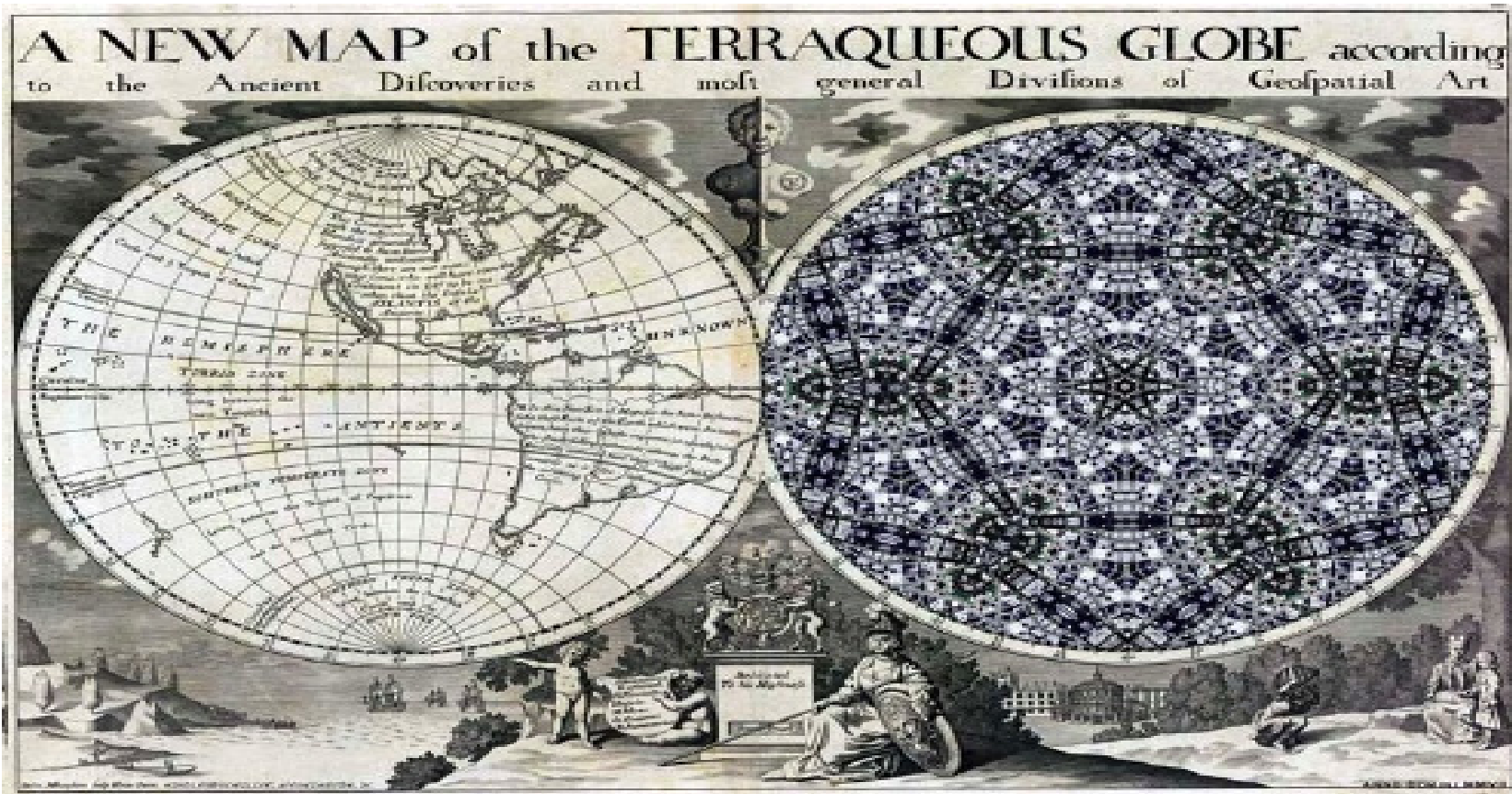
Integrated in select language.

Find a meetings that happened in July within 5 miles of Berkeley that was attended by the most important person in Jans' friends and friends of friends.

```
(select (?x)
  (ego-group !person:jans knows ?group 2)
  (actor-centrality-members ?group knows ?x ?num)
  (q ?event !fr:actor ?x)
  (qs ?event !rdf:type !fr:Meeting)
  (interval-during ?event "2009-07-01" "2009-07-25")
  (geo-box-around !geoname:Berkeley ?event 5 miles)
!)
```



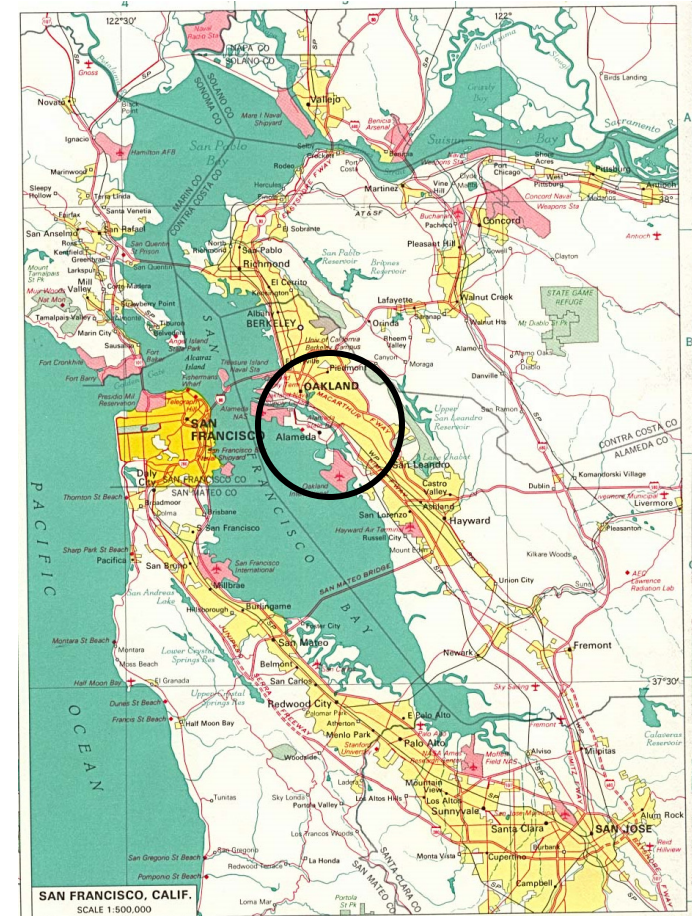

Geospatial Reasoning





GeoSpatial

- Make the following super efficient
 - Where did something happen?
 - How far was event1 from event2?
 - Find all the events that occurred in a bounding box or radius of M miles?
 - Do these two shapes overlap?
 - Find all the objects in the intersection of two shapes
- On a very large scale
 - when things don't fit in memory
 - millions of events and polygons





Sample Geospatial Primitives

(geo-bounding-box ?x +minlat +maxlat +minlon +maxlon)

(geo-box-around +x ?y +miles)

(geo-distance +x +y ?dist)

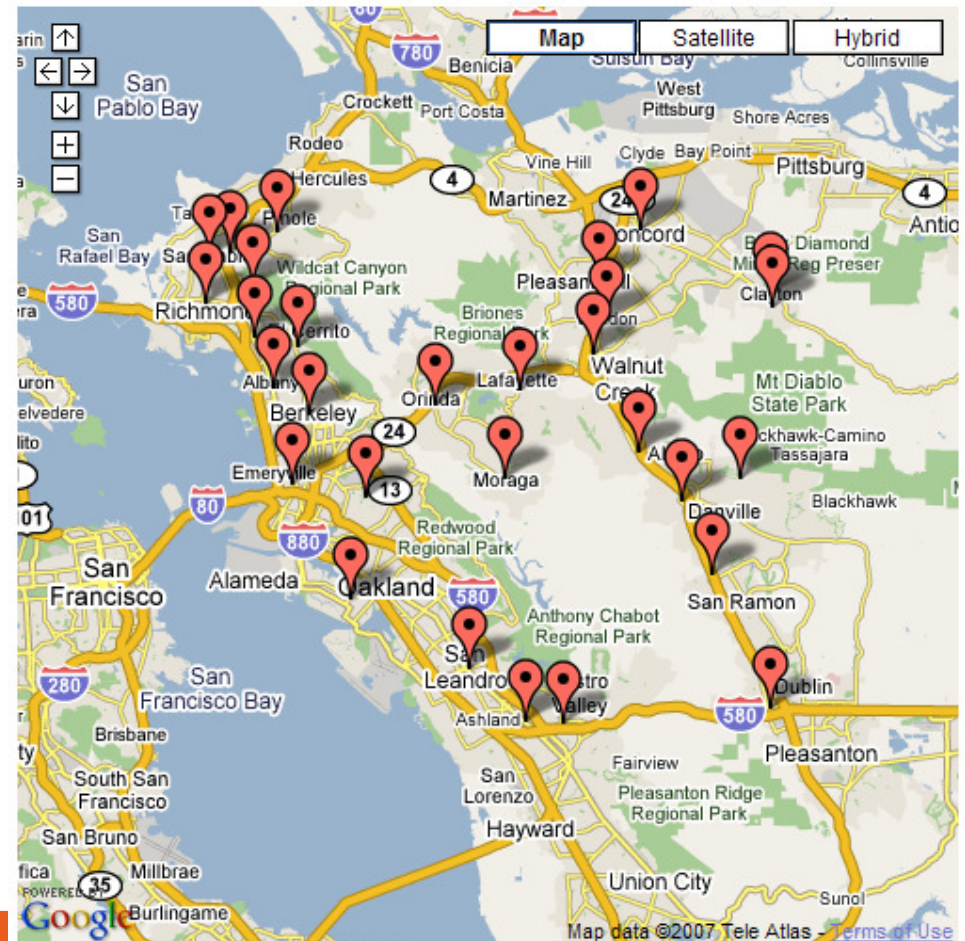
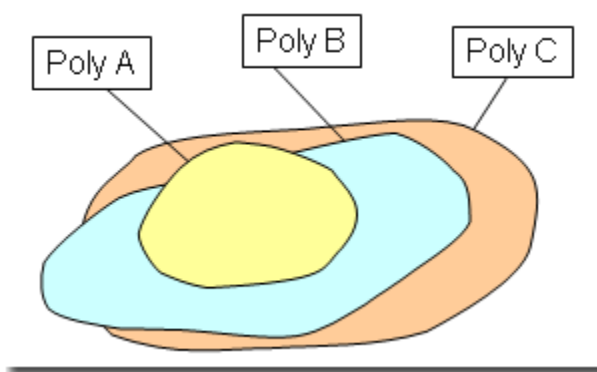
(geo-radius-around +x ?y +miles)

(polygon-in ?p1 ?p2)

(polygon-touch ?p1 ?p2)

(polygon-overlap ?p1 ?p2)

Etc.





Now with SPARQL Support [1]

```
PREFIX fr:
    <http://franz.com/ns/allegrograph/3.0/geospatial/>
PREFIX geo:      <http://www.geonames.org/ontology#>
PREFIX country: <http://www.geonames.org/Countries#>
```

```
SELECT ?placename ?population WHERE {
  GEO OBJECT
  HAVERSINE ( ?londonpos, 50 MILES ) {
    ?place fr:pos ?pos ;
    geo:name ?placename ;
    geo:population ?population ;
    geo:countryCode ?cc
  }
  WHERE {
    # Select London, UK.
    ?london geo:name 'London' ;
    geo:countryCode 'GB' ;
    fr:pos ?londonpos .
  }
  FILTER (?population > 25000)
```




Benchmarking on GeoNames.rdf

Using GeoNames

- coordinates for 6,445,201 places
- 109,568,417 RDF triples

AllegroGraph requires only 95 msec real time to return the 502 entries within a 3 mile radius of the Franz Inc offices.



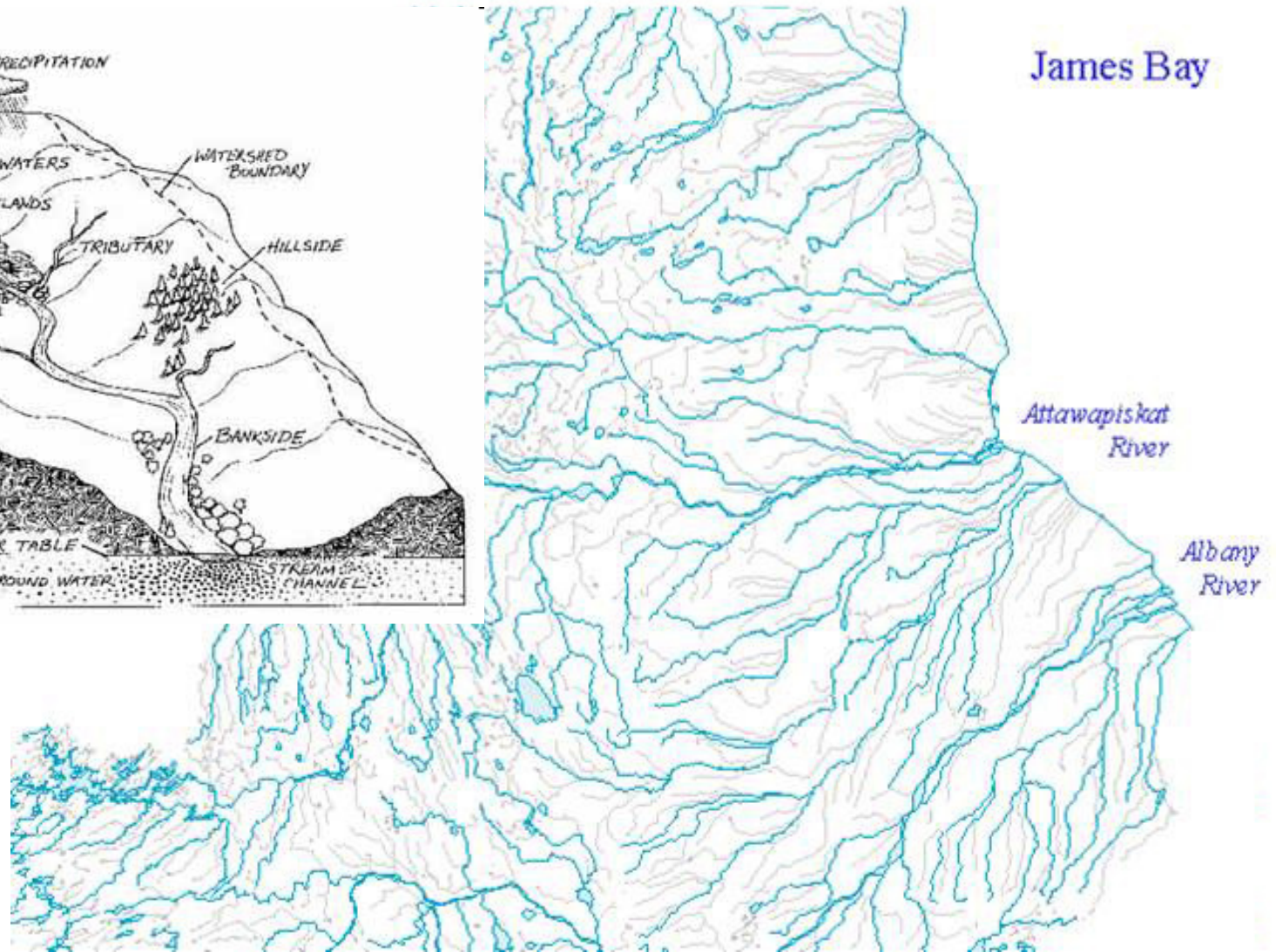
Example 1: dealing with a river network using KML and Google Earth

- Working with data from GeoBC





The graph in a River Network





Regular data with a graph

S1 type stream-segment

S1 upstream S2

S1 upstream S3

S1 left-drainage D1

S1 right-drainage D2

S1 isAt (-121.2, 12.1)

Given the polluted segment S1 find all the upstream segments within 50 miles of City1200

Given the polluted drainage D1 find all the schools in the rectangle $\langle x1, y1, x2, y2 \rangle$ that might be influenced



Queries

- Graph search
 - find all upstreams or downstreams of a segment
- Graph search + bounding box
 - Given the polluted drainage D1 find all the schools in the rectangle $\langle x1, y1, x2, y2 \rangle$ that might be influenced

Search

Fly To Find Businesses Directions

Fly to e.g., New York, NY

Search input field with a magnifying glass icon

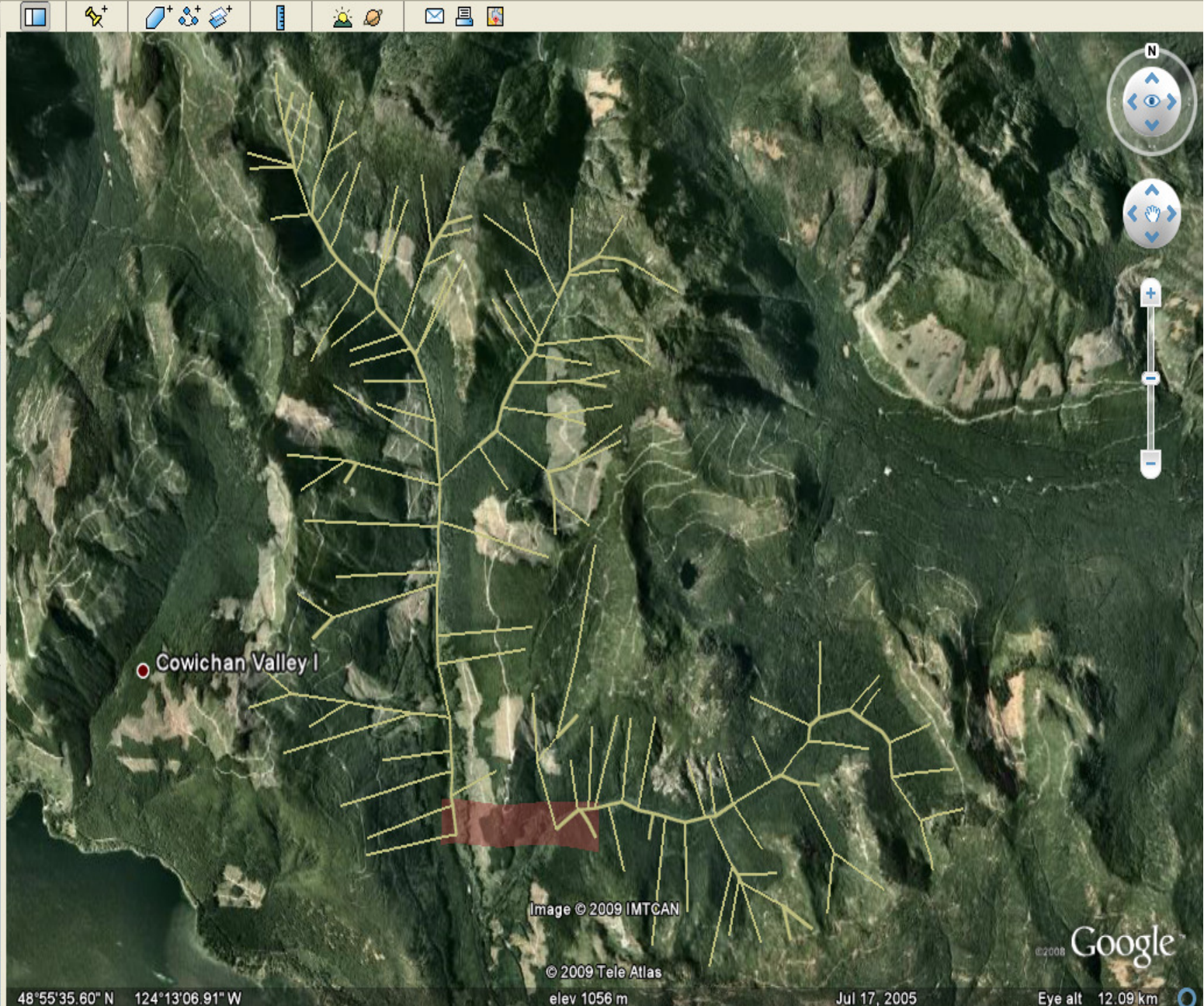
Places

Add Content

- My Places
 - Sightseeing
 - Select this folder and click on the 'Play' button below, to start the tour.
 - ~GE34CD.kmz
 - ~GEB0DB.kmz
 - Drainage for node X-124D25017...**
 - Drainage for node lbc:710281837
 - lbc:710281837
 - Temporary Places

Layers

- Primary Database
- Geographic Web
- Roads
- 3D Buildings
- Street View
- Borders and Labels**
- Traffic
- Weather
- Gallery
- Global Awareness
- Places of Interest
- More



48°55'35.60" N 124°13'06.91" W

elev 1056 m

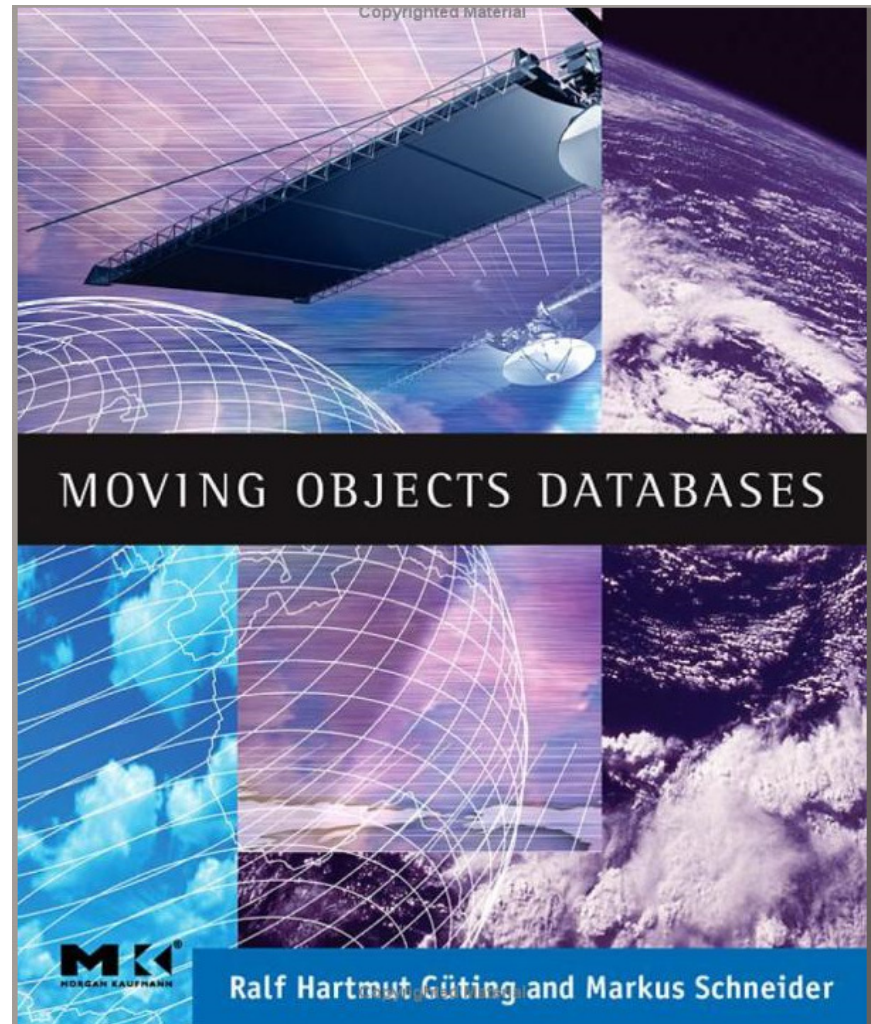
Jul 17, 2005

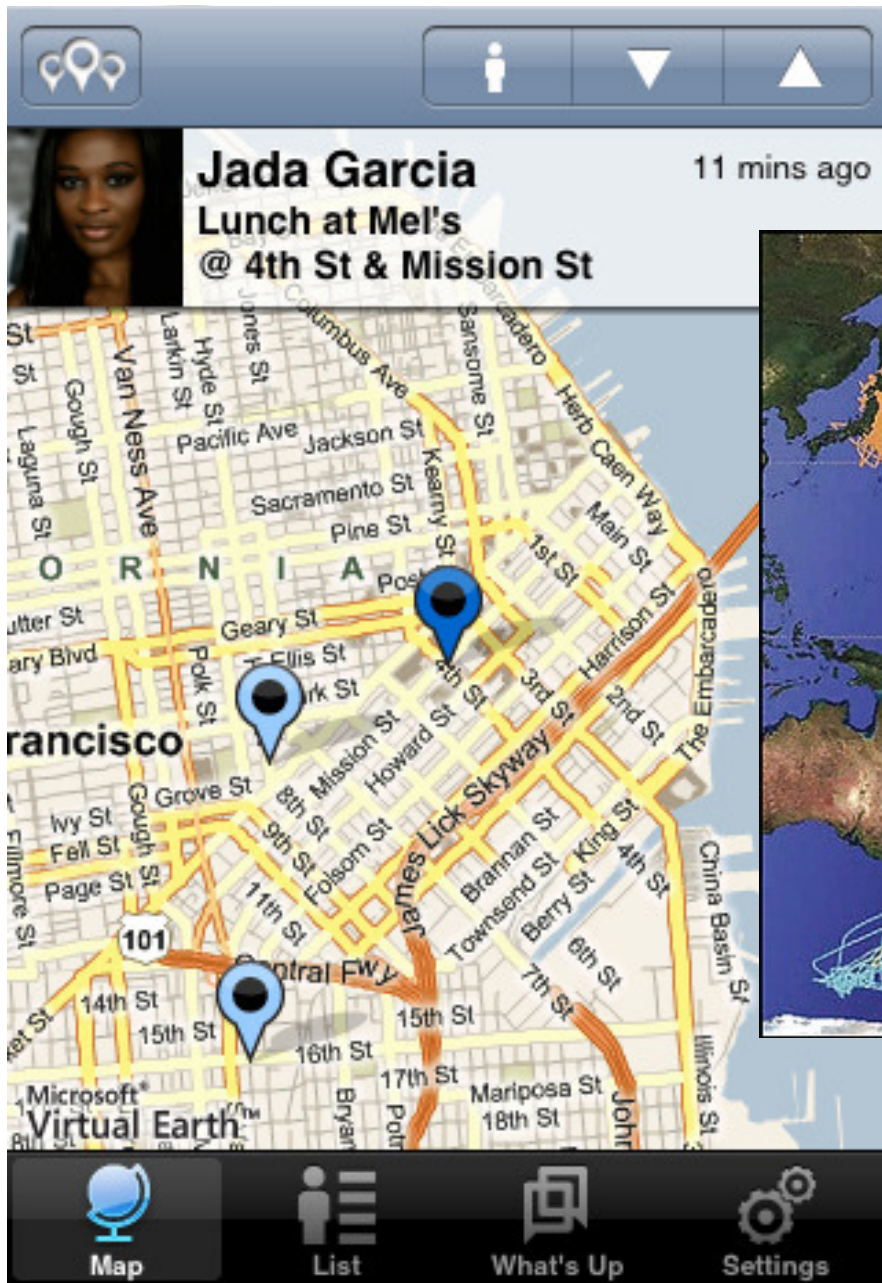
Eye alt 12.09 km



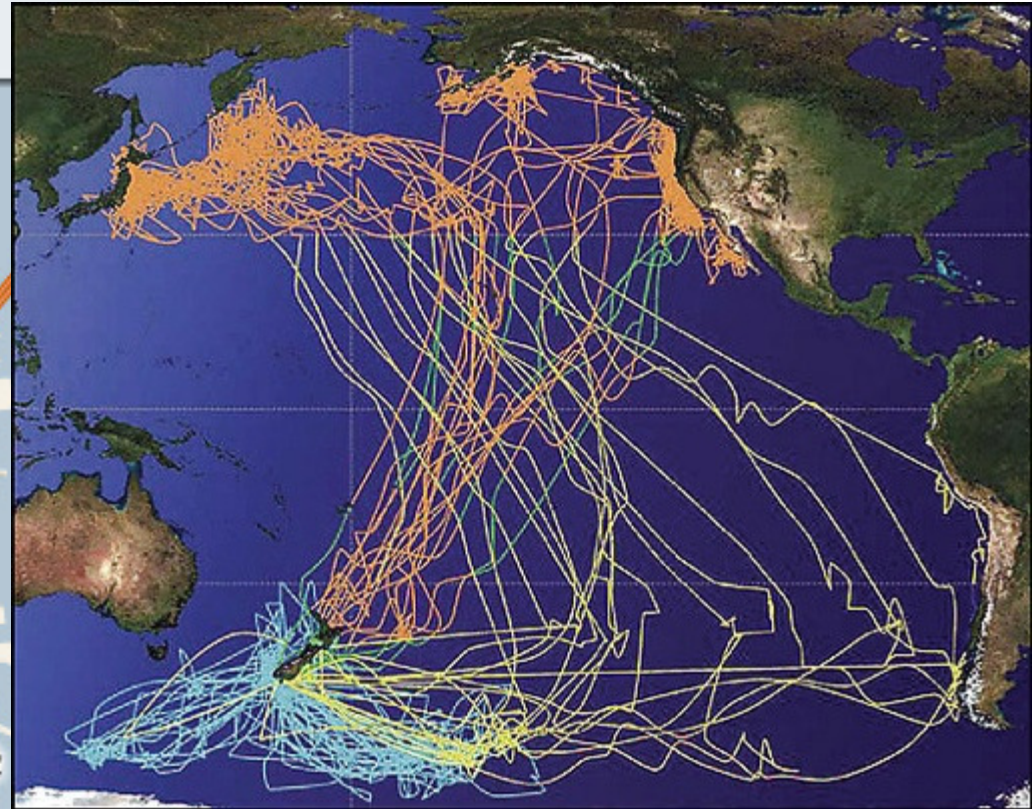
Moving Objects

- Track data for animals in biodiversity projects
- Track ships near the coast of Africa
- Fleets of trucks and swarms of airplanes
- Track data for soldiers
- New Telco Service:
 - Loopt and friends





Loopt

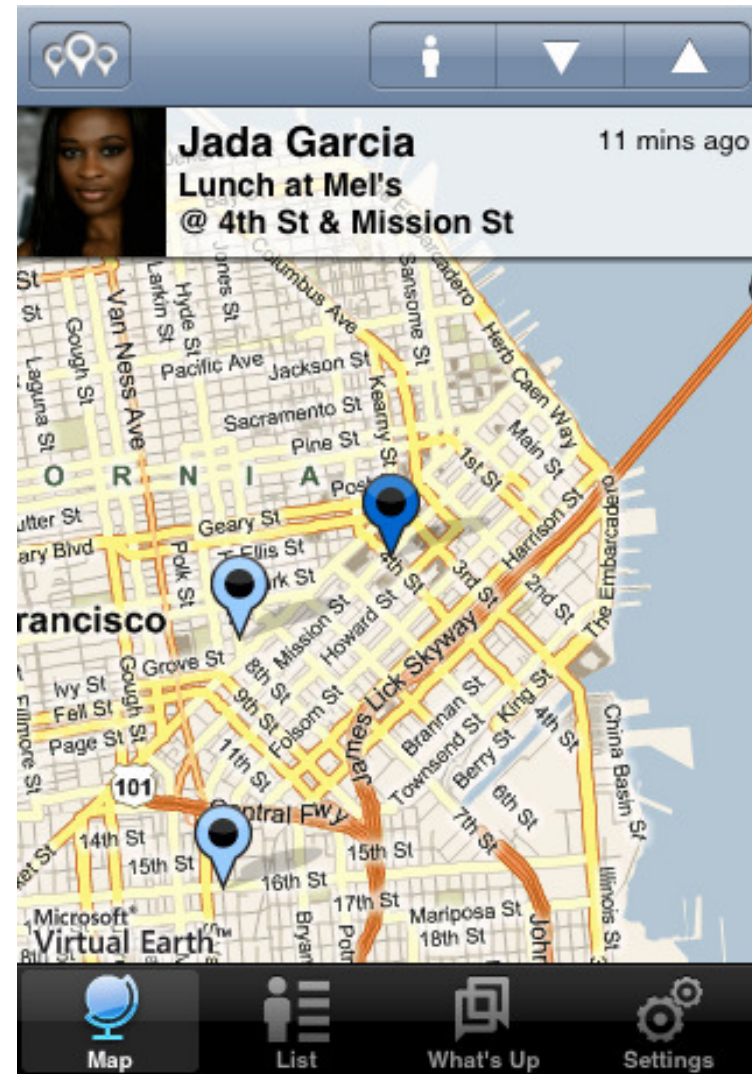


Tagged birds



The challenge

- Track a million objects
- At a resolution of once per minute or a change in x meters
- In real time
 - Use geotemporal (3D) indexing techniques from the game industry
 - Do complex event handling for current situation
- And historically
 - Use geotemporal (3D) indexing
- And do predictions





Questions we need to answer

- Do we have 5 soldiers that have been active less than 8 hours with training x and equipment y within a mile distance of each other
- Find another truck that can pick up package X at location Y so that I can pick up package A at location B so that we both will arrive at P before time T .



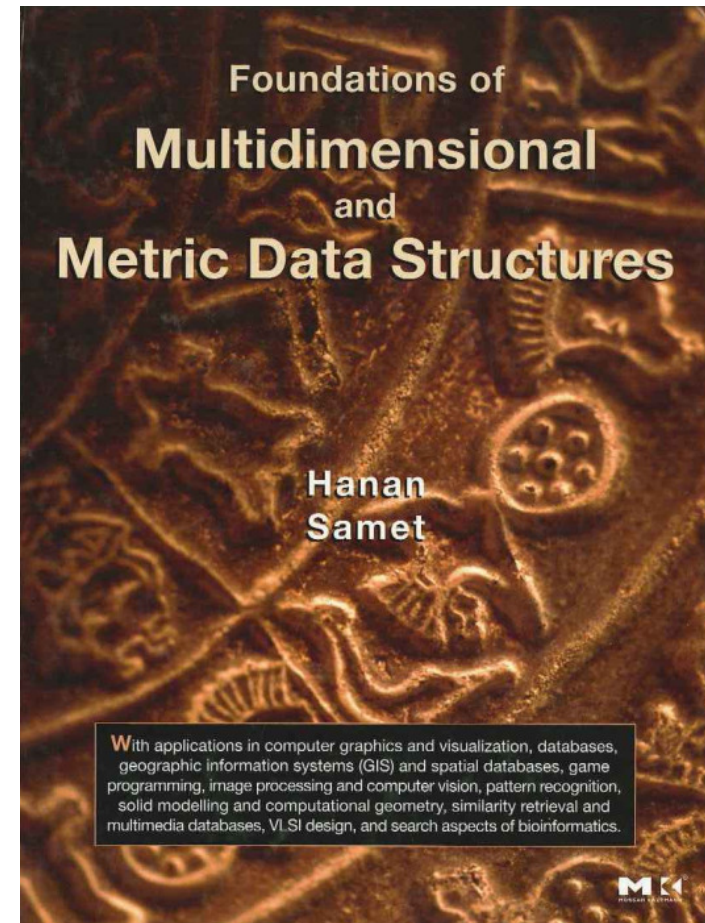
So far implemented – real time

- An object model for representing active objects
- In memory 3D indexing using a grid approach
- Simple rules to trigger actions based on in-memory objects
- Simple rules to aggregate real time data and store it in the triple store.



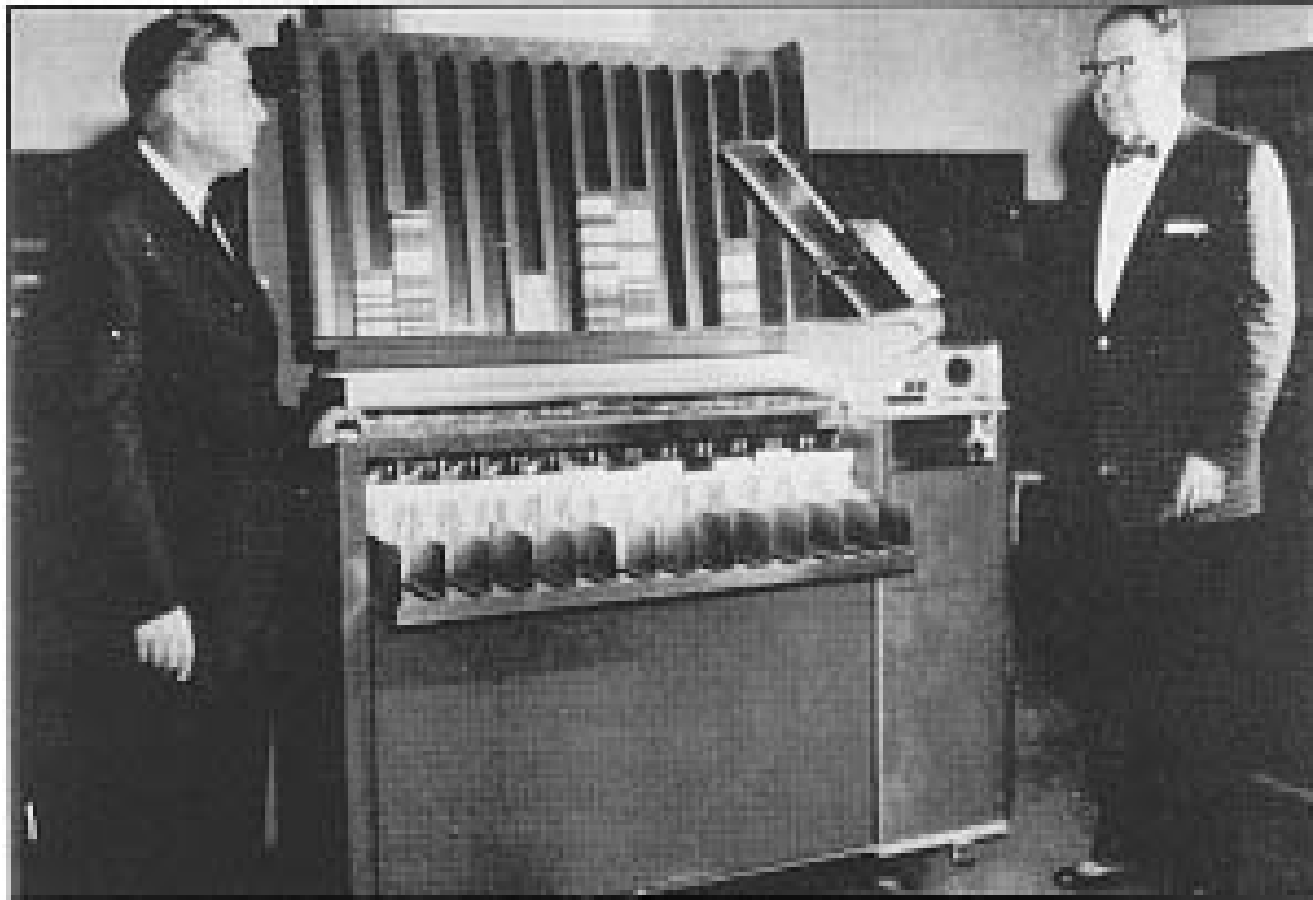
So far implemented - historic

- 2D indexing
 - (in the regular Agraph 3.2)
- 3D indexing
 - (in AG 4.0 or close thereafter)





**Happiness is sorting things in
 $n \log n$ time.**





“Happiness is finding things that are linear.”

- Computer main memory and disk are linearly addressable vectors.
- It is well known that a vector of length n can be
 - sorted in $O(n \log n)$ time.
 - searched in $O(\log n)$ time.
- AllegroGraph is designed to exploit machine speed, despite scaling requirements, by keeping everything linear.
- Maintains multiple sorted indexes (e.g. SPOG, POSG, GOSP).
- By selecting the proper index, triples variously related to others can be retrieved from a local region of that index.



Happiness is finding things linearly.

- If I want to retrieve everything about `<http://franz.com/employees#Jans>`, all triples with this Subject are sorted together in the SPOG index.
- All triples with Jans as the Object are together in the OSPG index.
- All triples with a particular Predicate e.g. `<http://franz.com/employees#isSupervisorOf>` are grouped together, sorted secondarily on Object, in the POSG index.
- And so on...



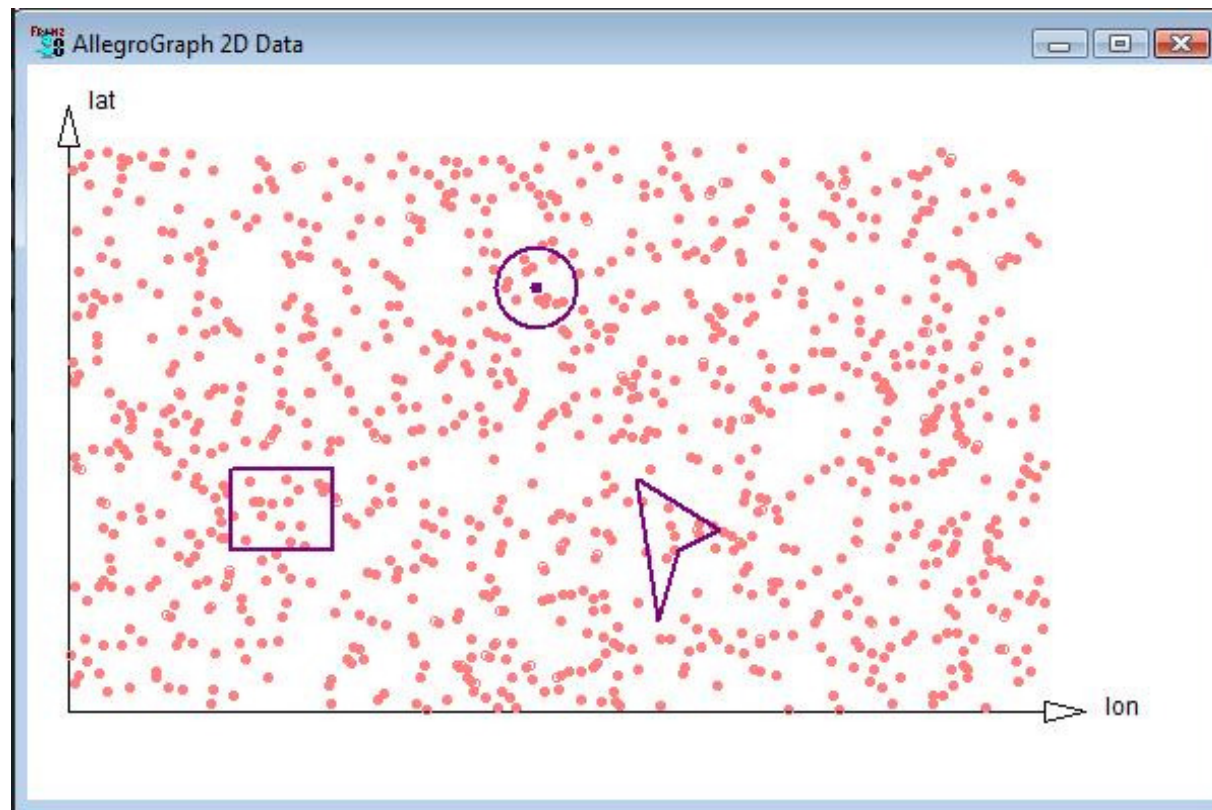
But 2-D and 3-D Aren't Linear!

- Age, date and/or time, currency, phone numbers, stock prices, license-plate numbers, and barometric pressure are all linearly orderable.
- Cartesian and spherical (e.g. geospatial) coordinates in two or higher dimensions are not immediately orderable and sortable. How to integrate these into the AllegroGraph model?



How to find linear happiness in higher dimensions???

- The important problem is proximity search.
- We want to optimize speed retrieving all triples with coordinates in a certain locality.





The Wrong Way

- Data in two dimensions could be sorted on the two separate dimensions in the obvious way, either first on Y/latitude or X/longitude, or the reverse. (*)
- But this causes search time over a locality to increase linearly with the the size of the data set.

() This is a gross misfeature in some RDF schema for geospatial data.*

Latitude and longitude should not be carried by separate triples:

<http://ex.com#HQ> <geo:isAtLat> "37.45"^^<geo:latitude> .

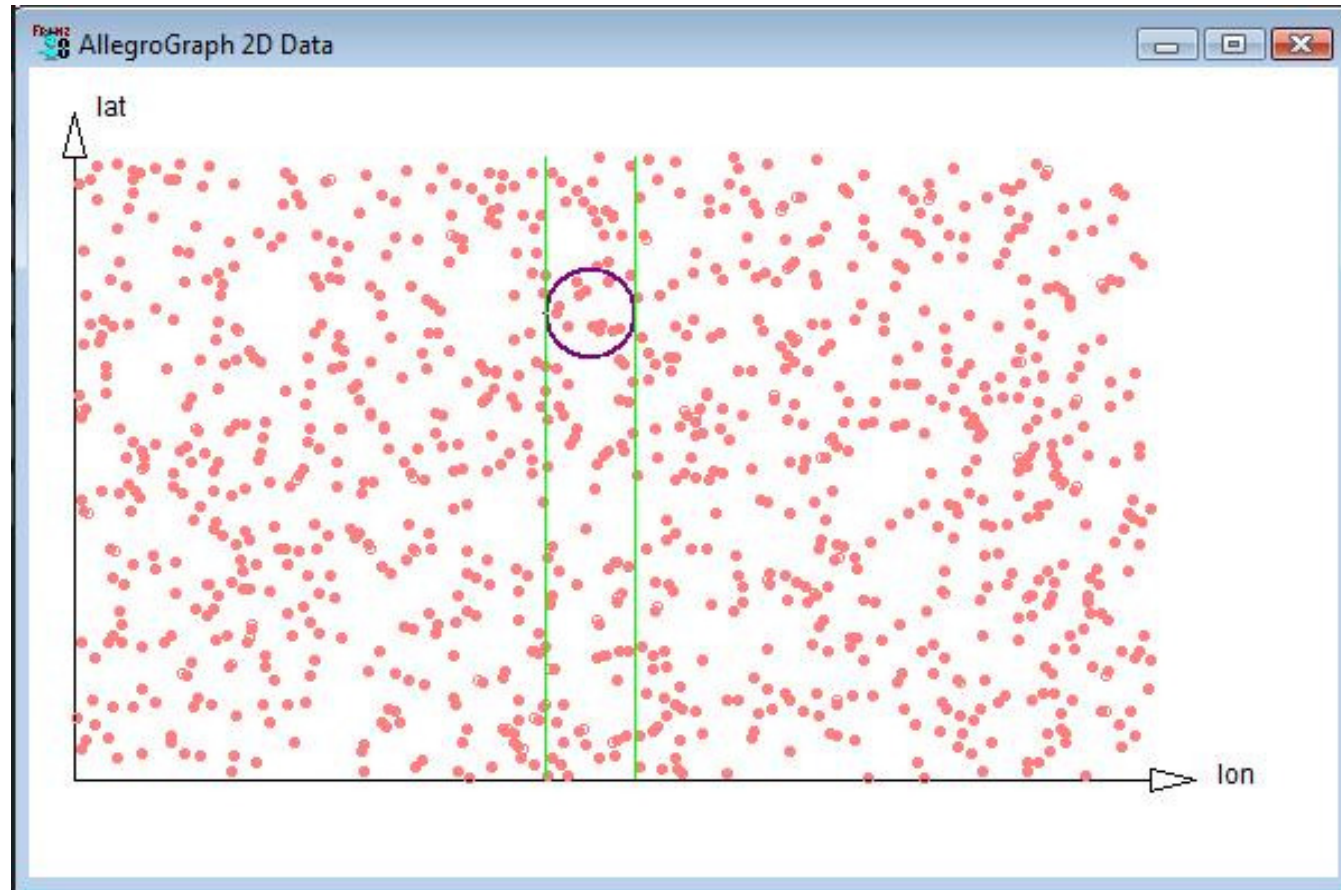
<http://ex.com#HQ> <geo:isAtLon> "-123.27"^^<geo:longitude> .

ISO6709, for example, would be better:

<http://ex.com#HQ> <geo:isAt> "+37.45-123.27"^^<geo:ISO6709> .



Sorting Lat and Lon Separately



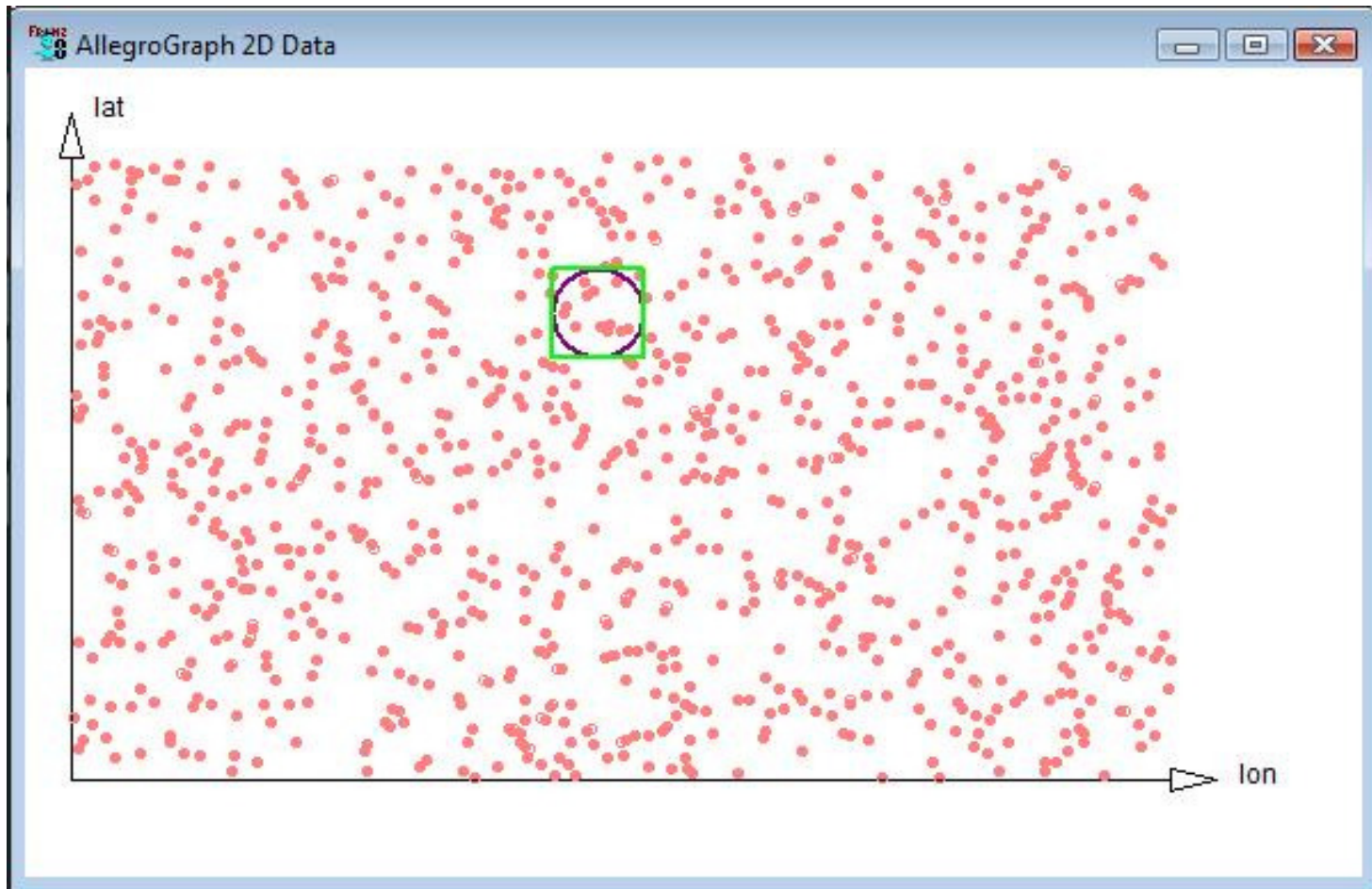


How?

- R-trees and numerous other schemes support very efficient search of localities.
- But if all the data won't fit in memory, performance can be unpredictable.
- And there is no obviously efficient way to reconcile 2-D and higher R-trees with the AllegroGraph linear indexing.
- But suppose we knew a little more about how we will use our data, specifically:
- The approximate size of typical regions to be searched.



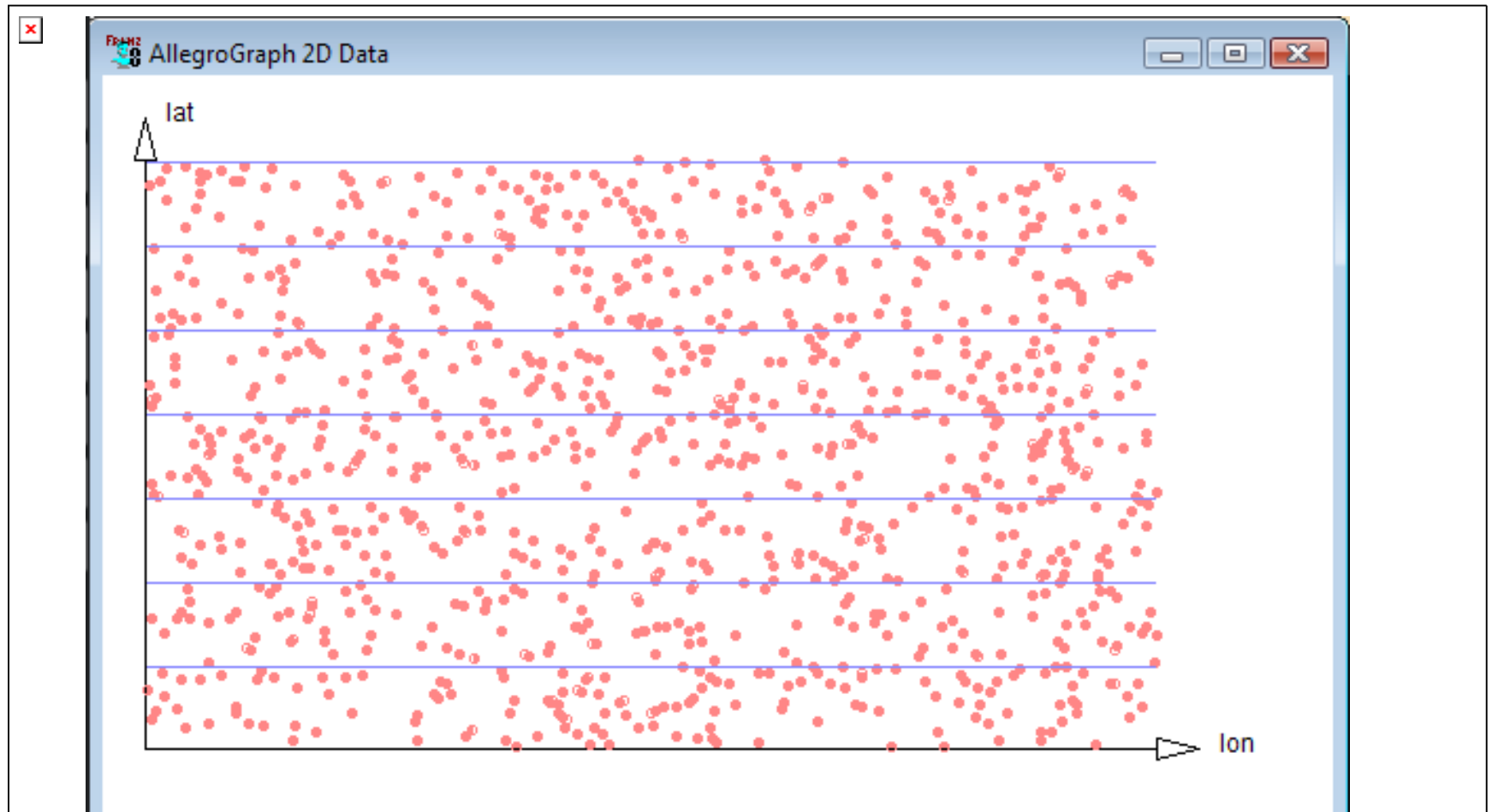
We'd Rather Search Just the Locality



- But how?



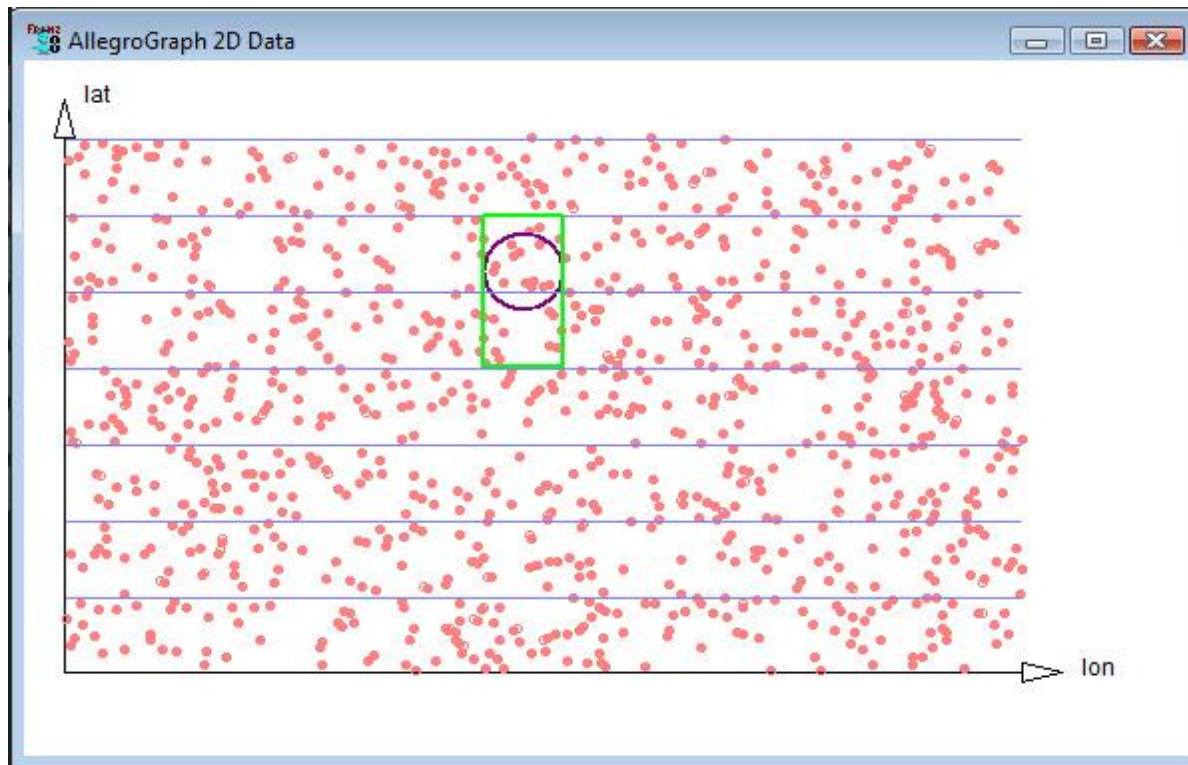
Eureka! Think In Strips





If Strip Width is the Same As Search Radius

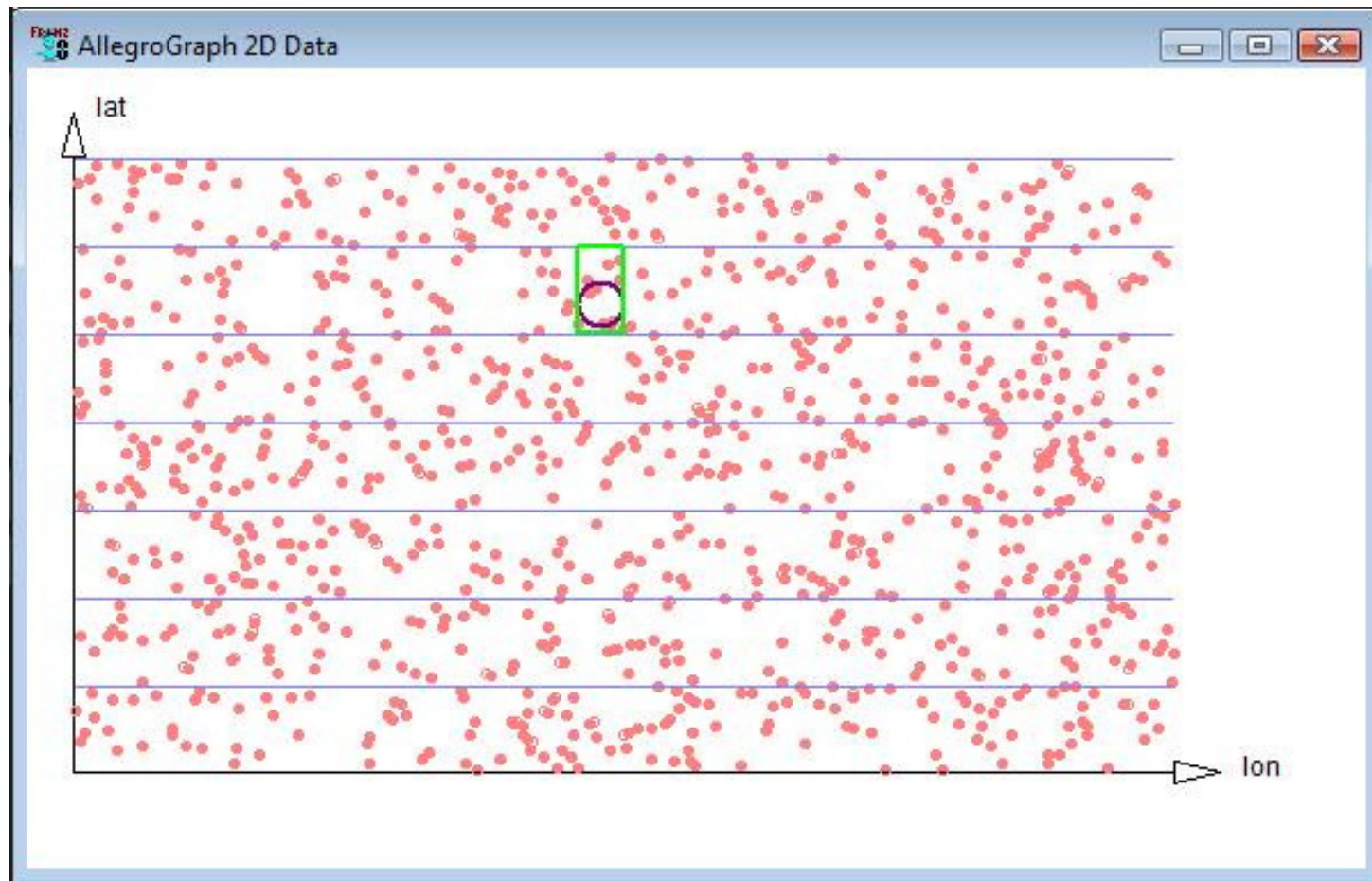
- Need search only short linear regions of two strips.





If Search Radius is Somewhat Smaller Than Strip Width

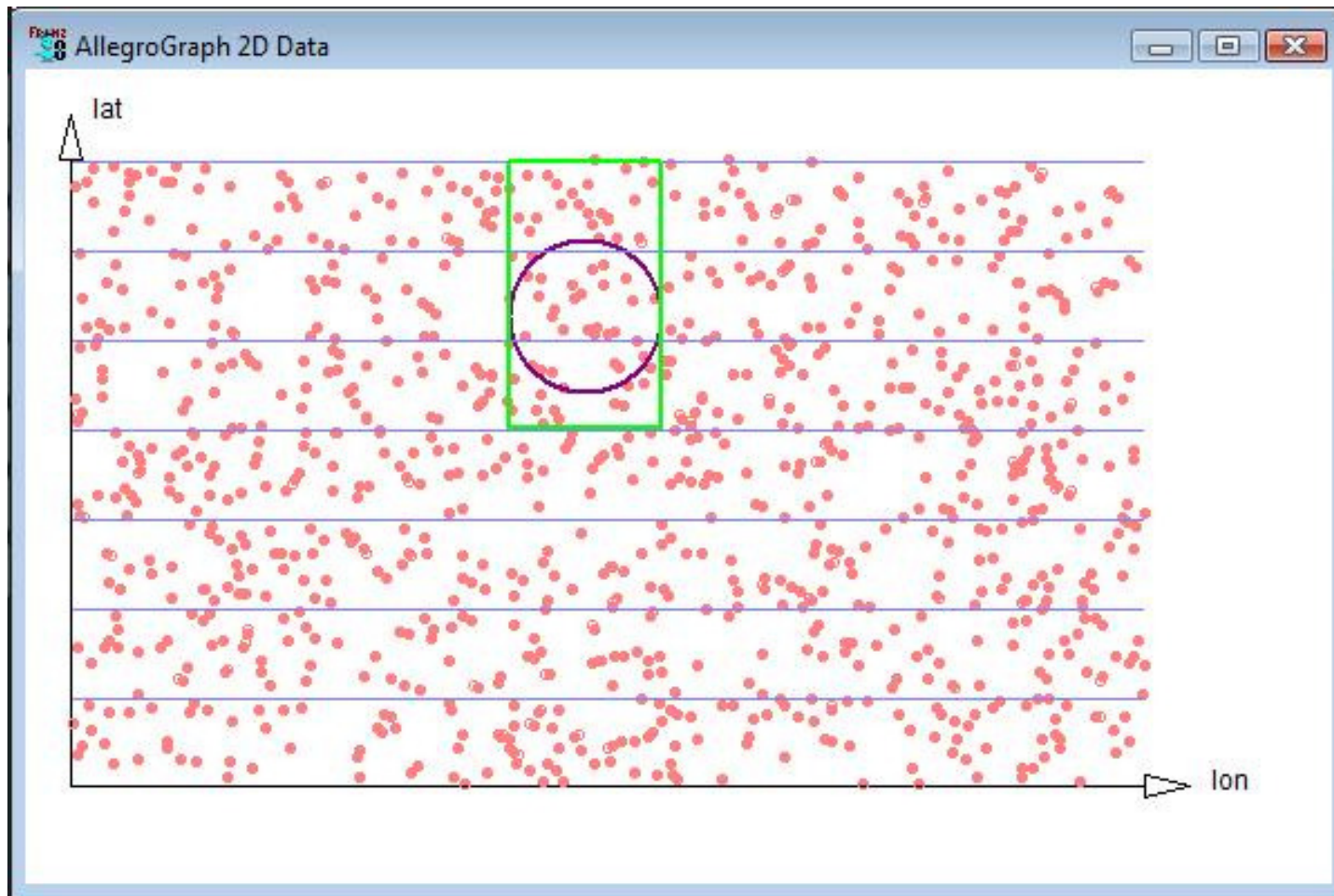
- Need search only short linear regions of one or two strips.





If Search Radius is Somewhat Larger Than Strip Width

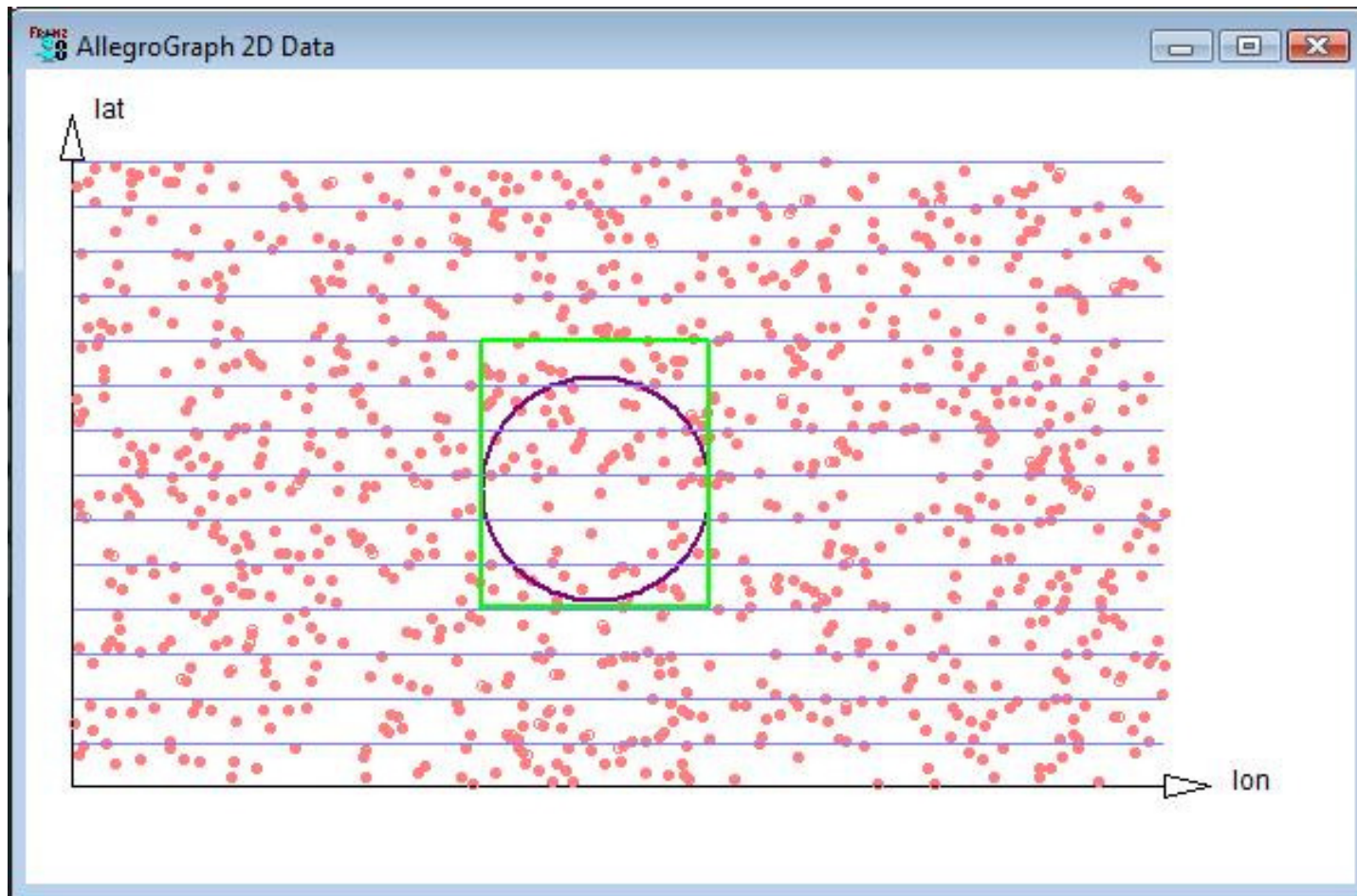
- Need search short linear regions of a few more strips.





If Search Radius is Much Larger Than Strip Width

- Need search longer linear regions of even more strips.



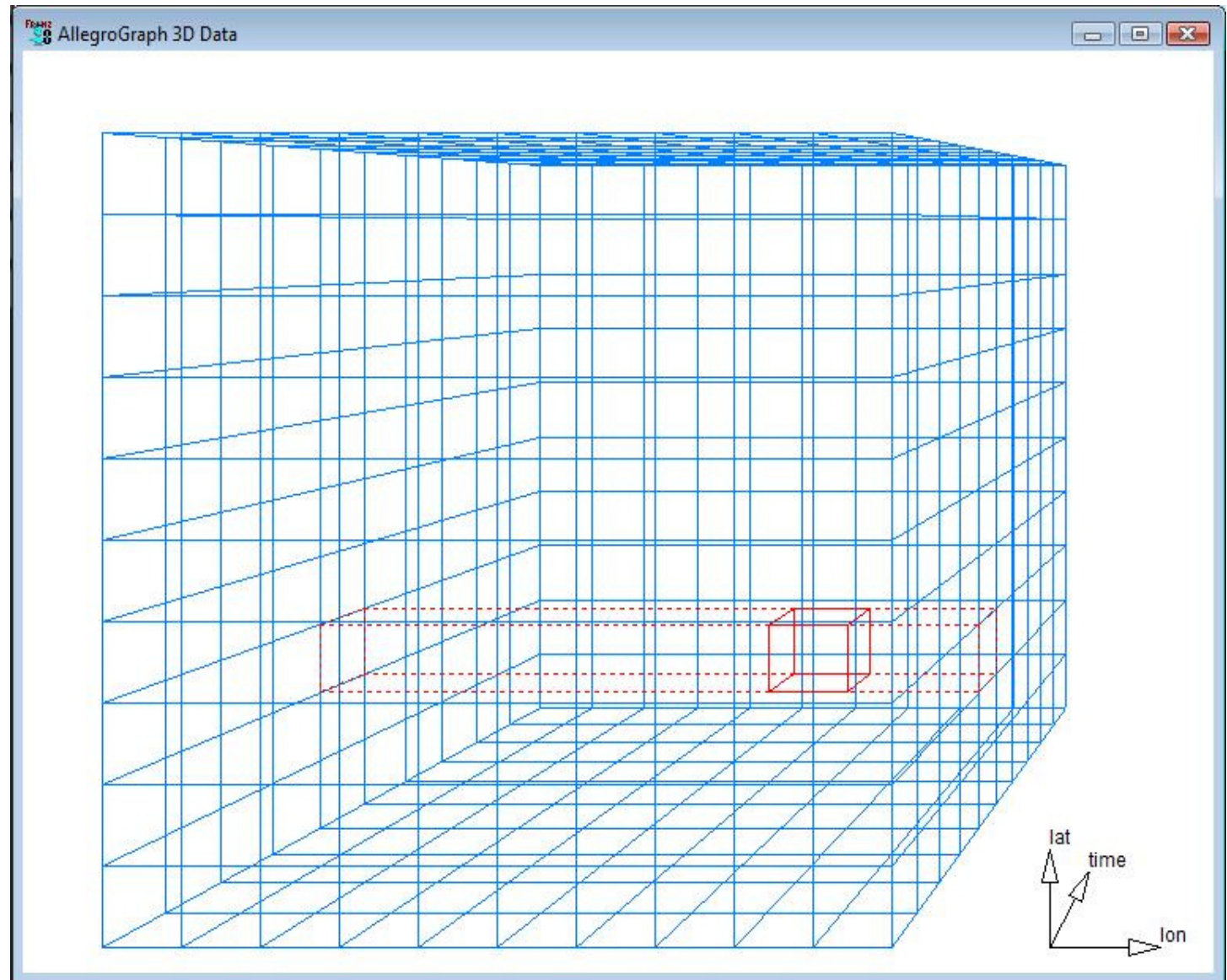


Summary of 2-D

- It is inconvenient to need to specify strip width in advance.
- But performance is still reasonable even with an order of magnitude error.
- In addition, if extremely different strip sizes are needed, the data can be stored twice with different strip sizes.
- The two coordinates can be anything:
 - pressure and temperature
 - distance and time.
- This last possibility suggests extension to 3-D and beyond, particularly MOB's in latitude/longitude/time.



Eureka In 3D! Think In Prisms



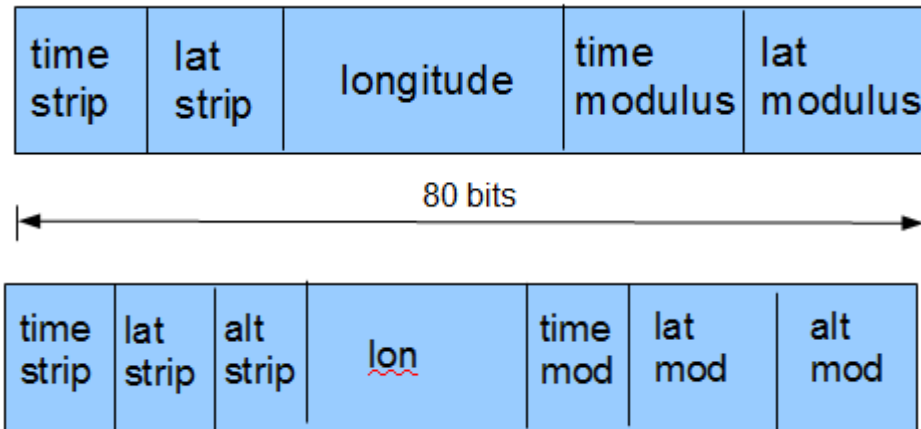


Extending to 3-D and MOBs

- Strip representation can be extended to 3-D and beyond using prisms.
- Again, the dimensions can be anything:
 - $[x,y,z]$, $[\text{lat},\text{lon},\text{alt}]$, $[\text{lat},\text{lon},\text{time}]$, $[\text{lat},\text{lon},\text{alt},\text{time}]$
- $[\text{lat},\text{lon},\text{time}]$ is very important, corresponding to Moving Objects on the Earth's Surface.



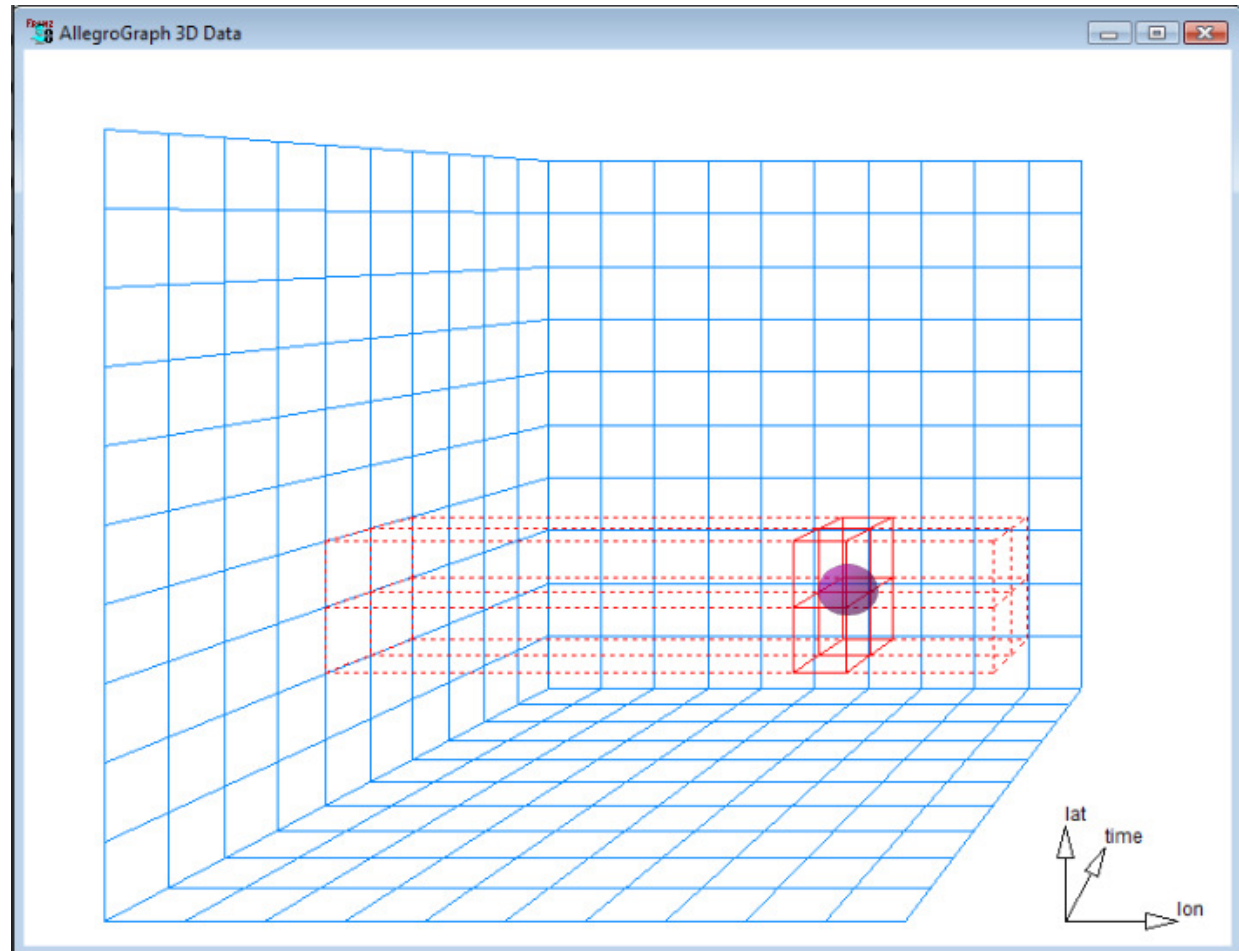
Layout of Latitude and Longitude



- 3 and higher dimensions can be encoded and linearized similarly.
- Map each ordinate into a nonnegative integer, then split the first $n-1$ coords into strip and modulus.
- Number of strips is the ordinate range divided by strip width.
- Modulus is remainder is ordinate divided by number of strips.
- The 80 bits can be allocated according to resolution needs.
- But the available resolution is less for higher dimensions.



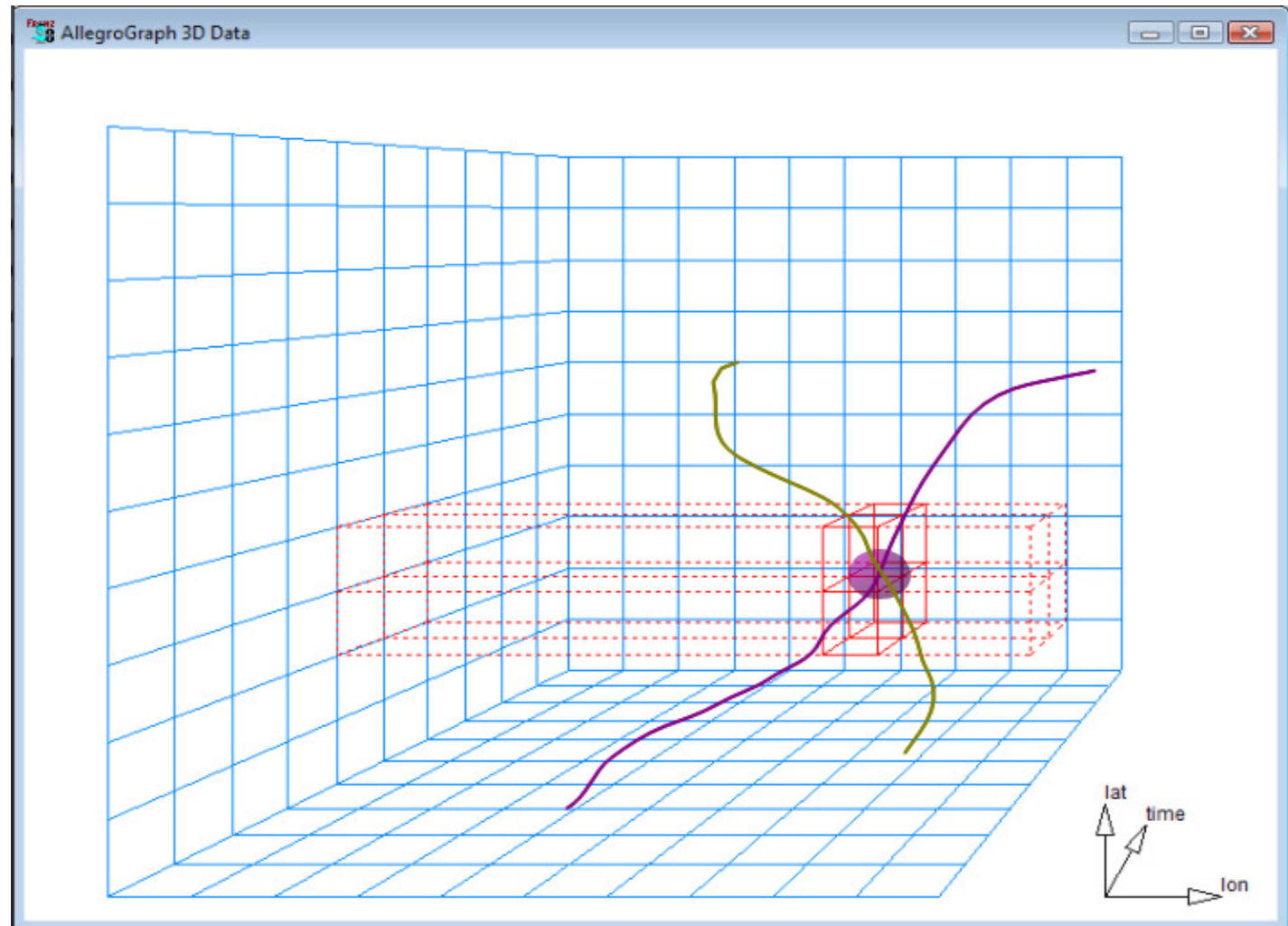
[1] What is within a given bound from a given lat/lon/time?





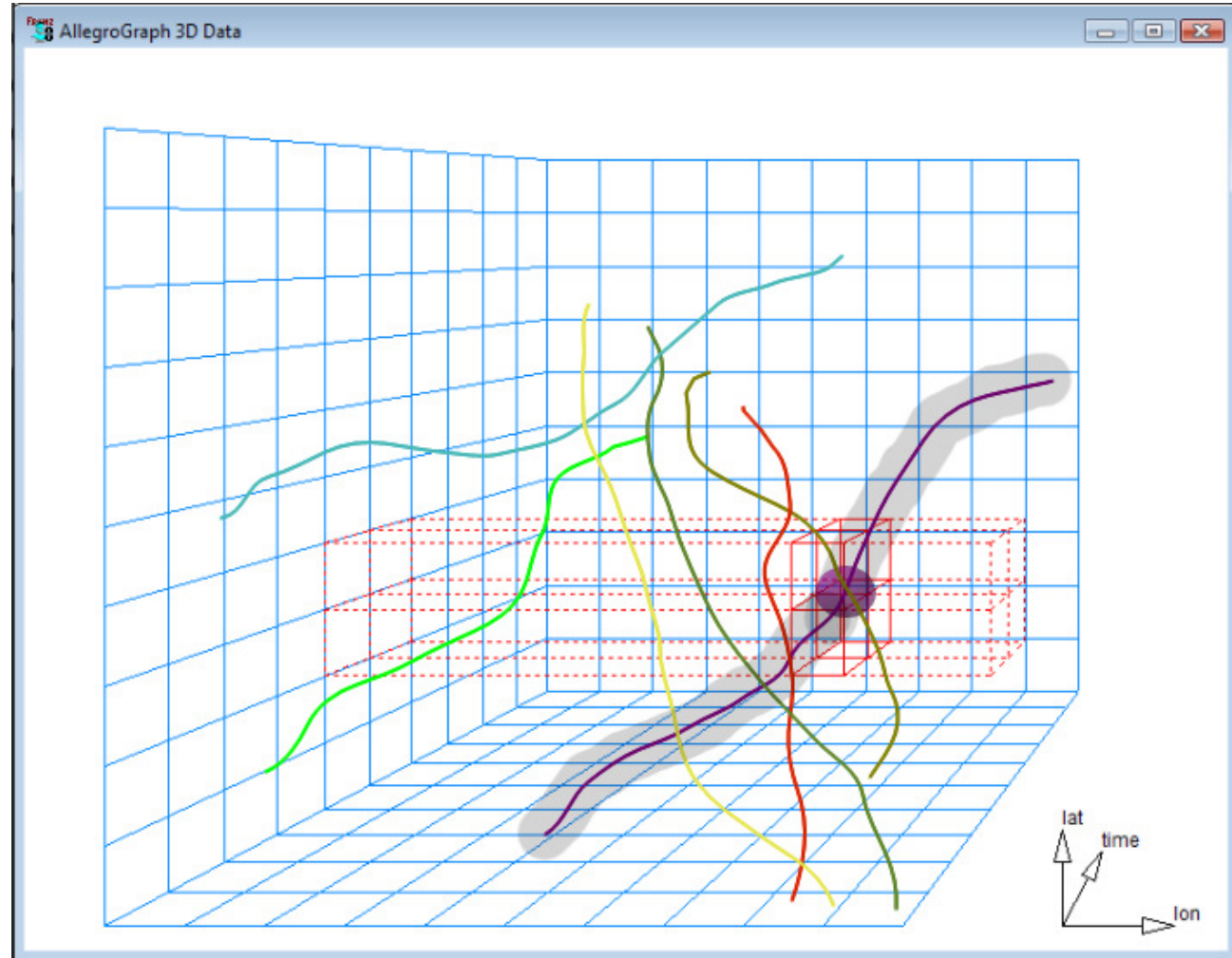
[2] Detect When Two Given MOB's Were Within a Given Distance

Requires
traversing
the entire
paths of the
two given
MOB's.





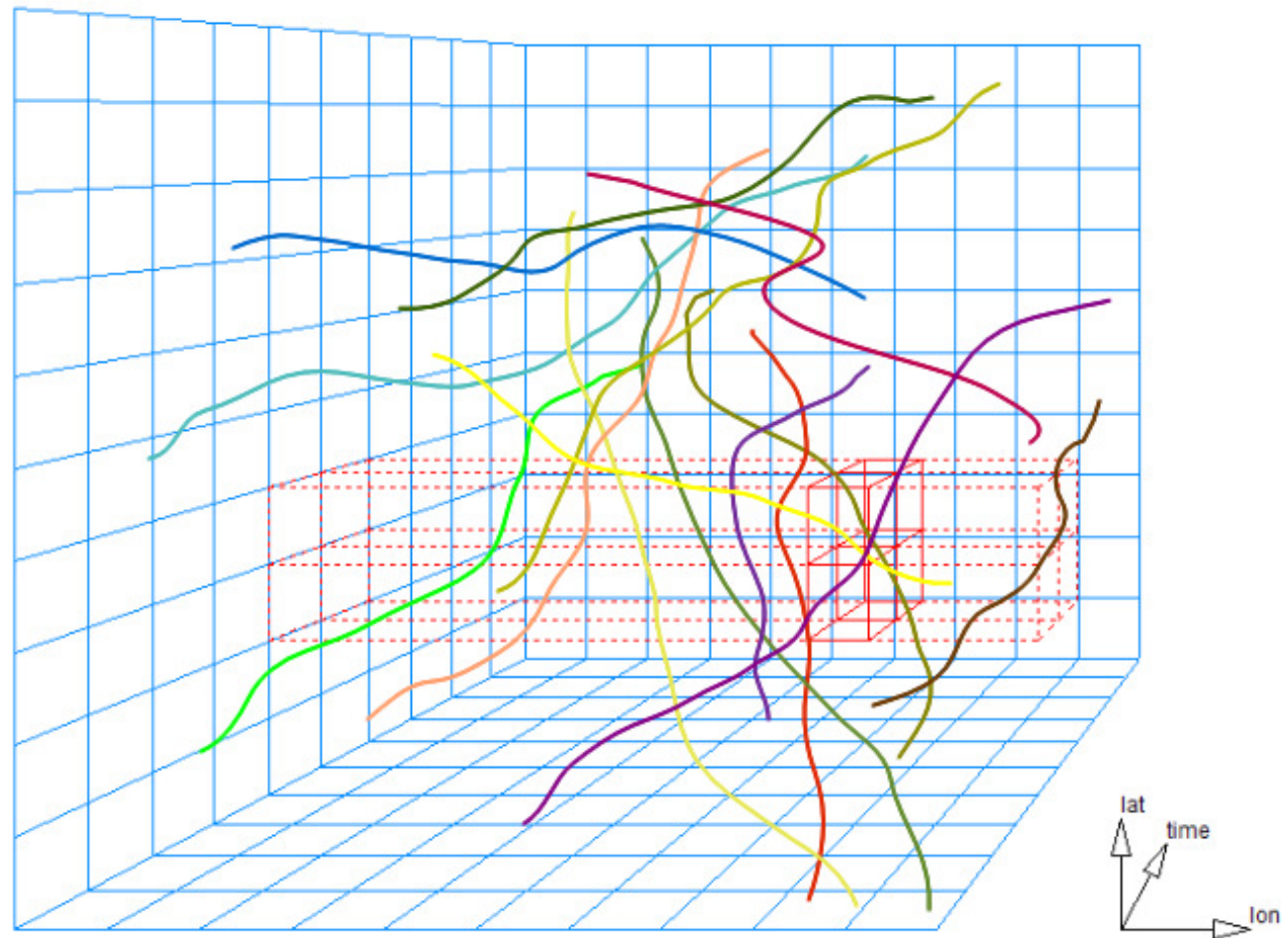
[3] Given a MOB, detect all MOBs ever within a given distance.





[4] Find all occurrences of two MOBs within a certain distance.

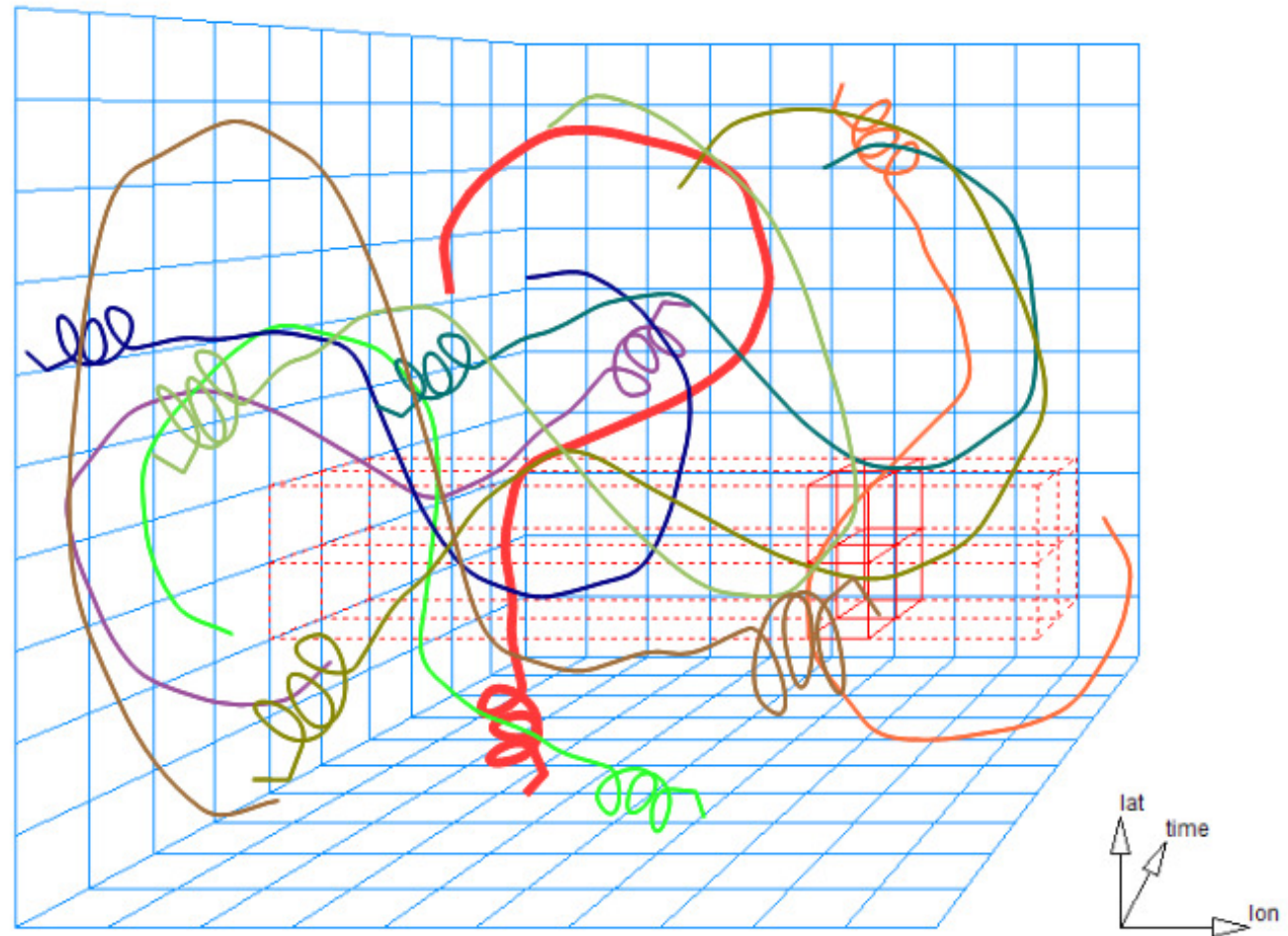
Don't know which MOB is important, so need to traverse and compare each pair of MODs, thus scanning the entire MOB component of the database.





[5] Detect Potential MOB Cliques

Both
expensive
and
difficult.
There is no
a priori
knowledge
what we
are looking
for...





And

- To summarize
 - You can do 2D geospatial and temporal now in AG32
 - Expect 3D indexing in 4.0 (or close thereafter)
 - We are looking to help people that want to do projects with moving objects
 - We are looking for 4.0 testers

Thanks

