

---

# Spatial Webs

Michael F. Goodchild  
University of California  
Santa Barbara

# The locations of computing

---

- User location **u**
  - the user interface
- Processing location **p**
  - $||\mathbf{u}-\mathbf{p}||$ 
    - 1960s < 10m
    - dedicated lines ca 1970 <10km
    - now no limit
- Data storage location **d**
  - independent of **u**, **p**
- Subject location **s**
  - independent of **u**, **p**, **d**

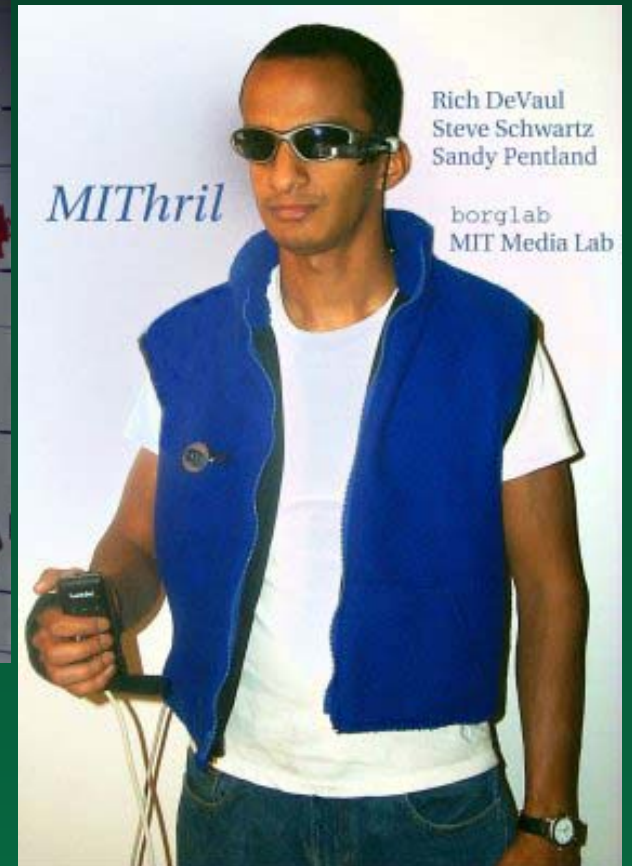
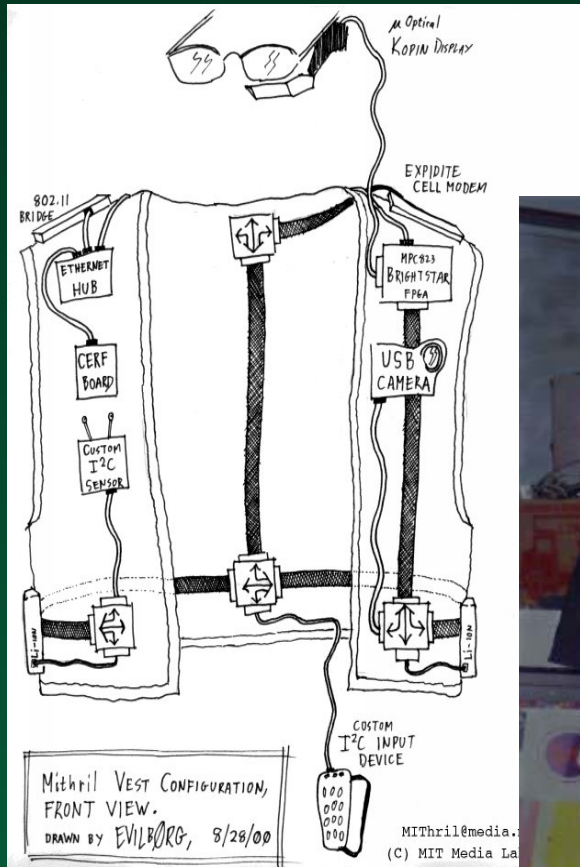
# User and subject

---

- $||\mathbf{u-s}|| = 0$ 
  - augmented reality (AR)
    - information system augments the senses
    - information system replaces impaired senses
  - LBS
  - contextual awareness
  - field work
  - emergency management
- $||\mathbf{u-s}|| \gg 0$ 
  - virtual reality (VR)
    - virtual tourist
    - desk-top GIS
    - Digital Earth







# Locations in field computing

---

- Many workers in the field (many **u**'s at **s**)
  - interacting with each other through the senses and through the technology
  - interacting with distributed services and data
- Possibly a hierarchical arrangement
  - a field manager also at **s**
  - a field base

# Options for $d$

---

- Where to store data in the SDI?
  - cost of dissemination goes to zero
  - close to  $s$ 
    - access to ground truth
  - level of interest determined by  $\|u-s\|$ 
    - information of geographically determined interest
    - geographic information is IGDI
    - but other information is not
  - convergence of  $u$  and  $s$

# Implications for finding data

---

- Heuristics for the SAP
- Geographic information is most likely to be found on a server located within its footprint
  - convergence of  $d$  and  $s$
  - but at what level in the hierarchy?
    - jurisdiction that most closely matches the footprint
    - $\max ||J \cap F|| / (||J|| ||F||)^{1/2}$

# Transitioning map libraries

---

- From central services to unique services
  - from general collections to special collections
  - from UCSB's Map and Imagery Laboratory to the Alexandria Digital Library
- There will always be more than one service
  - no amazon.com of geographic information
    - [www.alexandria.ucsb.edu](http://www.alexandria.ucsb.edu)
    - [www.geographynetwork.com](http://www.geographynetwork.com)
    - [www.fgdc.gov](http://www.fgdc.gov)
- Unique services must declare themselves
  - through collection-level metadata (CLM)
  - formalizing and publishing d

**WICK PLACENAME SEARCH**  
Search the entire world for...  
 Find  
Enter "Rome" if you want Rome, Italy.  
[Get more information](#)

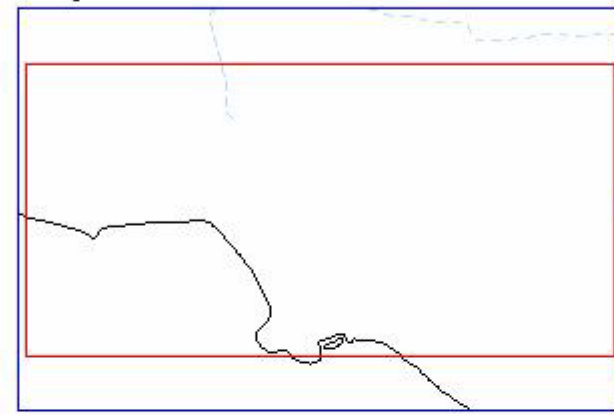
**GENERAL SEARCH**  
Select collection to search  
ADL Catalog  
[Browse collections](#)

**Set geographic region**  
Click the map to the right to set the geographic extent of the search, or directly enter bounding coordinates below.

N  
34.43  
W -118.97 -117.57 E  
33.73  
S

**Words to search for**  
  
Any of the above words  
All of the above words  
Exact phrase

## Map Browser



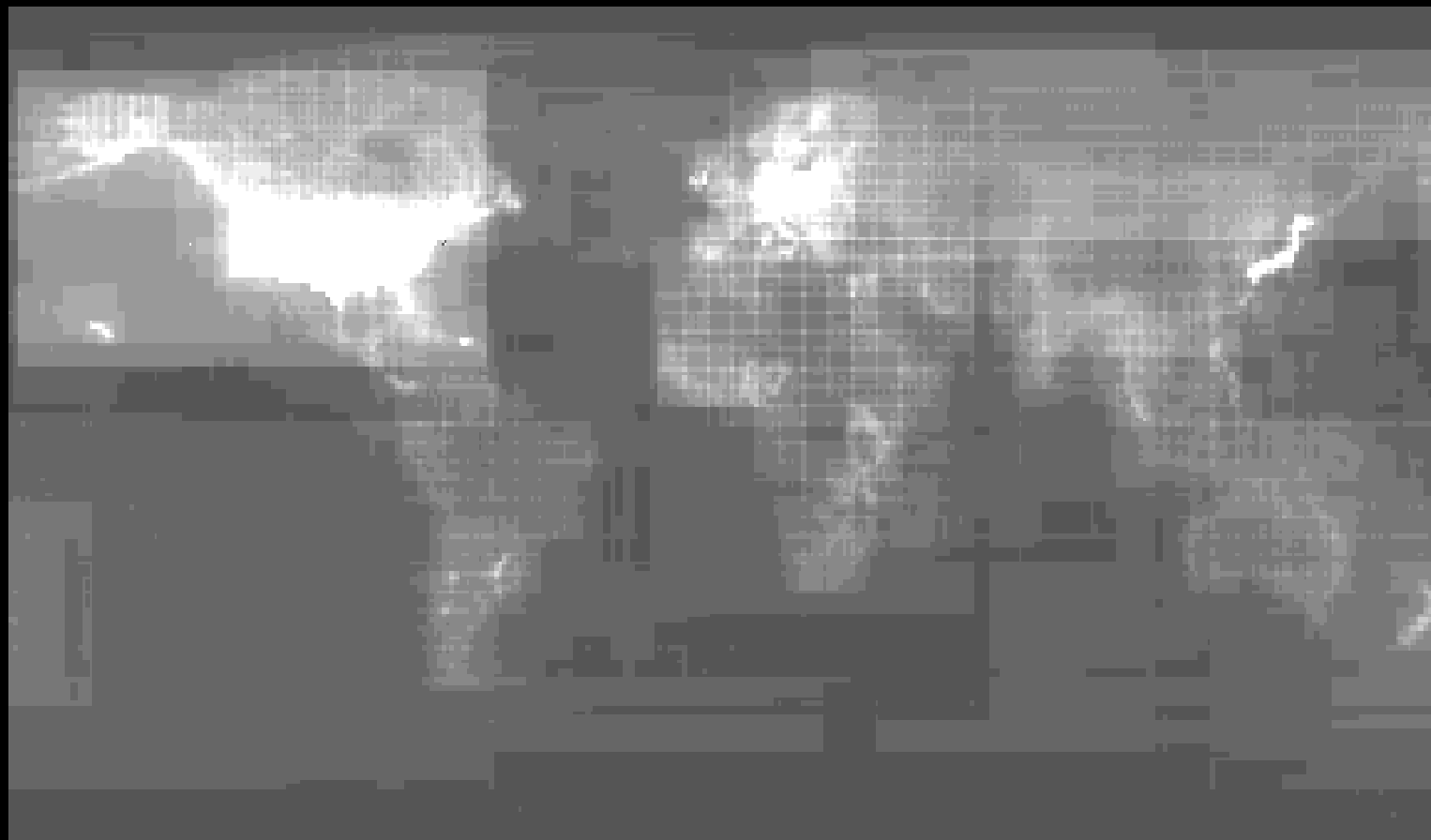
Click map to:  
Recenter & Zoom in  
Change location to:  
--

Navigation controls including directional arrows, zoom in/out buttons, a shape selection tool, and a Reset button.

## ADL Search Results

The query that produced these results can be found at [the bottom of this page](#).

- DRG o33117g6, Digital Raster Graphic of BLACK STAR CANYON, CA.**  
Type: maps. Format: TIFF. Date: 1988. ADL identifier: adl\_catalog:800279.  
[HIGHLIGHT IN MAP](#) • [COMPLETE DESCRIPTION](#) • [BROWSE GRAPHIC](#) • [ACCESS/DOWNLOAD](#)
- DRG o33117g7, Digital Raster Graphic of ORANGE, CA.**  
Type: maps. Format: TIFF. Date: 1981. ADL identifier: adl\_catalog:800280.  
[HIGHLIGHT IN MAP](#) • [COMPLETE DESCRIPTION](#) • [BROWSE GRAPHIC](#) • [ACCESS/DOWNLOAD](#)
- DRG o33117g8, Digital Raster Graphic of ANAHEIM, CA.**  
Type: maps. Format: TIFF. Date: 1981. ADL identifier: adl\_catalog:800281.  
[HIGHLIGHT IN MAP](#) • [COMPLETE DESCRIPTION](#) • [BROWSE GRAPHIC](#) • [ACCESS/DOWNLOAD](#)



CLM of the Alexandria Digital Library

# Knowing where to look

---

- Approaches to CLM
  - by data type
    - [ortho.mit.edu](http://ortho.mit.edu)
  - by area of the globe
    - Arctic Data Directory
  - the one stop shop
    - [www.fgdc.gov](http://www.fgdc.gov)
  - a new generation of search engines
    - identifying footprints

# Options for p

---

- Where to process?
  - server or client, which server?
  - published services
    - directories
    - [www.geographynetwork.com](http://www.geographynetwork.com)
    - evolving g.net
  - description standards
    - UDDI: Universal Description, Discovery and Integration
    - WSDL: Web Service Definition Language

# p and u

---

- $\|p-u\| = 0$ 
  - computing in the client
    - using local data,  $\|u-d\| = 0$
    - using remote data
- $\|p-u\| > 0$ 
  - send data to the service from the client
  - link a remote service to a remote data source,  $p \neq u$ ,  $d \neq u$

# Costs and benefits

---

- More cycles available remotely
  - integrating and exploiting waste cycles
  - the Grid
  - SETI
- Intellectual property issues
  - intellectual value of service
  - risk of dissemination
  - commercial value
- Update, versioning issues
  - distributed service has versioning problems
- Process coupled to data, well defined

# geography network explorer

search

browse

1 Type place name & press Go:

GO

or draw search area  :



2 Choose content type:

Geographic Services

Choose content theme:

<All Content Themes>

Optional Keyword (e.g., river):

SEARCH

Search NSDI Clearinghouse

Records Found: 7



## Content Found by Search

### Geographic Services

**Publisher:** Tele Atlas North America, Inc.

**Content Title:** EZ Locate

**Coverage Area:** United States

[Link to Content](#)

[View Details](#)

**Publisher:** Tele Atlas North America

**Content Title:** ArcIMS Route Server

GeoService Demo

**Coverage Area:** 50 United States

**Map Scale:** 1:24000

[Link to Content](#)

[View Details](#)

**Publisher:** ESRI

**Content Title:** Place Finder Sample Web

Service

**Coverage Area:** global

[Link to Content](#)

[View Details](#)

# High-priority geoservices

---

- Geocoding
  - tied to data, update issue
- Gazetteer
  - conversion between general or domain-specific placename and coordinates
  - geoparsing
    - identification and decoding of placename references in text
    - mapping and associating news stories
  - queries based on placenames
    - how far is the capital of Belgium from the capital of France?
- What else, is there a general model?

# The future spatial Web

---

- Computing embedded everywhere
  - ground-based sensor networks
    - nano-scale dust
  - networks of space-borne sensors, UAVs
  - mobile, ubiquitous computing
    - GIS capabilities everywhere
  - infinite wireless bandwidth
    - living in a soup of radiation

# Associated technologies

---

- Grid computing
  - harvested cycles
  - distributed services and data
    - transparent access
    - search engines, metadata
- Interoperability
  - semantic tools
  - ontologies

# But...

---

- Overlapping metadata domains
  - FGDC for geospatial
  - EML for ecological data
  - DDI for social data
    - but EML, DDI have spatial components
    - impossible to define non-overlapping domains
  - metadata light as the umbrella ontology
    - Dublin Core

# and...

---

- Mapping between ontologies, semantics
  - successful if 1:1, 1:n, n:1
  - but not if partial, fuzzy, uncertain
  - if it was simple enough to be automated it would have been done years ago
- The CLM issue
- The cost/benefit ratio for metadata
- IP address space
  - 2 billion assigned of 4 billion possibilities
    - 32 bit
  - IPv6 is on the way
    - 64 bit

# Four types of interoperability

---

## ■ Syntactic

- does the receiver understand the sender's coding?
  - mapping from  $\langle x, z \rangle$  to  $\{0, 1\}$
- coding defined by metadata
  - are there outstanding metadata issues
  - does FGDC do it all?

# Syntactic interoperability issues

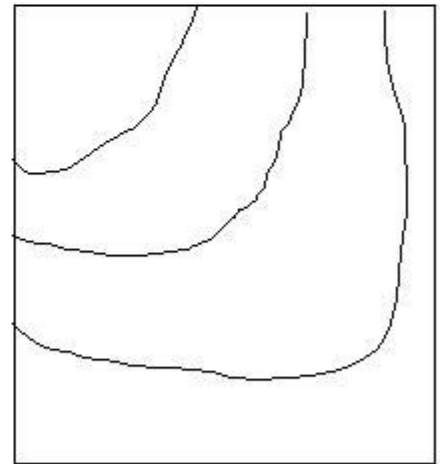
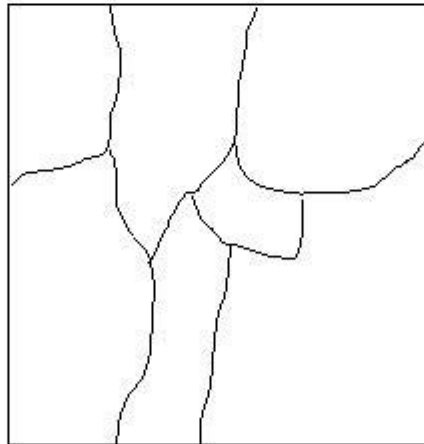
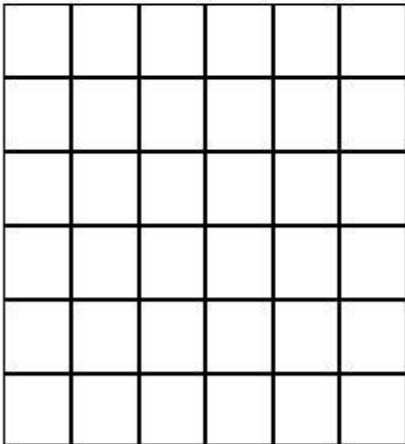
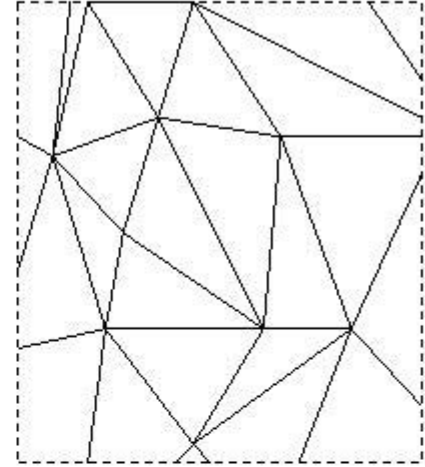
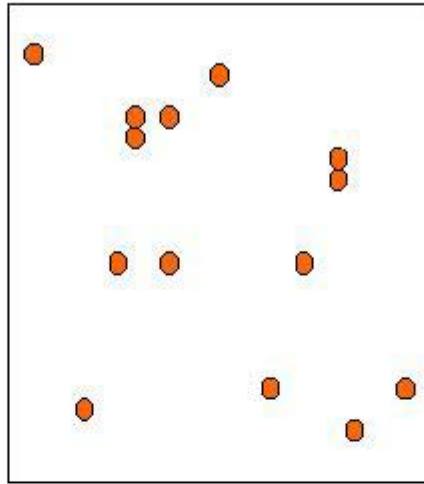
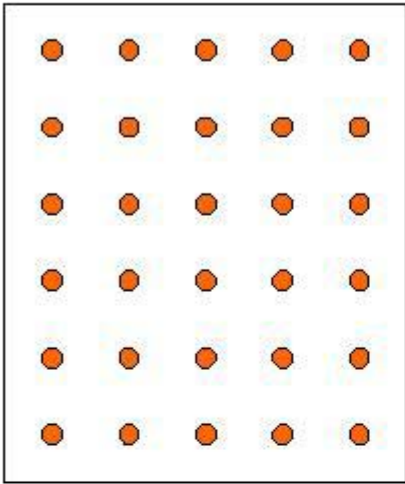
---

- FGDC as OLM
  - need for CLM
  - what is an object?
    - feature
    - class
    - feature data set
    - database
- Abstraction hierarchy
  - Tsou GIScience 2002 paper

# Spatial support issues

---

- Representations of a field
  - $z = f(\mathbf{x})$
  - six discretizations
  - inherent spatial resolution
    - explicit only for the two raster representations



# Techniques

---

- Areal interpolation
  - Phaeton, Enki
- Resampling
- Hiding the representation
  - Kemp 1997
  - Vckovski 1998
  - Van Deursen language

# Accuracy issues

---

- Two data sets with different lineage
  - positional accuracy
  - ISSDQ 1999 paper
    - M.F. Goodchild (2002) Measurement-based GIS. In W. Shi, P.F. Fisher, and M.F. Goodchild, editors, *Spatial Data Quality*. New York: Taylor and Francis, pp. 5–17.

# Positional uncertainty

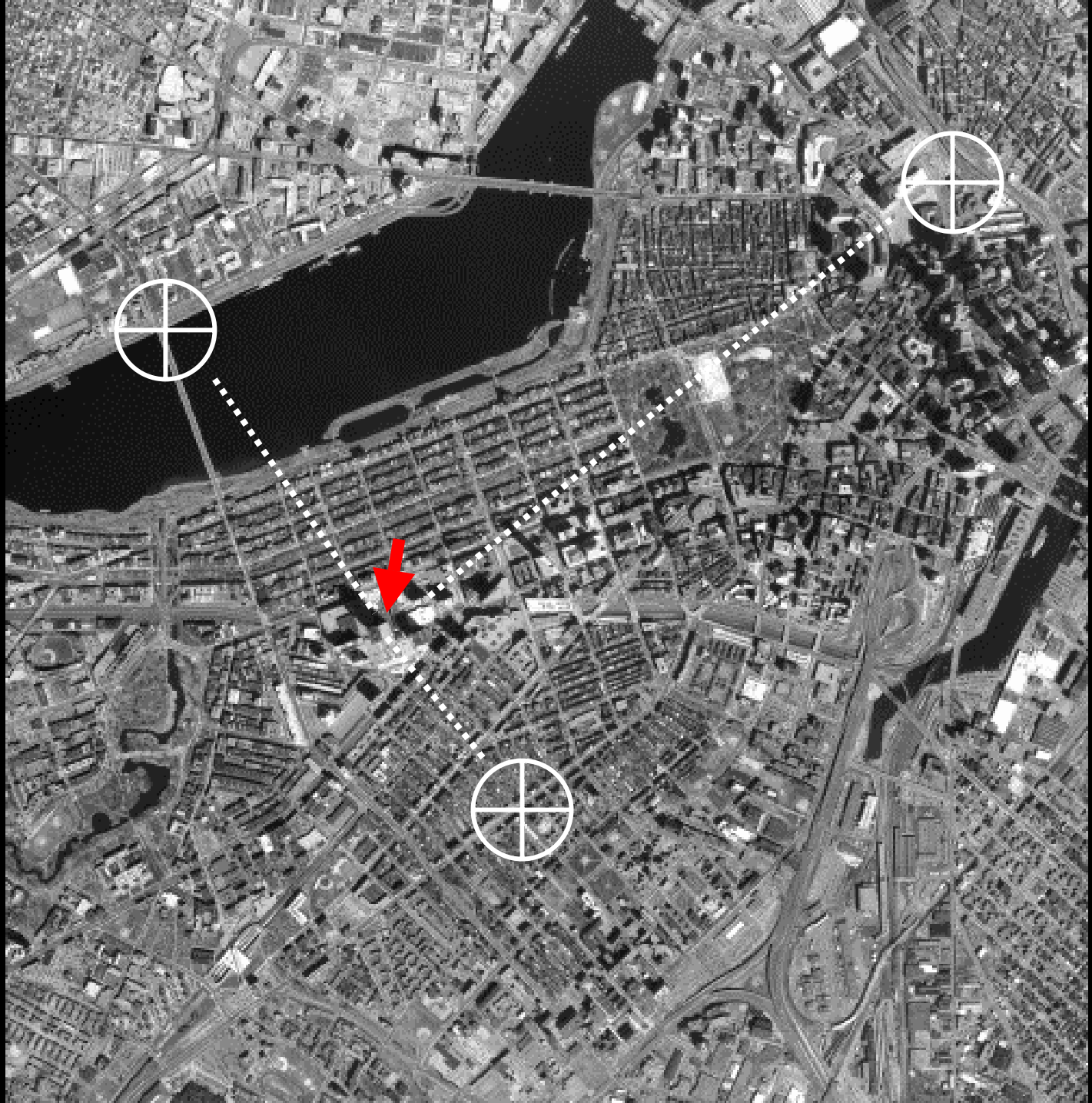
---

- The fundamental item of geographic information  $\langle x, z \rangle$ 
  - uncertainty in  $x$
- Geographic location
  - absolute
  - stored at point, object, data set level
  - return absolute position

# Measurement of position

---

- Position measured
  - $\mathbf{x} = f(\mathbf{m})$
- Position interpolated
  - between measured locations
  - surveyed straight lines
  - registered images
- The inverse function
  - $\mathbf{m} = f^{-1}(\mathbf{x})$



# Theory of measurement error

---

- Measured value = true value + distortion
  - $x' = x + \delta x$
  - some derived value  $y = x^2$
  - $y + \delta y = (x + \delta x)^2$
  - expanding and ignoring terms in  $(\delta x)^2$
  - $\delta y = 2 x \delta x$
  - more generally if  $y = f(x)$ ;  $\sigma_y = df/dx \sigma_x$
  - generalizes to several variables, variance-covariance matrices

# Errors of position

---

- Location distorted by a vector field
  - $\mathbf{x}' = \mathbf{x} + \varepsilon(\mathbf{x})$
  - $\varepsilon(\mathbf{x})$  varies smoothly
- Database with objects of mixed lineage
  - different vector fields for each group of objects
  - lineage may not be apparent
    - e.g. not all houses share same lineage

# Absolute and relative error

---

- Two points  $\mathbf{x}_1, \mathbf{x}_2$ 
  - perfect correlation of errors,  $\varepsilon(\mathbf{x}_1) = \varepsilon(\mathbf{x}_2)$ 
    - no error in distance
  - zero correlation of errors
    - maximum error in distance
- Absolute error for a single location
  - measured by  $\varepsilon(\mathbf{x})$
- Relative error for pairs of locations
  - value depends on error correlations



# Implications

---

- Most GIS operations involve more than one point
  - e.g. distance, area measurement, optimum routing
  - knowledge of error correlations is essential if error is to be propagated into products
  - joint distributions are needed
  - statistics such as the confusion matrix provide only marginal distributions

# The inverse $f^{-1}$

---

- An error is discovered in  $\mathbf{x}$ 
  - error at  $\mathbf{x}_1$  is correlated with error at  $\mathbf{x}_2$
  - both errors are attributed to some erroneous measurement  $\mathbf{m}$
  - to determine the effects of correcting  $\mathbf{x}_1$  on the value of  $\mathbf{x}_2$  it is necessary to know  $f$  and its inverse  $f^{-1}$

# Definitions

---

- *Coordinate-based GIS*
  - locations represented by  $\mathbf{x}$
  - $f$ ,  $f^{-1}$  and  $\mathbf{m}$  are lost during database creation
- *Measurement-based GIS*
  - $f$  and  $\mathbf{m}$  available
  - $\mathbf{x}$  may be determined on the fly
  - $f^{-1}$  may be available

# Partial correction

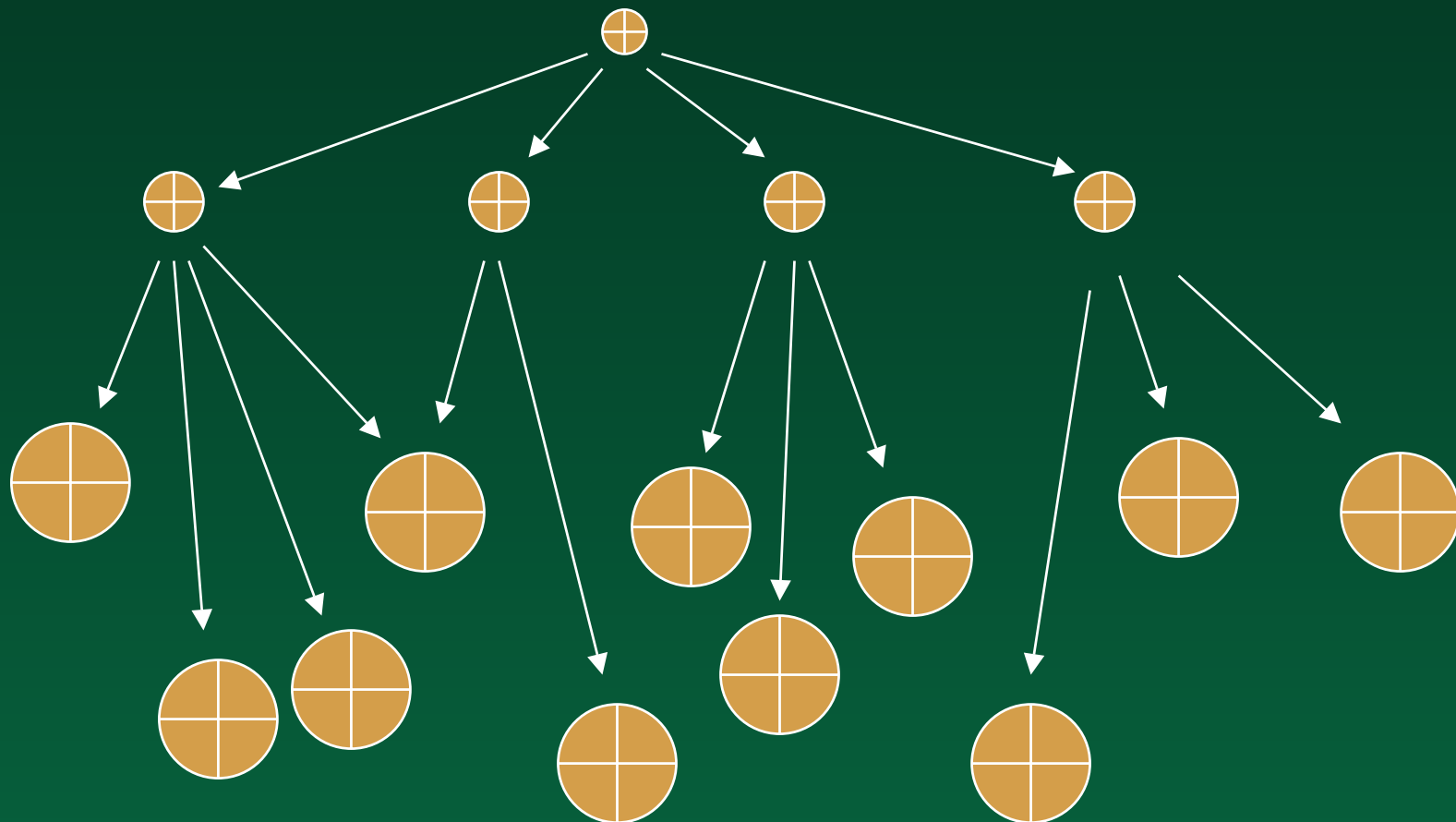
---

- The ability to propagate the effects of correcting one location to others
  - preserving the shapes of buildings and other objects
  - avoiding sharp displacements in roads and other linear features
- Partial correction is impossible in coordinate-based GIS
  - major expense for large databases

# The geodetic model

---

- Equator, Poles, Greenwich
- Sparse, high-accuracy points
  - First-order network
- Dense, lower-accuracy points
  - Second-order network
- Interpolated positions of even lower accuracy
- Locations at each level inherit the errors of their parents



# Formalizing measurement-based GIS

---

- Structured as a hierarchy
  - levels indexed by  $i$
  - locations at level  $i$  denoted by  $\mathbf{x}^{(i)}$
  - locations at level  $(i+1)$  derived through equations of the form  $\mathbf{x}^{(i+1)} = f(\mathbf{m}, \mathbf{x}^{(i)})$
  - locations at level 0 *anchor* the tree
  - locations established independently (GPS but not DGPS) are at level 0

# An example

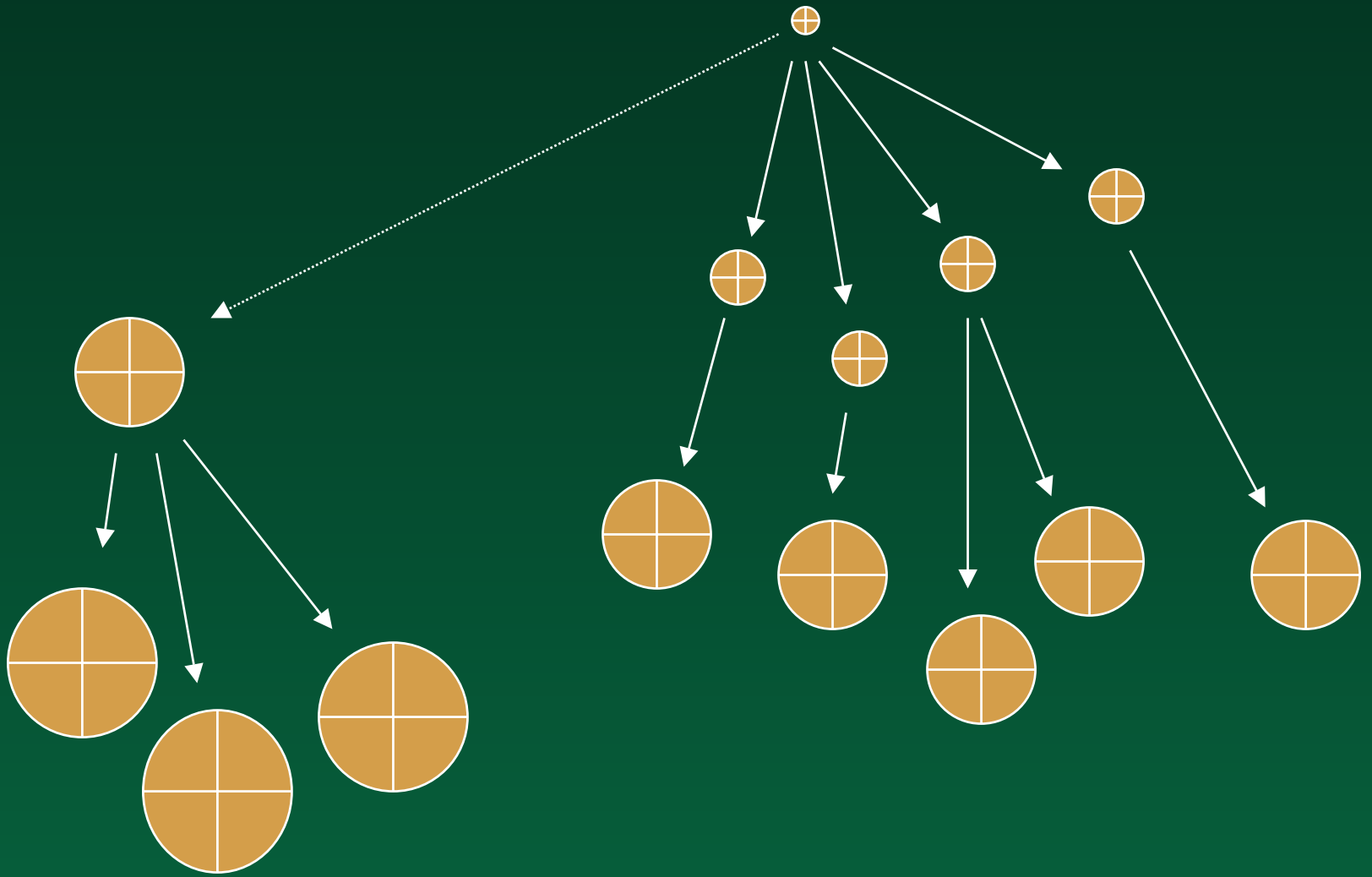
---

- A utility database
- Pipe's location is measured at 3 ft from a property boundary
  - $m = \{3.0, L\}$
  - property at level 3, pipe at level 4
- Property location is later revised or resurveyed
  - new  $m = \{2.9, L\}$
  - effects are propagated to dependent object

# Beyond the geodetic model

---

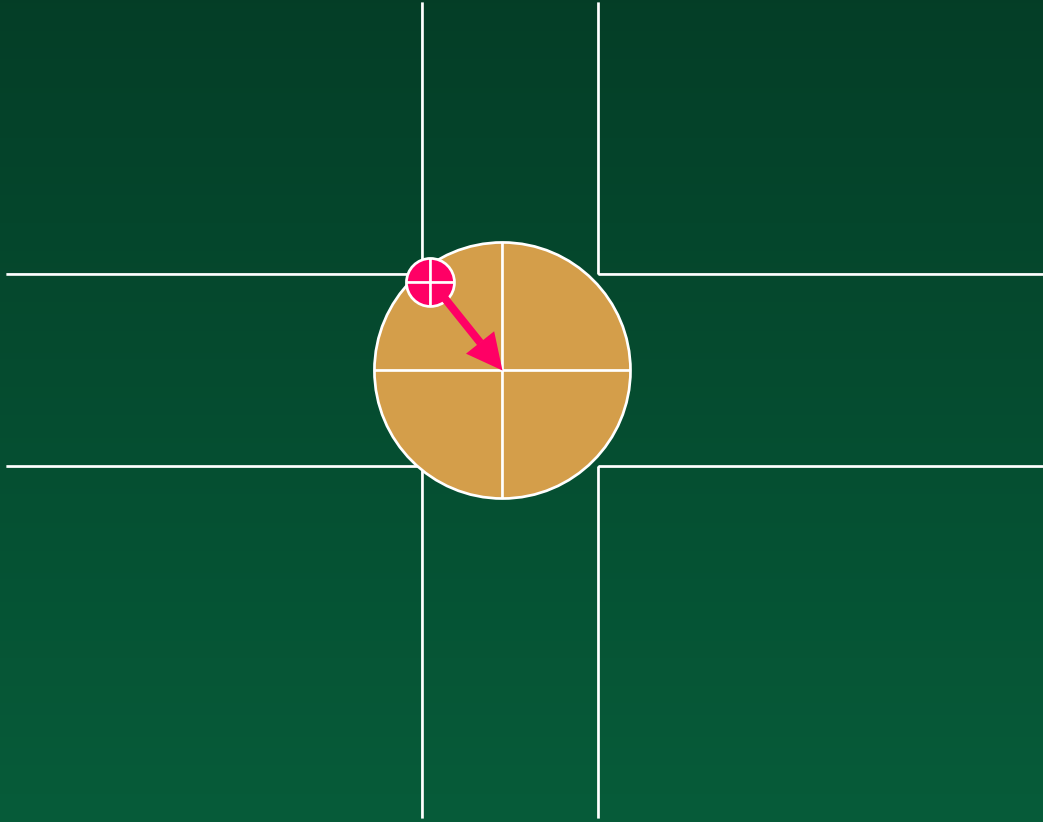
- National database of major highways
  - 100m uncertainty in position
    - sufficient for agency
    - relative accuracies likely higher, e.g. highways are comparatively straight, no sudden 100m offsets
- Local agency database
  - 1m accuracy required
  - two trees with different anchors



# Merging trees

---

- Link with a pseudo-measurement
  - displacement of 0
  - standard error of 100m
  - revisions of the more accurate anchor can now be inherited by the less accurate tree
    - but will normally be inconsequential



# Interoperability metadata

---

- OLM as  $n$  metadata
- Need for  $n(n-1)/2$  metadata
  - " $n^2$ " metadata

# Semantic issues

---

- Ontology mappings
- Existing products
  - how applicable are they to the geographic case?
  - how compatible with other types of interoperability?

# Immediate tasks

---

- Spatial support
  - Phaedon and Philipp
  - towards an integrated approach
  - building on previous work
  - prototype
  - implemented in Arc9

# Syntactic interoperability

---

- Mike and Matt
- Collect all literature on hierarchical metadata
  - OLM, CLM, hierarchical schemas
- Examine state of metadata in Arc9
  - wrt Geodatabase
- Devise generic approach

# Other tasks

---

- December meeting
  - Upham
- Web site
  - Gamaiel