GDI 2010:
Generalization and Data integration

University of Colorado - Boulder
20-22 June 2010

Organizers

- Barbara Buttenfield U. Colorado – Boulder
- Cynthia Brewer Penn State University
- Keith Clarke U. California – Santa Barbara
- Michael Finn USGS-CEGIS
- Lynn Usery USGS-CEGIS

Focus

- Significant progress in knowledge about generalization
  - Scale-change and topographic multi-scale base mapping
  - Design and construction of multi-representation databases (MRDBs)
  - European national mapping agencies and researchers especially active
- Context is national mapping
- Special emphasis on data modeling and data integration

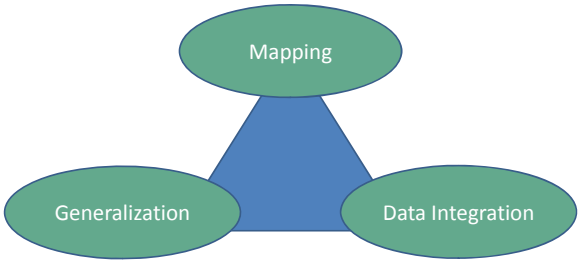
Challenges

- Impediments to building fully functional MRDBs
  - Multiple resolution databases
  - Multiple representation databases
- Integration of varied data representations
  - Establishing links between multiple representations
  - Data fusion
  - Conflation
  - Conflict resolution
  - Other data modeling and mapping tasks

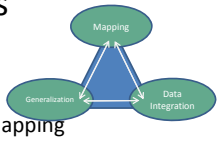
Goals of the Symposium

- Catalyze discussion between data integration and generalization communities
  - Role of submitted abstracts – context and scope
  - Share expertise on problems of common interest
  - Benefits to both scientific communities
  - Benefits to national mapping
- Identify problems which can be addressed given current states of knowledge
- Prioritize challenges which persist
- Opportunities for collaboration
- Research volume

One Way to Look at It



Mechanics



- Breakout sessions
  - Integration and Generalization (processing and data organization)
  - Generalization and Multiscale Mapping (data modeling and data display)
  - Integration and Multiscale Mapping (national and international frameworks)
- Focus discussion on:
  - Share states of current knowledge
  - Challenges and solutions (short, long term
  - Areas for collaboration?

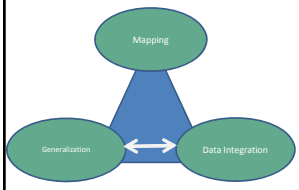
Integration and Generalization

Processing and Data Organization

Integration and Generalization

Processing and Data Organization

1. Do we need multiple representations or not?
2. What is the balance between automation and quality?
3. Generalization vs. Integration Methods



Integration and Generalization

Processing and Data Organization

1. Do we need Multiple Representations?
   - Disadvantages MRDBs:
     - Expensive and difficult to maintain, possibly inconsistent
     - Inflexible, e.g. difficult to change database schema
     - Redundancy increases update work and errors
   - Ideal: one database at "best" scale

   BUT THIS CARRIES PROCESSING PENALTIES
   - Need fully automated, reproducible, deterministic generalization process
   - Difficult to maintain relations:
     - with historical versions (maps for planning or decision tracking)
     - between features over scales (unique, long-persisting ID)

   Mobile and static databases:
   - Don't maintain multiple versions in mobile, integrate on-the-fly
   - Challenges = real-time integration, geographic IR, harvest VGI content

Integration and Generalization

Processing and Data Organization

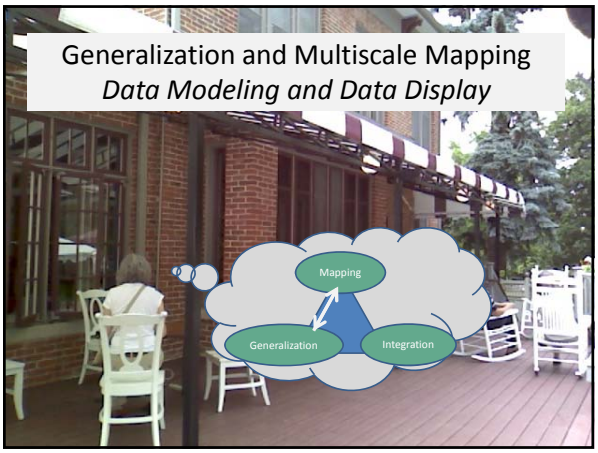
2. Balance between automation and quality
   - Limits to automated generalisation processing persist. Human interaction still needed to reach high quality generalization
   - Standing assumption that paper must be higher-quality representation
   - Minimum requirement = maintain topological relationships but sometimes underlying imagery is enough
- Consequences of mismatch vary in different application domains
  - Severity of mismatch relative to severity of potential consequences.
  - Quality is closely related to intended application (fitness for use)
- Common drivers for expected quality of generalization
  - Cost constraints - relaxed specifications when automating production lines
  - Legal aspects
  - Error in manual process carries less weight than in automated process (human error vs. whole system put into question)
  - Decreasing quality expectations of end user – nothing to compare against

Integration and Generalization

Processing and Data Organization

3. Integration Methods that support generalization
   - Resolving data sets at different accuracies / scales (align roads w/ imagery)
   - Align across data layers (rivers w/ bridges; car accidents lie upon roads)
   - Resolving data across boundaries (national, int'l) and edges (of data sets)
   - Alignment w/ ancillary data (DEM, landcover, ortho as referents for hydro)
   - Attribute integration across layers relates geometry and semantics
   - Integrate data collected at different times or time scales
     - Waterways at flood season

Integration challenges during generalization:
- Anecdotal information w/ geospatial data (precision)
- VGI/UGC data (unstructured and heterogeneous)
- Generalize geometry to facilitate integration (e.g., bldgs)
- Enrichment to facilitate generalization – add metrics and semantics

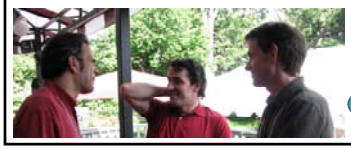


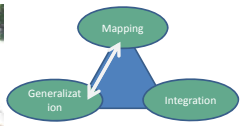
Generalization and Multiscale Mapping*

Data Modeling and Data Display

1. What generalization methods will serve future needs for multi-scale mapping?
2. What are the trade-offs b/t automation & quality in mapping?
3. What map user benefits/costs of generalized data?
4. How to generalize data for multiscale national mapping from private data, images, and volunteered- data contributors?

* aka DLM / DCM products





Generalization and Multiscale Mapping

Data Modeling and Data Display

1. What generalization methods will serve future needs?
   - Much of Europe works with richer object oriented model than N. America and this creates additional support for mapping
   - Consistency of input data critical, methods to validate UGC not developed
     - Schema modification: Variety of data models in local US governments is a problem, especially considering they have common tasks
   - Standard target scales (Google/Microsoft caching as de-facto standard)
   - Data items can vary in quality within one database
     - Assess and report item level uncertainty
     - Crowd sourcing for data updating as suggested option
   - LiDAR, point clouds and other improved (point) source data
     - Contours can be replaced by digital derivatives from DEMs
     - Automatic classification of point data through ontologies
     - Surface modeling improvements
     - Frameworks for generalizing sensor network data
   - For many databases, many methods already exist (especially for 3D)

Generalization and Multiscale Mapping

Data Modeling and Data Display

2. What trade-off between automation /quality?
- Wrong to think more automation means lower quality… or higher.
  - Cost often determines level of quality which can be attained
  - Sometimes quality cannot be measured readily (UGC)
  - Unique/extreme cases will happen and need to be handled; finding exceptions is a challenge (symbols to facilitate error detection)
- Quality historically interpreted as accuracy -- Shift to timeliness /currency (OpenStreetMap *best* map available for Haiti post-quake)
- Need to better formalize generalization, improve metrics
- Automation takes longer to code and develop, but faster in long run
  - Challenge: Engage cartographers in the automation process
  - Mobile applications and streaming input requires automation
- Can / should UGC/VGI be generalized?

Generalization and Multiscale Mapping

Data Modeling and Data Display

3. What map user benefits/costs of generalized data?

BENEFITS
- Smaller data volume is advantage for data managements users
- Consistently (generalized) data that is authoritatively compiled is the most beneficial
- Generalized data gives prominence to features needing emphasis – better cartographic communication

COSTS
- Sharing commercial data, maps for commercial purposes
- We stop showing certain information at given smallest scales
- Rules for aggregation are needed for absorbing small entities into larger entities of differing types
- Star vs. incremental generalization -- processed outputs differ

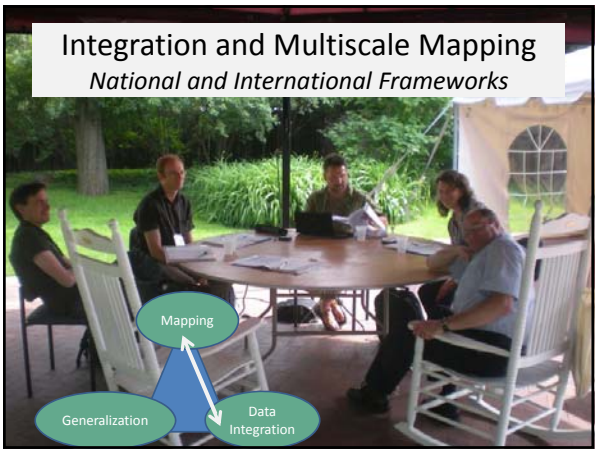
Generalization and Multiscale Mapping

Data Modeling and Data Display

4. How to generalize data from private data, images, and user-generated content?
   - Guidance about how to volunteer data is critical, also educating users about how data stewardship operates
   - Vendors allow use of commercial data by NMAs if it is generalized
   - Research community contributions:
     - Mashup is an example of how easy contributing should be
     - Tools to validate contributions for accuracy (location, orientation, topical, etc.)

Integration and Multiscale Mapping

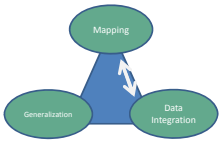
National and International Frameworks



Integration and Multiscale Mapping

National and International Frameworks

1. What integration methods or mapping methods can detect scale-inherited problems automatically?
2. What changes to mapping strategies which will emerge as new sources of geospatial data become commonplace?
3. To what degree can heterogeneous information be fused and conflated to support future national mapping efforts?
4. What software tools, use cases, and existing national frameworks best illustrate the future of data integration for mapping needs?



Integration and Multiscale Mapping

National and International Frameworks

1. Integration or mapping methods which can detect scale-inherited problems automatically

CURRENT CAPABILITIES FOR AUTOMATIC DETECTION LIMITED
- Edge matching to blend across data tiles
- Mapping tools to detect conflicts between features e.g., data overlap
- Raster data to verify vector data
- Agent-based approaches to detect conflicts b/t hydro and terrain
- Reason about some features to generate others (transportation nets to city bdy)

NOT AUTOMATICALLY DETECTABLE
- Features not resolved in original data (below measurement scale)
- Temporal conflicts in data (e.g. Broomfield County not in census data)
- Pattern recognition to detect discrepancy: offset in waterbody at tile edge

FUTURE DEVELOMENTS WHICH ARE PROMISING
- Harvestable data "heaps" will evolve: linked heterogeneous collections of data
- Improved data mining tools
- Spatial information tools for visually impaired

Integration and Multiscale Mapping

National and International Frameworks

2. Changes to mapping strategies as new sources of geospatial data become commonplace
- Event-based update strategy - periodic to continuous updates
- 3D building footprints derived from oblique Imagery
- Data with finer temporal resolution than spatial
- Ability to capture process, not just form
- Augmented reality, real-time annotation on national footprint
- Geotag photos without specifying positional precision or orientation
- Integrating real time assets e.g., GPS/Direction enabled cameras
- Crowd sourcing for faster updates: how to verify UGC data efficiently?

Integration and Multiscale Mapping

National and International Frameworks

3. Heterogeneous map fusion/conflation to support mapping
- Enriched capture methods inform automated tasks
  - Evaluate off-shore contractors' worker ID -- preferred worker list
- 3D views, semi-transparency, augmented reality
- Exploit collected imagery that a human never sees
- 'Enabled' sensors: wildland fire scanners, early warning systems
- Improve relational and geometric matching
- Efficient matching methods esp. networks, point data
- Heterogeneity across capture zones(USGS NHD, Inspire Road net)
  - Homogenize to coarser level of detail
  - Ground heterogeneity vs. heterogeneity due to capture process
- Semantic models to georegister features (deictic location)
  - petrol stations should be (re-)positioned along roads, but not points representing wind sensor
  - "the accident took place at the junction/after the bend " rather than just 'at location (x,y)' or 'between the two intersections'

Integration and Multiscale Mapping

National and International Frameworks

4. What tools, use cases, and existing national frameworks best illustrate the future of data integration for mapping needs?

- Ease of web maps: accessible API and kml key developments
- Evolution from discoverable data to services
- Integration across time important for historical analyses
- Increased importance of (item-level) metadata
- Attach coordinates to everything: not yet done for RDF triplets

EXAMPLES

- Germany, France, UK NMAs – consistency, links b/t scales
- China's 5-year Plan to integrate data
- OGC , INSPIRE, GeoBase
- RadiusStudio from 1Spatial - topology use
- ESRI helping local governments integrate da
- Walmart inventory
- FedEx GPS for efficient street mapping

GDI 2010: Generalization and Data Integration

Research Volume

Working Title: Generalization and Data Integration

- Full length papers (max 6,000 words)
- Target date submissions: 22 Oct 2010
- Covering any of the topics discussed at Symposium
- Utilize the abstract submitted as starting point, or submit a new paper. Some additional papers will be invited.
- Co-edited research volume: babs, Cindy and Keith
- Referee process – babs, Cindy and Keith plus one of the GDI 2010 participants, plus one external; period for revision
- babs soliciting a publisher (MIT Press, Taylor & Francis)