The Challenges of Geospatial Analytics in the Era of Big Data

Dr Noordin Ahmad
National Space Agency of Malaysia (ANGKASA)
CITA 2015: 4-5 August 2015
Kuching, Sarawak
Big data is an all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process them using traditional data processing applications.

http://en.wikipedia.org/wiki/Big_data
Big Data – A New Data Type for Geospatial?
Progression of Geospatial Information

- Device-Centric Geospatial Information
- Human-Centric Geospatial Information
- Feature-Centric Geospatial Information
- Region-Centric Geospatial Information

Number of users:
- IoT Space
- Indoor Space
- ?? <1m
- Pedestrian Navigation
- Car Navigation/Web Map
- Urban Planning/Civil Engineering

Size of space:

Source: S. Liang

Copyright © 2013, Open Geospatial Consortium
Big Geospatial Data ...........

Big processing of big imagery data
Imagery analysis in the cloud has become a reality through web services.

Flood of Big Data comes from mobile devices.
Big Data analytics is an effective way to enhance the power of location with many comes from mobile devices.

GIS moves to processing in the cloud.
Apache Hadoop, an open source framework based on Google’s MapReduce, has been used by many projects to perform geospatial processing in the cloud or in cloud-like enterprise data centers.

Transaction oriented analytics: processing of streaming spatial data.
Relationships become more apparent with each user interaction, each addition of data or each new algorithm that's introduced in a discovery session.
SATELLITE TRANSPORTING BIG DATA

DISTRIBUTED DATA CENTRES

- 12 Years of data
- 700 Terapixels of data
- 654k Landsat scenes
- 1M Hours of computation
- 10,000 CPUs used
- 4 Days to complete

Source: Hansen, Potapov, Moore, Hancher et al., Science, 15 November 2013
http://earthenginepartners.appspot.com/science-2013-global-forest
Till now Geospatial Big Data analytics has focussed on Data visualisation.
Reliable sources of geospatial information are essential to...

...support broad national objectives such as economic growth, social cohesion and well-being, and environmental management
<table>
<thead>
<tr>
<th>Phase</th>
<th>Data Type</th>
<th>Example Data Sets</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparedness and Early Warning</td>
<td>User-generated</td>
<td>Twitter (food crisis, earthquake), web traffic (Flu)</td>
<td>Requires machine learning, classification algorithms</td>
</tr>
<tr>
<td></td>
<td>Sensor</td>
<td>Precipitation (PERSIAN, TRMM), evapotranspiration, soil moisture, temperature, vegetation density and water content (MODIS, LANDSAT), groundwater levels (GRACE)</td>
<td>Should be paired with validated and calibrated biophysical models. Used for droughts, floods, fires, epidemics.</td>
</tr>
<tr>
<td>Impact and Response</td>
<td>User-generated</td>
<td>CDR, Flickr, Twitter</td>
<td>Requires verification, algorithms to separate signal from noise. CDR data requires agreement with cell provider</td>
</tr>
<tr>
<td></td>
<td>Sensor</td>
<td>Imagery (LANDSAT, MODIS, Geoeye) thermal (LANDSAT, MODIS), radar (RADARSAT-1, CARTOSAT), spatial video</td>
<td>Crowdsourcing aids damage detection. Crisis map mash ups by volunteer cartographers help publicize and visualize data.</td>
</tr>
<tr>
<td>Mitigation, Risk and Vulnerability Modeling</td>
<td>User-generated</td>
<td>CDR, emergency call content, facebook</td>
<td>Data may not be representative of population.</td>
</tr>
<tr>
<td></td>
<td>Sensor</td>
<td>Nighttime Lights (NTL), Imagery, thermal, Radar, spatial video, Temporal Flood Inundation Mapping (GIEMS)</td>
<td>Populations without electricity not identified by NTL. Must be paired with socioecological vulnerability models. Uncertainty and scale issues.</td>
</tr>
<tr>
<td></td>
<td>institutional, public</td>
<td>GCM (Global Climate Model), Transportation data (subway, bikeshare), census, Worldpop, Open Cities</td>
<td>Issues in uncertainty- GCMs are most uncertain at predicting extremes, Worldpop combines data scales for modeling and has associated uncertainties</td>
</tr>
</tbody>
</table>
The process .....
The challenge of making the Technology trends into “Reliable” Data (Values)

**Technological/Trends**
- Open Data
- Volunteered Geographic Information (VGI)
- Open Source
- Web 2.0 and the GeoWeb
- Cloud Computing
- Mobile and Location-based Services
- High Resolution Imagery
- Mass Market Geomatics
- “Big data”/data Integration

**Legal/Administrative Issues**
- Ethical Legal Practices
- Confidentiality, Security, and Sensitive Information
- Privacy
- Intellectual Property
- Copyright
- Licensing
- Data Sharing
- Liability
- Archiving and Preservation
- Data Quality
Collaborative Decision Making

Geospatial Data Sharing & Collaboration:
- Leveraging Agency Datasets
- Promotes Open Government

Benefits of Collaborative Decision Making from Geospatial Big Data:
- Same Map at the Same Time
- Shared Data Across Platforms
- Accelerates Situational Awareness
- Accelerates Decision Making
- Leverages Common Operating Pictures already in use
Complex Event Processing (CEP)
...rapid event processing, in-depth impact analysis, pattern matching etc.

**INSIGHT**
1. Capture: data capture and collection is where the cycle begins
2. Understand: data is only useful as the raw material of insight and understanding.
3. Model: requires more robust analytics capability.

**ACTION**
1. Decide: the power of real-time data analysis is its ability to enable real-time decisions.
2. Act: once the system have determined a particular response to this complex chain of events, it must be acted upon quickly.
3. Monitor: the efficacy of actions must be measured.

Complex event processing revolves around two poles: insight and action (http://spotfire.tibco.com/)
3 main type of Geospatial analytics

1. Predictive analytics

Likelihood of future events or happenings.
2. Operational Intelligence

Dynamic, business analytics that delivers visibility and insight into business operations. For instance, a wireless telephone company that monitors its network could use location data to determine which cell tower to fix based on where their high value customers are located.
3. Situational intelligence

Integrates and correlates large volumes of multidimensional real time and historical data to identify and act on a problem. Visualizing and analyzing this data can help answer questions like what happened, where did it happen, and why did it happen.
Concept of Sentinel Asia

Observation System

Archived Data Observation Data

Capacity Building

Information Sharing Platform

Data

Disaster Management Agencies

Digital Asia

Disaster Prevention

Residents in disaster areas

Warning Refuge Rescue