Big Data and Smart City

Dr. Winnie Tang
Chairman and Founder, Esri China (HK)
Agenda

• Smart City: Overview
• Big Data in Smart City
• Big Data Analytics
• Geospatial Big Data Applications and Privacy
Smart City: Overview
Our World
Is Increasingly Challenged

The Evidence Is Clear . . .
We Need to Be Smarter
We Need to Understand
. . . And Act Together

- Natural Disasters
- Energy
- Climate Change
- Migration
- Urbanization
- Loss of Nature
- Biodiversity
- Population
- Pollution
- Inequality
- Social Conflict
- Water
- Consumption

Consumption
The Challenges That Cities Are Facing

- By 2050, 2/3 of the earth’s population (6.4 billion) will live in cities
- Aging population leads to decreasing productivity
- Resource strain of energy, housing, transportation, jobs, public services, etc.
- Citizens demand higher transparency of Government
The Challenges of Hong Kong

- About 326,000 private housing units aged 70 or above by 2046, ~300 times of those in 2015
- 1/3 population would be aged 65 or above by 2040
- Overall labor force participation rates projected to decline from 59.3% in 2014 to 48.6% in 2064
Smart City

- Ultimate Goals:
  - To enhance quality of life of citizens
  - To foster knowledge economy
  - To create a vibrant eco-system
Smart City Components

Mixed-modal access
Clean & non-motorized option
Integrated ICT

Smart Mobility
21st Century Education
Inclusive society
Embrace Creativity

Smart People

Culturally vibrant & happy
Safe
Healthy

Smart Living

Entrepreneurship & innovation
Productivity
Local and global interconnectedness

Smart Economy

Enabling supply & demand size policy
Transparency & open data
ICT & eGov

Smart Government

Green buildings
Green energy
Green urban planning

Smart Environment
Key Technology Drivers of Smart City

- Big Data
- Cloud Computing
- Geographic Information System (GIS)
- Green Technologies
- Internet of Things (IoT)
- Mobile Technologies
Examples of Smart City Applications

- Smart Living: eHealth
- Smart Economy: Fintech
- Smart Mobility: Connected Cars
- Smart Environment: Green Building
Worldwide Smart Cities Examples

• New York: LinkNYC Wi-Fi Station
• Singapore: Smart Parking Signage
• Barcelona: Smart Bus Network
Worldwide Smart Cities Examples

- San Francisco: Crime Map
- Minneapolis: Find My Solar Suitability
- Singapore: Smart Urban Planning of Jurong Lake District
Hong Kong as a Smart City

- Smart City @ Kowloon East (EKEO)
- HKSAR Smart City Blueprint (OGCIO)
- Hong Kong 2030+ (Development Bureau)
Smart City Consortium

- Smart City Consortium (SCC) is formed by a group of professionals from different corporations and organizations to provide opinions and suggestions to the Government for formulating related policies and standards in the development of Hong Kong as a world-class smart city. We encourage worldwide collaboration with different stakeholders to create the right ecosystem which fosters innovation and sustainable economic growth for Hong Kong.
Vision of Smart City Consortium

• To build Hong Kong as the world’s leading Smart City to foster knowledge-based economy, enhance the quality of life and to create a vibrant ecosystem leveraging relevant Information and Communication Technologies and adopting effective resources management.
Works of Smart City Consortium

- Organized and participated in over 60 events
- 13 MoU with other Smart City organizations signed
- Views and opinions on smart city blueprint for Hong Kong submitted
Big Data in Smart City
Worldwide Data in 60 Seconds...

- 1 New definition is added on Wikipedia
- 1,600 new reads on Scribd
- 13,000 hours of music streaming on Pandora
- 12,000 new ads posted on Craigslist
- 370,000 voice calls on Skype
- 98,000 tweets
- 100 new LinkedIn accounts
- 6,600 new pictures uploaded on Flickr
- 510,040 comments

- 20,000 new posts on Tumblr
- 13,000 new iPhone applications downloaded
- 600 new videos
- 70 new domains registered
- 60 new blogs
- 168 million emails are sent
- 694,445 search queries
- 1,700 Firefox downloads
- 125 plugin downloads
- 79,364 wall posts
- 100 new articles published

- 40 new questions asked on Answers.com
- 100 new articles published

- 25+ hours total duration

- 1,500 new blog posts

- 100 new LinkedIn accounts

- The largest social reading company

- The world's largest community created content

- The largest social reading company
Data Growth

- Data is growing at 40% annually, reaching nearly 44ZB by 2020
- Exponential growth of data requires new methods of collecting, storing and processing
The Type of Data in Various Sectors

<table>
<thead>
<tr>
<th>Sector</th>
<th>Video</th>
<th>Image</th>
<th>Audio</th>
<th>Text/numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banking</td>
<td></td>
<td></td>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Insurance</td>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Securities and investment services</td>
<td></td>
<td></td>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Discrete manufacturing</td>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Process manufacturing</td>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Retail</td>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Wholesale</td>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Professional services</td>
<td></td>
<td></td>
<td></td>
<td>Medium</td>
</tr>
<tr>
<td>Consumer and recreational services</td>
<td></td>
<td></td>
<td></td>
<td>Medium</td>
</tr>
<tr>
<td>Health care</td>
<td></td>
<td></td>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Transportation</td>
<td></td>
<td></td>
<td></td>
<td>Medium</td>
</tr>
<tr>
<td>Communications and media</td>
<td></td>
<td></td>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Utilities</td>
<td></td>
<td></td>
<td></td>
<td>Medium</td>
</tr>
<tr>
<td>Construction</td>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Resource industries</td>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Government</td>
<td></td>
<td></td>
<td></td>
<td>Medium</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td>High</td>
</tr>
</tbody>
</table>

1. We compiled this heat map using units of data (in files or minutes of video) rather than bytes.
2. Video and audio are high in some subsectors.

SOURCE: McKinsey Global Institute analysis
Big Data Market ($US Billions)

- Forecasted market size US$43.4 billion in 2017, to reach US$84.69 billion by 2026
China to Build National Big Data Center

- President Xi (Nov 2016):
  - The central government will establish a Big Data center at national level, so as to promote the convergence of technology, business and data, and provide coordinated public services based on data sharing across government departments at various levels.
Big Data: Definition

- Big data is a term for data sets that are so large or complex that traditional data processing application software is inadequate to deal with them. Challenges include capture, storage, analysis, data curation, search, sharing, transfer, visualization, querying, updating and information privacy. // (Wikipedia)
Characteristics of Big Data

- 5 Vs
  - Volume
    - The vast amounts of data generated every second
  - Variety
    - The different types of data can now be used
  - Velocity
    - The speed at which new data is generated and the speed at which data moves around
  - Veracity
    - The messiness or trustworthiness of the data
  - Value
    - Having access to big data is no good unless we can turn it into value
### How Big Data is Different

<table>
<thead>
<tr>
<th>Traditional Data</th>
<th>Big Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Large scale</td>
<td>• Massive scale</td>
</tr>
<tr>
<td>• Highly centralized</td>
<td>• Highly distributed</td>
</tr>
<tr>
<td>• Structured</td>
<td>• Unstructured</td>
</tr>
<tr>
<td>– Files</td>
<td>– Emails</td>
</tr>
<tr>
<td>– Records</td>
<td>– Audio/Video</td>
</tr>
<tr>
<td>– Databases</td>
<td>– Documents</td>
</tr>
<tr>
<td>• Sequential</td>
<td>– Spreadsheets</td>
</tr>
<tr>
<td>• Indexed</td>
<td>– Log files</td>
</tr>
<tr>
<td>• Processing transactions</td>
<td>– Sensor data</td>
</tr>
<tr>
<td></td>
<td>– Geo-spatial data</td>
</tr>
<tr>
<td></td>
<td>– Books</td>
</tr>
<tr>
<td></td>
<td>– Journals</td>
</tr>
<tr>
<td></td>
<td>– Blogs</td>
</tr>
<tr>
<td></td>
<td>– Text messages</td>
</tr>
<tr>
<td></td>
<td>– Chat sessions</td>
</tr>
<tr>
<td></td>
<td>– Search data</td>
</tr>
<tr>
<td></td>
<td>• Random</td>
</tr>
<tr>
<td></td>
<td>• Looking for patterns and relationships</td>
</tr>
</tbody>
</table>
Major Sources of Big Data

- IoT and Sensors Networks
- Smartphone Apps
- Social Media
Big Data from IoT and Sensors Networks

- Sensor technologies have existed for long time
- How the vast amount of data can be streamed to backend via communication network?
5G – More Data from Billions of Devices

- The fifth generation of cellular networking
- Early products available as soon as in 2018
- Dramatically increase:
  - Speed of data transfer
  - Response time
  - Capacity for billions of devices to be connected
Water Intelligent Network

- Being implemented progressively by Water Supplies Department (WSD)
- Continuous monitoring of water mains over the territory by installing sensors in the underground water supply network
- Determination of the most effective network management measures:
  - Active Leakage Detection & Control
  - Pressure Management
  - Quality & Speedy Repair
  - Re-provisioning of Water Mains
Big Data from Smartphone Apps

- Sensors within Smartphone are connected to Internet via the phone’s data connection
Big Data from Social Media

- Unstructured
- Texts, Check-ins, Hashtags, Voices, Pictures, Videos, etc.
Get Geographic Insights from Social Media

- Putting social media data on a map helps users track a dynamic situation.
  - Retailers display data feeds on maps to monitor and protect their brands
  - Banks use geographic analysis to detect fraud
  - Governments worldwide can gauge public sentiment in real time during significant events, such as elections and uprisings.
Big Data and Location

- All IoT sensors have locations
- The most common IoT sensors in Smartphones – GPS receiver
- All posts, photos and messages are tagged with phone or IP locations in social media
Geospatial Big Data

• “80% of data is spatial”
• Smart City services rely heavily on Geospatial Technologies
• For many big datasets, location is a crucial component to truly understand underlying patterns and trends
• Without location, datasets are less valuable, or in extreme circumstances - meaningless
Big Data Analytics
Why Big Data?

- Explore trends and patterns hidden in big data
- Model the phenomena leading to big data
- Predict what will happen in where and at when?
- Require Big Data Analytics
  - Turn big data into big understanding, insight and wisdom
Enabling Technologies for Big Data Analytics

- Grid Computing
- Parallel Processing
Geospatial Big Data Analytics

- Requires new science of Spatial Statistics
- GIS as a tool for spatial statistical analysis
# Big Data Analytics

<table>
<thead>
<tr>
<th>Traditional Analytics (BI)</th>
<th>vs</th>
<th>Big Data Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Focus on</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Descriptive analytics</td>
<td></td>
<td>• Predictive analytics</td>
</tr>
<tr>
<td>• Diagnosis analytics</td>
<td></td>
<td>• Data Science</td>
</tr>
<tr>
<td><strong>Data Sets</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Limited data sets</td>
<td></td>
<td>• Large scale data sets</td>
</tr>
<tr>
<td>• Cleansed data</td>
<td></td>
<td>• More types of data</td>
</tr>
<tr>
<td>• Simple models</td>
<td></td>
<td>• Raw data</td>
</tr>
<tr>
<td><strong>Supports</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Causation:</strong> what happened, and why?</td>
<td></td>
<td><strong>Correlation:</strong> new insight More accurate answers</td>
</tr>
</tbody>
</table>
Explore Geographic Patterns

- Use maps to see the story behind Big Data
  - Retailers can see where promotions are most effective and where the competition is.
  - Banks can get an explanation of why loans are defaulting and where there is an underserved market.
  - Climate-change scientists can see the impact of shifting weather patterns.
Find Spatial Relationships

• Seeing spatially enabled big data on a map allows users to answer questions and ask new ones
  - Where are disease outbreaks occurring?
  - Where is insurance risk greatest given recently updated population shifts?

• Geographic thinking adds a new dimension to big data problem solving and helps users make sense of big data
Perform Predictive Modeling

- Predictive modeling using spatially enabled big data helps users develop strategies from if/then scenarios
  - Governments can use it to design disaster response plans
  - Natural resource managers can analyze recovery of wetlands after a disaster
  - Health service organizations can identify the spread of disease and ways to contain it
Predicting Crime

- Rutgers University’s School of Criminal Justice: Risk Terrain Modeling
  - Examined how the environment affects illegal activity
  - Analyzed maps displaying spatial relationships between shooting incidents and nearby cluster of bars, clubs, liquor stores and fast food restaurants
  - Compared the migration paths of shooting incidents between 2 years
  - Conclusion 1: The presence of above businesses caused higher crime rates
  - Conclusion 2: Criminals shift illegal activities from one high-risk area to another over time
Improving Patient Care

- Louisiana Department of Health and Hospitals
  - To map epidemiological issues such as babies with low birth weights
  - Correlations between low birth weight rates and crime-riddled neighborhoods were discovered
  - Preventative healthcare measures can be introduced to reduce high-risk births, thereby cutting healthcare costs
Optimizing Banking Network

- Bank of America
  - Uses GIS to track every single transaction by customer, location, time and channel
  - Stores the information for 60 months of history (more than 20 TB of data)
  - An optimum plan of opening and closing branches and ATMs was drafted by combining capacity of each branch and ATM, cost and revenues of each facility location
  - Reduced its annual expenses by US$800 million while losing very few customers
Anti-Money Laundering

- To detect and discover fraud and suspected money laundering activities from billion numbers of transaction records daily by tracking and analyzing the flow of money from different locations at different periods of time using Geospatial Analytics.
Geospatial Big Data Applications and Privacy
Big Data Applications and Privacy Issues

- Most Big Data involves personal data especially that from Smartphones and Social Media
- Big Data Applications and Privacy are not mutually exclusive
- Privacy issues caused by technologies can be solved by technologies
- Just like balancing economic development with environmental conservation results in Sustainable Development, we can facilitate the technological & economic development brought by Big Data Applications while ensuring proper protection of privacy
Real Time Traffic Information

- To compute traffic speed GPS locations of Smartphones and base stations signals
- Data captured will be used for 24/7 speed and event data, location of each traffic congestion, traffic load level, Estimated Time of Arrival (ETA), etc.
- All personally identifiable information from mobility data is removed and Privacy is secure at all times, safeguarded at every human touch point
Mobile Mapping System

- To acquire road inventory data
- 3D Geo-Referenced images of roads will be captured
- Images of individuals faces and car plates number will be masked
Resolution of Census Data

• To strike a balance between the usefulness of Big Data and the right of individuals to privacy
Chief Data Officer (CDO)

- International Society of CDO:
  - CDO is a senior executive who bears responsibilities of creating data governance policies and using Big Data strategically to identify new business opportunities and reduce operating costs.
  - CDO must ensure that users have easy access to relevant data and reporting tools, and be able to trust in the data’s quality.

- Gartner:
  - 50% of all companies in regulated industries will have a CDO this year.
City of Los Angeles

GeoHub

Eric Garcetti
Mayor

Lilian Coral
Chief Data Officer
Story Maps and Apps Based on GeoHub

LA Clean Streets Index

Tech Industry of LA

Vision Zero LA
Mayor’s Dashboard

- To display a range of daily statistics covering areas such as public safety, city services, human services and economic development
- Let both the Mayor and the public to track daily operations and progress on major strategic objectives
- Data from various Open Data sources including GeoHub
Good Practices for Organizations

• To attain industry certifications
• Get advices from responsible authorities including Privacy Commissioner for Personal Data
• Employ consultants to conduct a Privacy Impact Assessment (PIA) or a Privacy Audit to assess the privacy risks involved and adopt measures to safeguard personal data when engaging in Big Data analytics or personal data sharing.
Privacy and Security Compliance of Esri’s Products

Esri Corporate Compliance

- ISO 27001
  - Esri’s Corporate Security Charter

- Privacy Assurance
  - US EU/Swiss SafeHarbor self-certified
  - TRUSTed cloud certified

Privacy & Compliance
Summary Across ArcGIS Online

Privacy
- TRUSTe
- EU-U.S. Privacy Shield
- Under Review

Compliance
- FISMA
- FedRAMP

Answers
- CSA Security Alliance
- ArcGIS Online | EMCS

Trust.ArcGIS.com
Conclusion

• **Big Data in Smart City** can be visualized from a Geospatial perspective that reveal patterns, trends, and relationships

• **GIS can pull Big Data** all together to help focus Smart City citizens’ decision making

• **GIS can help to analyze Big Data**
  - To solve problems easier by visualization through maps
  - Geospatial analysis can improve the accuracy of Big Data analytics

• **Big Data Applications** can be developed and implemented by striking a proper balance between economic & societal benefits and the right of individuals
Thank You

Dr. Winnie Tang
iam@winnietang.hk
www.esrichina.hk