BEYOND THE DATA:
Extracting Knowledge, Deriving Insight, Delivering Intelligence

Hani S. Mahmassani

NUTC Data Driven Business Workshop
October 2012
Drinking From A Fire Hose: Real-time Data And Transportation Decision-making

Hani S. Mahmassani
The University of Texas at Austin

UCTC Student Conference, Irvine, CA
February 2001
<table>
<thead>
<tr>
<th>CONVENTIONAL WORLD</th>
<th>ITS ENVIRONMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Steady - state</td>
<td>– Time varying</td>
</tr>
<tr>
<td>– Equilibrium</td>
<td>– Evolutionary paths</td>
</tr>
<tr>
<td>– Static</td>
<td>– Dynamic</td>
</tr>
<tr>
<td>– Data poor</td>
<td>– Data rich</td>
</tr>
<tr>
<td>– Uncertainty about past/current events</td>
<td>– Known past/current events (to varying degrees)</td>
</tr>
<tr>
<td>– Component level</td>
<td>– System level</td>
</tr>
<tr>
<td>– Long lead time between solution and implementation</td>
<td>– Immediate action</td>
</tr>
<tr>
<td>– Limited “accountability” of decisions</td>
<td>– Performance monitoring and feedback</td>
</tr>
<tr>
<td>– “A priori” solutions</td>
<td>– Real-time adaptive strategies</td>
</tr>
</tbody>
</table>
In Closing.....

• Danger of getting overwhelmed with vast amounts of real-time data (not seeing forest from the trees): essential role for database management and processing, and interactive query and visualization systems...

• Danger of over-relying on data as a panacea and substitute for seeking fundamental understanding of underlying processes

• Danger of confusing information with knowledge!
ACID TEST

How is more data allowing me to

Do things differently  (better- faster, cheaper, safer, higher impact, customer-pleasing…)

Do different things  (grow activities, revenue, improve image, employee retention…)

?
Convergence of location, sensing, telecommunication and automotive technologies for better transportation system safety, efficiency, and user convenience.
If I had a million dollars
We wouldn't have to eat Kraft Dinner
But we would eat Kraft Dinner
Of course we would, we’d just eat more

Barenaked Ladies, Lyrics for live version
Data Generators
Sources of information

PASSIVE SOURCES: Automated, No or Limited Human input

- Vehicle sensors (mechanical, safety, fuel)
- Location sensors
- Shipment sensors (location, tampering, ID)
- Inventory tracking (liquid/gas, units)
- Transaction data
  - Customer interactions (order details, fulfillment)
  - Supplier interactions
  - Shipment events
  - Vehicle events
Data Generators
Sources of information

**ACTIVE SOURCES: Requires Human Input**

- Status reports augmenting passive sensors
  Driver updates
- Agent notes about customers at time of sale
- Follow-up surveys for lost sales
- Follow-up post-service surveys to gauge satisfaction with service, competition, purchase determinants...
- Community ratings and social media (e.g. Rand McNally for truckers)
- Confidence levels with data items
Data Generators

Sources of information

**PASSIVE SOURCES: Automated, No or Limited Human input**

- Vehicle sensors (mechanical, safety, fuel)
- Location sensors
- Shipment sensors (location, tampering, ID)
- Inventory tracking (liquid/gas, units)
- Transaction data
  - Customer interactions (order details, fulfillment)
  - Supplier interactions
  - Shipment events
  - Vehicle events

- Fleet maintenance
- Fuel strategies

- Real-Time Visibility into Supply and demand chains
  - Fleet and shipment status
  - Carrier perspective

- Online Operational Control
- Service Planning
- Resource Optimization
- Demand Forecasting
Data Generators
Sources of information

**ACTIVE SOURCES: Requires Human Input**

- Status reports augmenting passive sensors
  Driver updates

- Agent notes about customers at time of sale

- Follow-up surveys for lost sales

- Follow-up post-service surveys to gauge satisfaction with service, competition, purchase determinants...

- Community ratings and social media (e.g. Rand McNally for truckers)

- Confidence levels with data items
THE FREIGHT DATA FLOW CYCLE

Immediate ACTION:
Process fulfillment

PREDICTIVE ANALYTICS:
REAL-TIME Operational Processes

BUSINESS INTELLIGENCE
Supports multiple levels of mgmt decision-making

PREDICTIVE ANALYTICS:
STRATEGIC PLANNING Processes Decision support

Creation
Filtering/Tagging/Initial processing

Aggregation

WAREHOUSING

Death/delete
Data comes with bias

1. **Selectivity bias**: observe only one portion of the population distribution, those who have made a certain choice or opted to self-report ("choice-based" sample)
   - Observe only own customers (not potential ones, those who are using the competition)
   - Only winning bids, not lost sales
   - Complaints, not kudos

Solution: Invest in better data collection; Apply correction in deriving insight and developing predictive tools
Data comes with bias

2. **Random incidence bias**: *distorts* target population experience measures

- On-time performance statistics (without corresponding information on impacted shipments or customers)
- Service quality indicators (operator perspective vs. user perspective)
- Combining off-peak with peak period data

**Solution**: Be aware of distortion (insidious); Associate weight to properly recover target population stats and indicators
Data comes with bias

3. **Strategic bias**: responses may reflect only what respondent wants you to know (generally for own strategic advantage)
   - Public listings of
   - Status self-reports re: available carrier capacity
   - “Other” price offers

**Solution:**
- Incentivize respondent; Seek independent source or benchmark
- Correct for bias and exploit added strategic value through strategy formulation
“Today the data we have available to make predictions has grown almost unimaginably large: it represents 2.5 quintillion bytes of data each day, enough zeros and ones to fill a billion books of 10 million pages each. Our ability to tease the signal from the noise has not grown nearly as fast. As a result, we have plenty of data but lack the ability to extract truth from it and to build models that accurately predict the future that data portends.”
Why So Many Predictions Fail — but Some Don’t

“we are fooled into thinking that random patterns are meaningful; we build models that are far more sensitive to our initial assumptions than we realize; we make approximations that are cruder than we realize; we focus on what is easiest to measure rather than on what is important; we are overconfident; we build models that rely too heavily on statistics, without enough theoretical understanding; and we unconsciously let biases based on expectation or self-interest affect our analysis.”
THREE BIG STRATEGIC THEMES ENABLED BY DATA AND ANALYTICS

• **Integration**
  – Transportation with logistics processes (sourcing, manufacturing, warehousing, distribution)
  – Supply and demand chains

• **Collaboration: Collaborative Logistics**
  – Information sharing
  – Improve efficiency by reducing dead miles
  – Coupled with online information

• **Dynamic (Real-time) Information and Optimization**
  – Visibility through supply chain at all times
  – Tracking technologies
  – Reduce inefficiency, exploit opportunities
Takeaways

• Volatile environment place higher premium on good data
• Explosion of data sources from freight processes (various sensors, transaction logs), online surveys, and external sources (third parties, social media): challenge of extracting simplicity from complexity
• Data is power—critical competitive asset; leaders will exploit opportunities, laggards will increasingly fall behind
• Business Intelligence and Analytics—broad umbrellas for creating information and knowledge from big data masses
• More data and greater data dependence creates vulnerabilities: bias, security, obsolescence
• Three continuing big themes enabled by data: Integration, Collaboration, Dynamic (real-time) Optimization
• Make data your friend