

# Challenges and Solutions in Data Analysis and Visualization

Dan Haught  
FMS Advanced Systems Group  
&  
Ed Waigand  
FBI

# Welcome!

- ◆ 9:00 – 9:30 Breakfast and networking
- ◆ 9:30 – 10:45 Dan Haught,  
FMS Advanced Systems Group
- ◆ 10:45 – 11:00 Break
- ◆ 11:00 – 12:15 Ed Waigand,  
FBI NDEX
- ◆ 12:15 – 1:30 Lunch and networking

# Challenges and Solutions

- ◆ Let's talk about some hard problems
- ◆ and then look at ways to solve some of them
- ◆ and look forward to future solutions

# Data

## ◆ Problems

- There is no shortage of data. There is only a shortage of information.
- Tools on data is not so good, tools on information is good.
- Tower of Babel: language and text
- Tower of Babel: formats
- Connectivity and security
- Disambiguation and de-cloaking

# Data

- ◆ There is no shortage of data. There is only a shortage of information.
- ◆ Solutions
  - intelligent, automated grading
    - ◆ statistical/temporal/geospatial/source gives relevancy
  - machine understanding of metadata
    - ◆ metadata means something: let's train the machine
  - communities of interest and targeting
    - ◆ generating hairballs is so 1990's



# Data

- ◆ Tower of Babel: language and text
- ◆ Solutions
  - machine translation
    - ◆ gives us a head-start on the fire hose
  - entity extraction
    - ◆ the next step, turning text into things
  - linguistic analysis
    - ◆ getting entities *and* relationships

# Data

- ◆ **Connectivity and security**
- ◆ **Solutions**
  - political and organizational mindset change
    - ◆ happens slowly, builds with trust
  - workable security controls
    - ◆ A-Space as an example
  - reaching across classifications
    - ◆ if a asks for b, but a can't see b, should c know?

# Data

- ◆ Tower of Babel: formats
- ◆ Solutions
  - ETL
    - ◆ tools to build repeatable load processes
  - federated queries
    - ◆ don't import, ask for in transient form
  - metadata models and standards
    - ◆ ongoing work to establish the standards



# Data

- ◆ Disambiguation and De-Cloaking
- ◆ Solutions
  - take a cue from the commercial world
    - ◆ list de-duping
  - metadata comparison
    - ◆ does bob's ssn match robert's ssn?
  - deep linguistic name technology
  - de-cloaking dark network players
    - ◆ network analysis, structural similarity, et al.

# Visualization

## ◆ Problems

- more information = less perception
- hiding metadata from the machine
- context switching causes problems
- poor integration between actual and hypothetical arenas

# Visualization

- ◆ **More information = less perception**
- ◆ Solutions
  - smart layouts
  - dynamic filters
  - centrality ID through Social Network Analysis
  - time is the key
  - next-gen path analysis
  - show me dead-ends before I go there

# Visualization

- ◆ Hiding metadata from the machine
- ◆ Solutions
  - attach weights to metadata
  - entity type, relationship type, status, etc.
  - balance SNA and other analytics with metadata
  - time = money (or accuracy)
  - result = radically increased accuracy

# Visualization

- ◆ Context switching causes problems
- ◆ Solutions
  - don't switch between apps, integrate them
  - don't switch between windows, integrate them
  - back, next and breadcrumbs
  - automatic post-its for the brain



# Visualization

- ◆ Poor integration between actual and hypothetical arenas
- ◆ Solutions
  - applications need to support both seamlessly
  - mix 'real' data with hypothetical: let the system track it for you
  - create a workable promote/demote model

# Next?

- ◆ As we build systems:
  - we solve problems
  - we iterate
  - we learn our lessons
  - our users demand more
  - the problems get harder

# Future

- Cognitive systems that feed visualizations
- Systems we teach and that learn on their own
- Immersive environments
- Prediction and accuracy
- New visual metaphors
- What else?



Q&A