



Characterizing the Intelligence Analysis Process: Informing Visual Analytics Design through a Longitudinal Field Study



Youn-ah Kang and John Stasko
Georgia Institute of Technology

Visual Analytics

Business

Natural Science

Medicine

Finance

Social Science

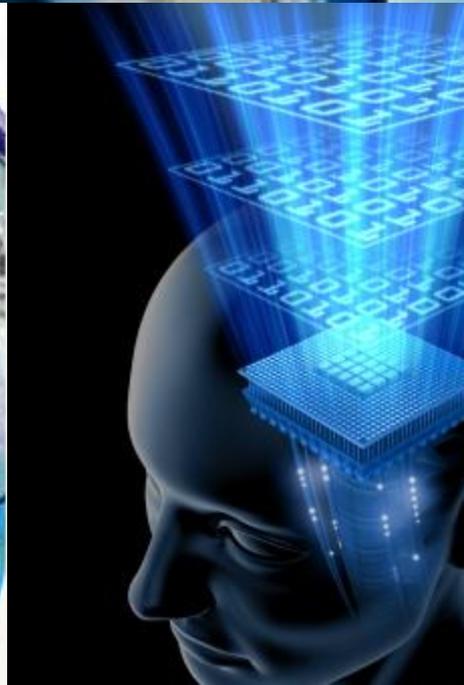
...

National Security

Law Enforcement

Criminal Analysis

Intelligence analysis



Background

Lessons from HCI

Explore

Understand users
and their problems

**System
Development**

Evaluate

Assess the utility of
the tool and refine

Current Status

Explore

**System
Development**

Evaluate





Mercyhurst
COLLEGE

Institute for Intelligence Studies



- One of the most prestigious programs in Intelligence Studies
- 300+ future intelligence analysts in government agencies and private sectors
- Discussion with Prof. Kris Wheaton
 - Inaccurate models of the intelligence analysis process
 - Little understanding of how intelligence is produced
 - Issues in adoption because of the difficulty of integration

Intelligence analysis practices

Knowledge Gap

Visual analytics systems



Our Approach

- Longitudinal field study
- Analysts-in-training at Mercyhurst
- 10-week intelligence project



VAST
2011

Department of Intelligence Studies

Are you curious?

Intelligence studies refers to the collection, the analysis and the counterintelligence and covert action operations that are necessary to maintain the country's ever-changing national security needs. At Mercyhurst, we're proud to offer one of the most well-regarded intelligence studies programs in the nation for both undergraduate and graduate students.

Goals of the Study

- What they do, challenges, and how we can help
- Processes and methods of intelligence analysis
- Reflect beliefs about their processes
- Draw design implications

Study Protocol

- An intelligence analysis project as part of a class
- Ten-week course, studied two separate semesters
- Three teams of 4-5 students (one team in fall, two teams in spring)
- Real intelligence problem with external clients
- Project status and progress on the wiki pages



Predictive
Analytics
TOP

*Technologies,
Organizations,
People*

★ home

PAGE ▾

DISCUSSION

HISTORY

NOTIFY ME

-  Wiki Home
-  New Page
-  Recent Changes
-  Manage Wiki

Home

[Table Of Contents](#)

Key Findings

Key Players

- [Technologies](#)
- [Organizations](#)
- [People](#)

Methodologies

- [Multi-Criteria Intelligence Matrix \(MCIM\)](#)

Mercyhurst College Institute for Intelligence Studies

Technologies, Organizations, and People (TOP) of the Predictive Analytic Industry

Requirement:

Who are the key people, technologies and organizations that likely currently have or will develop the potential to disrupt or replace traditional US national security Intelligence Community (IC) analytic work flows and products with commercially available products available over the next 24 months?

Criteria that will be used to identify these key players are:

- Those that are not beholden to the IC or US Government as primary sources of funding
- Those that have the potential to solve IC-like analytic problems,
- Those that are looking at future based events or actions that are outside the control of the forecaster/predictor
- Those that are focused on forecasting or predicting future courses of action

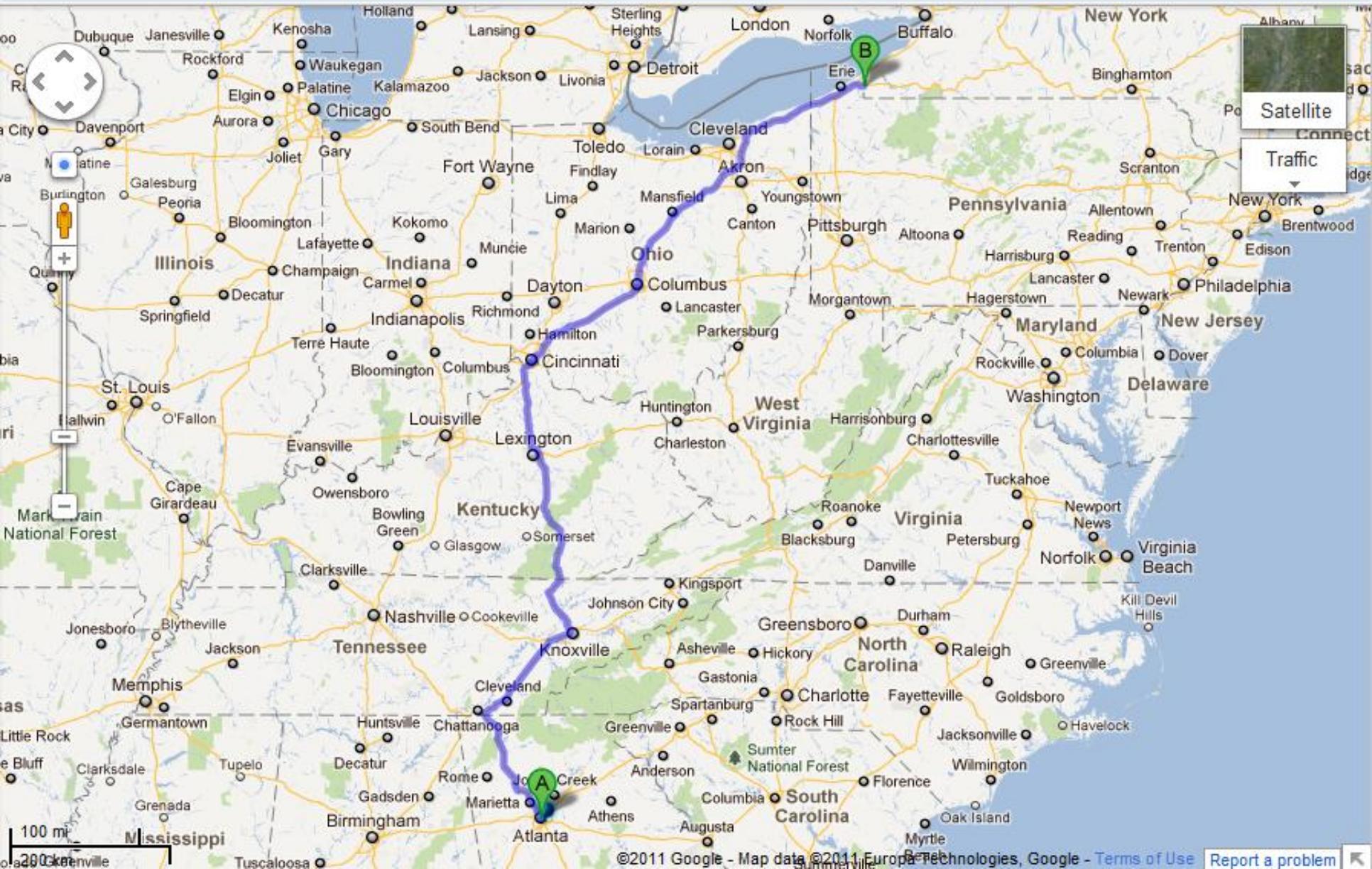
Methodology:

Analysts used open-source resources to construct a web-based Wiki, which includes any or all of the following: key findings, discussions, notes, res spreadsheets and interactive maps.

- Methods for analysis may include but are not limited to:
 - [Geographic Analysis](#)



Atlanta, GA -> Erie, PA (800miles)



Study Protocol

- Two face-to-face interviews with each team
- Observations on team meetings



Intelligence Task

- Strategic intelligence tasks – exploratory, long term
- “Relevant and important questions to the organization's success or failure but outside their control”

Examples:

Who are the key people, technologies and organizations that likely currently have or will develop the potential to disrupt or replace traditional US national security Intelligence Community (IC) analytic work flows and products with commercially available products?

What are the most consistent and identifiable characteristics displayed by potential insider threats to (a defense department)?

Overall Analysis Components

2. Collection

distance learning



Advanced search

About 70,900,000 results (0.18 seconds)

[Distance Learning Online | Phoenix.edu](#)

Ads

www.phoenix.edu

University of Phoenix® - Earn Your Degree from Home. Flexible Courses.
Business Degrees - Technology Degrees - Psychology Degrees

[DeVry Distance Learning | devry.edu](#)

www.devry.edu

Start Your **Distance** Degree Program Now at DeVry.

[We Are. Distance Learning | WorldCampus.PSU.edu](#)

www.worldcampus.psu.edu

A Recognized Leader in **Distance Learning** - Penn State University.

[Scholarly articles for distance learning](#)



[... and the online classroom: Vol. 3. Distance learning](#) - Berge - Cited by 545

[Issues in distance learning](#) - Sherry - Cited by 654

[The Effects of Distance Learning: A Summary of ...](#) - Moore - Cited by 305

[Distance education - Wikipedia, the free encyclopedia](#)

en.wikipedia.org/wiki/Distance_education

Distance education or **distance learning** is a field of education that focuses on teaching methods and technology with the aim of delivering teaching, often on an ...

History and development - Technologies used in delivery

[Top Distance Learning Universities and Online Degrees | Earn Your ...](#)

www.distancelearningu.com/

Sep 5, 2011 – Research top **distance learning** universities and earn your online degree from your home by enrolling in **distance learning** courses that best fits ...

- “All-source intelligence”
- Push & pull sources
 - Internet search (pulled)
 - RSS news feeds (pushed)

3. Analysis

“Information to knowledge”

MCIM : Entities

Entity Name	Unstructured Data	Unreliable Data	Incomplete Data	Detect Deception in Data	Data Extracted	Application Compatibility	Impact Score* (see Impact Matrix)	Total Score
Text Analytics	3	0	3	1	3	3	3	16
Dependency or Association Analysis	0	2	2	2	0	3	3	12
Network Analysis	0	2	2	2	0	3	3	12
Data profiling and transformations	2	2	1	2	0	3	1	11
Prescriptive Analytics	0	2	1	2	0	3	3	11
Sequential Pattern Analysis	2	0	2	0	2	3	2	11
Bayesian Analytics	0	2	0	2	0	3	3	10
Clustering or Segmentation	0	2	0	2	0	3	2	9
Optimization Models	0	1	1	1	0	3	3	9
Game theory	0	1	1	0	0	3	3	8
Regression Analysis	0	0	0	2	0	3	3	8
Basket Analysis	0	0	2	0	0	3	2	7
Classification	0	0	0	1	0	3	3	7
Simulations	0	0	1	0	0	2	3	6
Time series forecasting	0	0	0	0	0	3	3	6
Time series tracking	0	0	0	0	0	2	2	4
Behavior Learning Software-Machine Learning								
Behavior Learning Software-Machine Learning	3	2	3	2	3	3	3	19
Neural Networking	3	2	3	3	3	2	2	18
Complex Event Processing	3	2	2	3	2	2	2	16
Statistical Package for the Social Sciences (SPSS)	3	1	2	2	2	3	3	16
Temporal Analytics	3	1	3	1	3	3	2	16
Reality Mining	3	1	3	1	1	3	3	15

4. Production

Predictive Analytic Industry [Social Network Analysis](#)

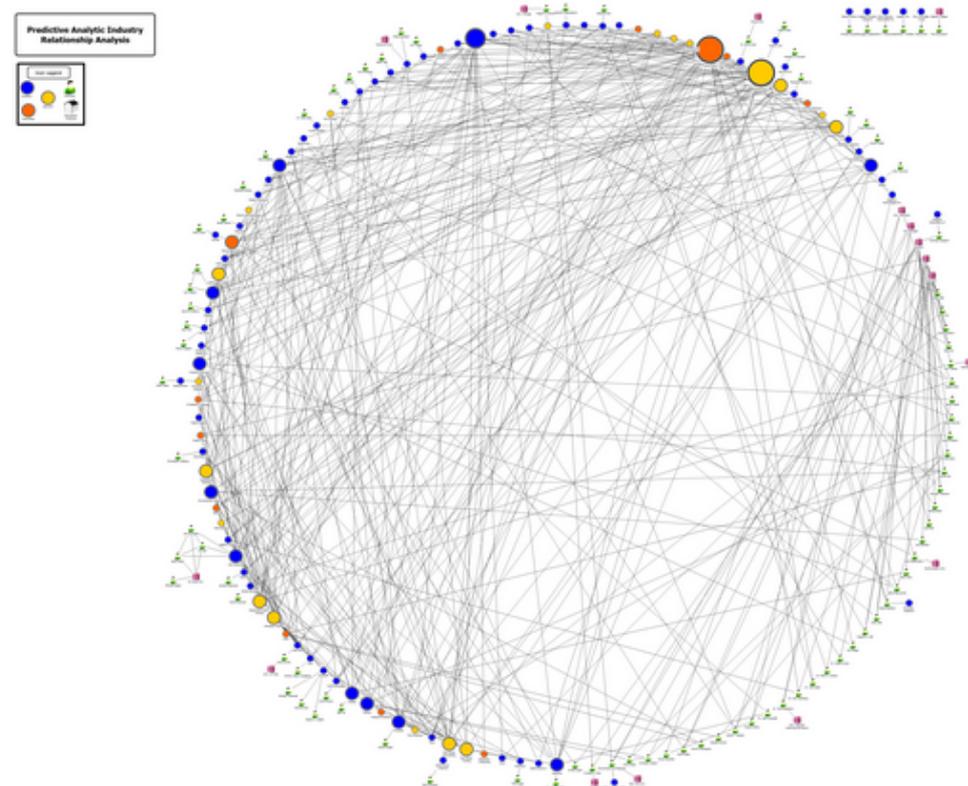
- It is highly likely that there are entities that have the potential to disrupt or replace IC work flows and products due to their relationship to key technologies, organizations, and/or people within the industry.
- Entities were placed into a network using i2 Analyst's Notebook. Links were established among any entities that were found as being related to each other through functionality, business, production, or application. The purpose of the Industry [social network analysis](#) was to highlight those entities that have the potential to possess the most influence or disruption within the predictive analytics industry. The analysts found that there were entities that did not score high on the [Multi-Criteria Intelligence Matrix](#), but had high industry influence.

- **Top 10 Betweenness Scores:**

1. Clustering or Segmentation (18.33)
2. Data Mining (18.12)
3. Simulation Models (8.80)
4. [SAS Analytics](#) (7.71)
5. Classification (6.61)
6. [Neural Networking](#) (6.07)
7. [IBM](#) (5.55)
8. Time Series Forecasting (5.54)
9. Apache (4.56)
10. Microsoft (4.42)

- **Top 10 Eigenvector Scores:**

1. Data Mining (49.65)
2. Clustering or Segmentation (43.81)
3. [IBM](#) (36.04)
4. [SAS Analytics](#) (30.12)
5. Microsoft (29.75)
6. Classification (27.81)
7. [Oracle](#) (24.64)



Understanding the Process: Misconceptions

Understanding the Process

- Misconception #1: Intelligence analysis is about finding an answer to a problem via a sequential process.

=> The process is more parallel and organic.

“Intelligence analysis is not about getting from point A to point B along a travel route, but it is better associated with basic research where you don’t necessarily know where you are going to go.”

“Our conceptual model is changing. As it changes, we’re changing our analytic focus, making decisions about production, and that’s being informed by new information that comes in.”

Understanding the Process

- Misconception #2: The key part of the intelligence process is the analysis of a specific set of data.

=> The process of “constructing a frame” is more important.

“Intelligence is about determining how to answer a question, what to research, what to collect, and what criteria to use”

“Intell analysis is not like that you have a set of data in hand and run a program. It’s like a conundrum from the very beginning. You have to learn how to learn, how to frame the question, and how to answer it through collecting and evaluating sources.”

Understanding the Process

- Misconception #3: Analysts do not often collaborate.

=> Multiple layers of collaboration exist in intelligence analysis and the degree of collaboration differs depending on the type of task and the group dynamics.

“Zotero is a good example of one way we collaborate. Each person creates a group library on the Zotero server. If I find a website that I think is useful, whether for my topic or someone else’s topic, I add it to our group collection, and then other members can see it before they go searching the Internet.”

Understanding the Process

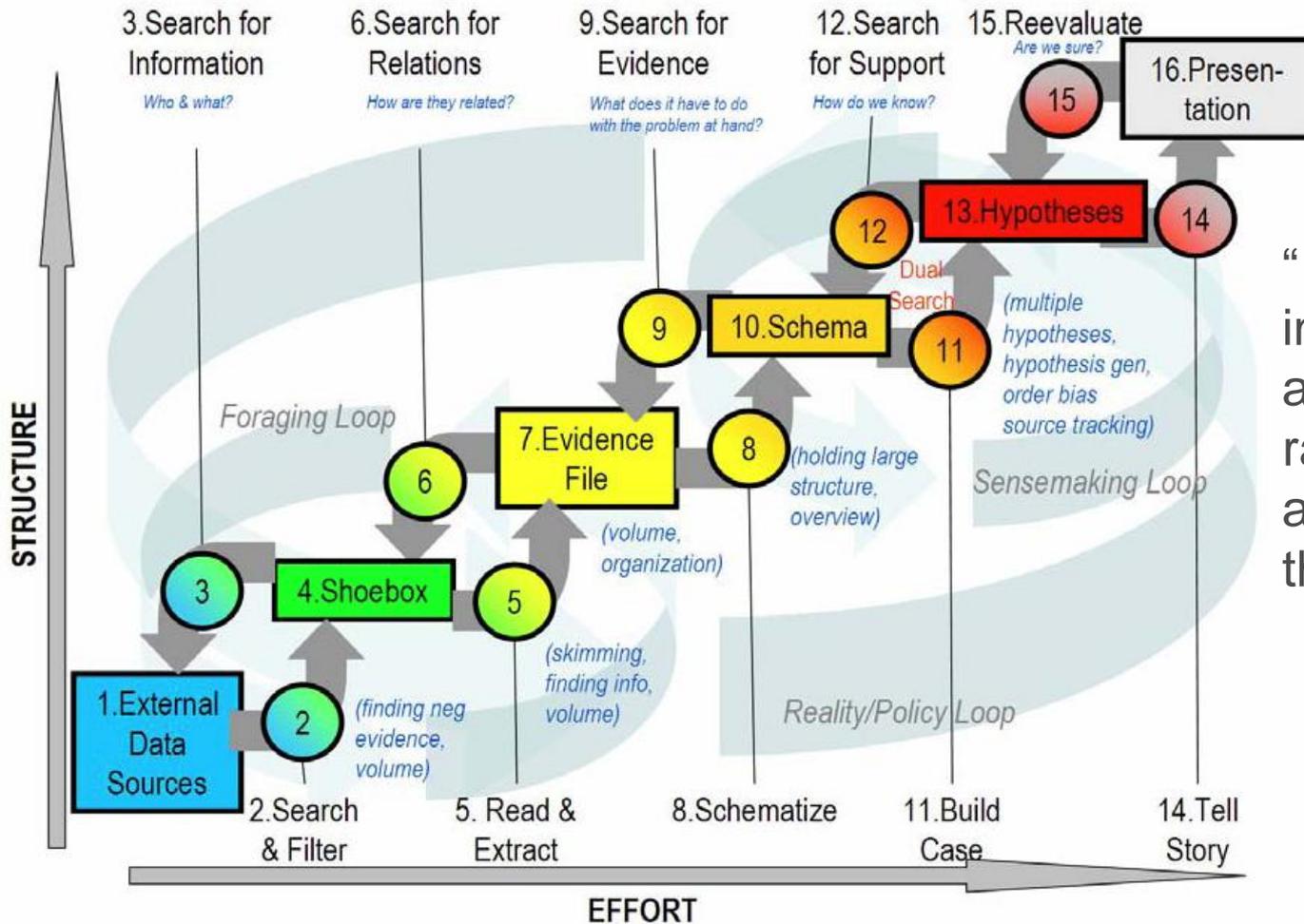
- Misconception #4: We can help intelligence analysts by developing sophisticated analytic tools that support specific types of analysis.

=> Analysts want an integrated system that can manage the intelligence process effectively and employ various analytical methods and tools quickly.

“Everything is fragmented...But there isn’t any set that ties all these. The pieces are here and there, but nothing links all that in one seamless thing so I can go from the requirement to a product in a single package, in a single way.”

Rethinking the intelligence analysis process - Linear vs. Parallel

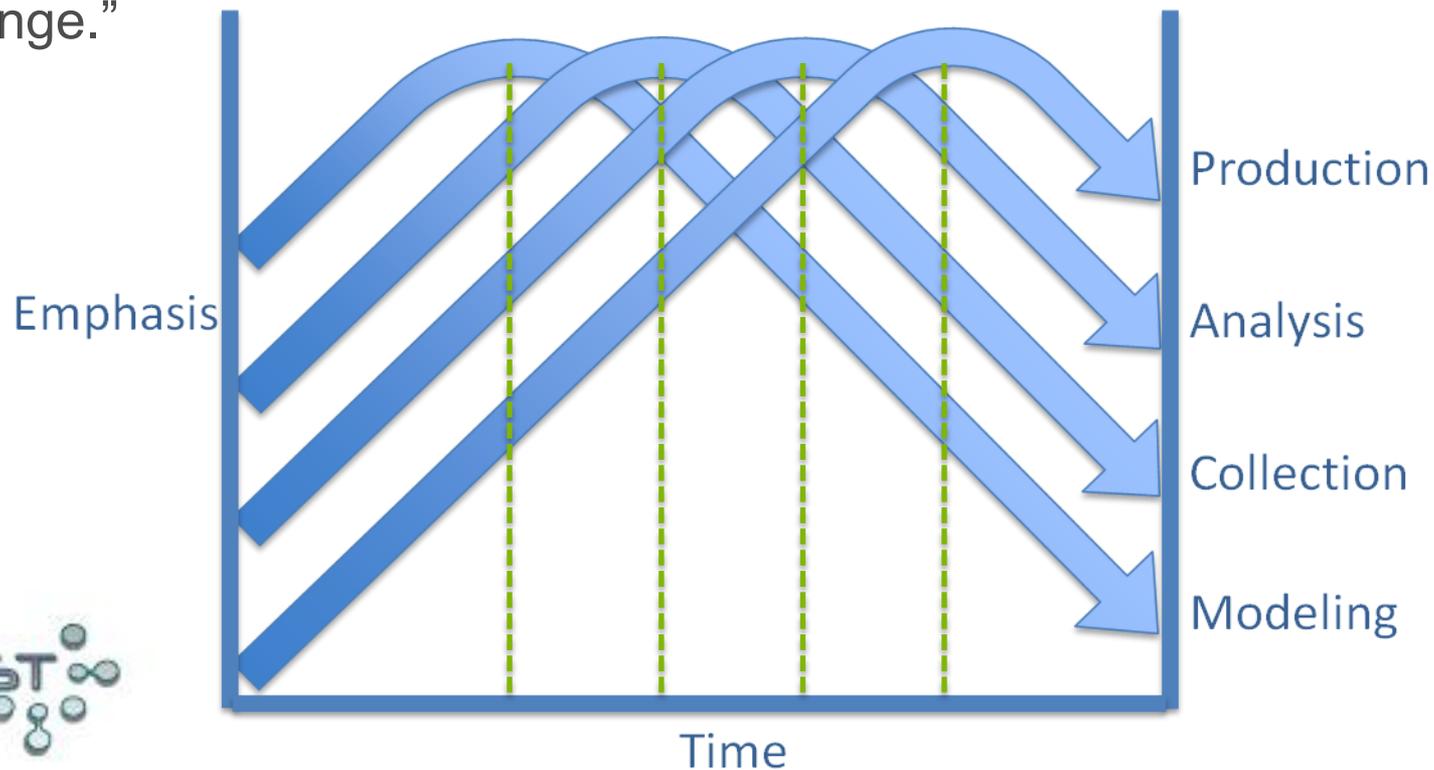
Pirolli and Card's Sensemaking Model



“... describes how information transforms and how data flows, rather than how analysts work and how they transition.”

Rethinking the intelligence analysis process - Linear vs. Parallel

- Wheaton's Multi-phasic model of the process
“All four functions begin almost immediately, but through the course of the project, the amount of time spent focused on each function will change.”



How Visual Analytics can Help: Design Implications

- Externalize the thinking (conceptual modeling) process
- Support source management
- Support analysis with constantly changing information – seamless transition between collection and analysis
- Help analysts create convincing production - support insight provenance and sanity checks
- Support asynchronous collaboration rather than synchronous collaboration for exploratory analysis

Conclusions

- Described an empirical study to understand intelligence analysts and their processes
- Documented the processes and methods they followed
- Clarified issues regarding the intelligence process
- Identified design implications for visual analytics systems for intelligence analysis

Acknowledgements

Professor Kristan Wheaton, Department of Intelligence Studies, Mercyhurst College for his support and input on this research.

This work is supported by the National Science Foundation and the VACCINE Center, a Department of Homeland Security's Center of Excellence in Command, Control and Interoperability.



National Science Foundation
WHERE DISCOVERIES BEGIN

