
Visual Analytics for Cyber Security: Observations and Opportunities for Large Display Workspaces

Endert, A., Fink, G. A., North, C.



Overview

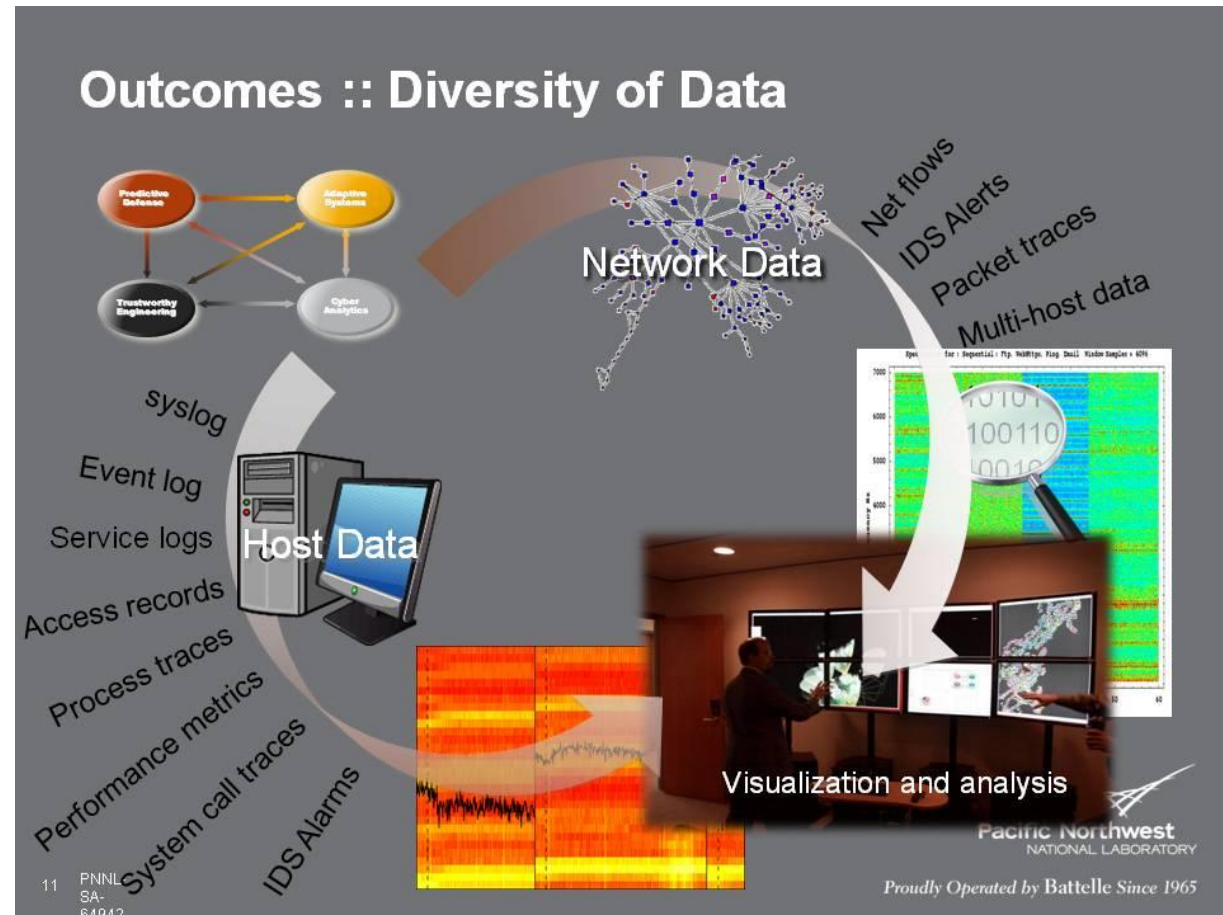
- Project Background
 - Domain Description
 - Studies Overview
- Findings
- Prototypes & Analyst Feedback



Background

- How can we design **Visual Workspaces** that aid **Cyber Security**?

- Monitor networks for intrusions
- Analyze network logs, process logs, email logs, etc.
- Tons of data?
- Lots of windows and tools?



Large, High-Resolution Displays

- (8) 30-inch high-res LCD Panels
- 33 Megapixel total resolution (10,240 x 3,200)
- “Single PC” Architecture
- Curved for optimal individual use



Methods

1. Semi-Structured Contextual Interviews (8 professional cyber analysts)

- Typical tasks and data?
- Work style?
 - E.g., Collaboration? Multi-tasking? Time constraints?
- Office setup
- What does your finished analysis product contain?
- **Constant Contact with analysts (>2 months, daily)**

2. Observational Lab Study (4 cyber analysts, VAST09 dataset)

- 2 sources of data: Building/room access records (Prox) and simulated computer network flows
 - HINT: making connections between the sources is key! 😊
- Tools provided: Excel, Spotfire, Windows XP

3. Feedback and Further Investigation/Analysis

- 4 cyber analysts at VT, 3 at PNNL, 1 manager

Lab Study

- Confirm general findings from contextual interviews.



Ethnographic
Study

“Controlled
Ethnography”

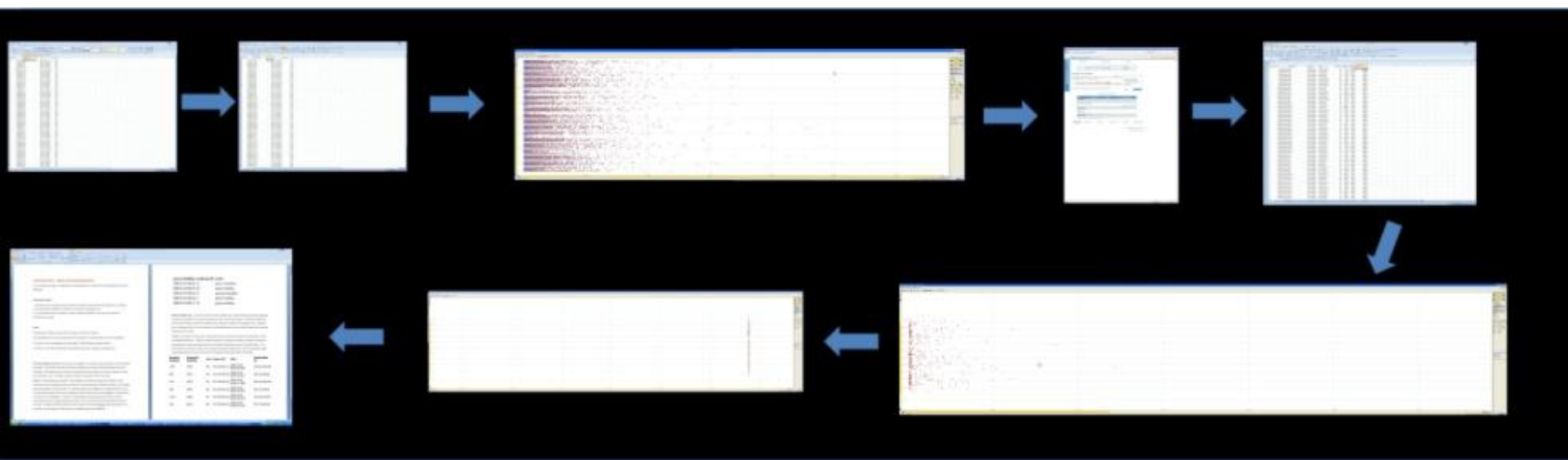
Lab Study

Key “Ethnographic” Discoveries

1. Data sources reside in separate tools
2. Analysts spend much time doing low-level tasks
3. They distrust visualizations
4. They are on a “Quest for a Query”
5. Cyber data comes in huge volumes and velocities
6. Cyber data comes from many diverse sources
7. Analysts seek direct access to the data
8. Analysts routinely conduct a large number of tasks in parallel (multi-tasking)

1. Data Resides in Different Tools

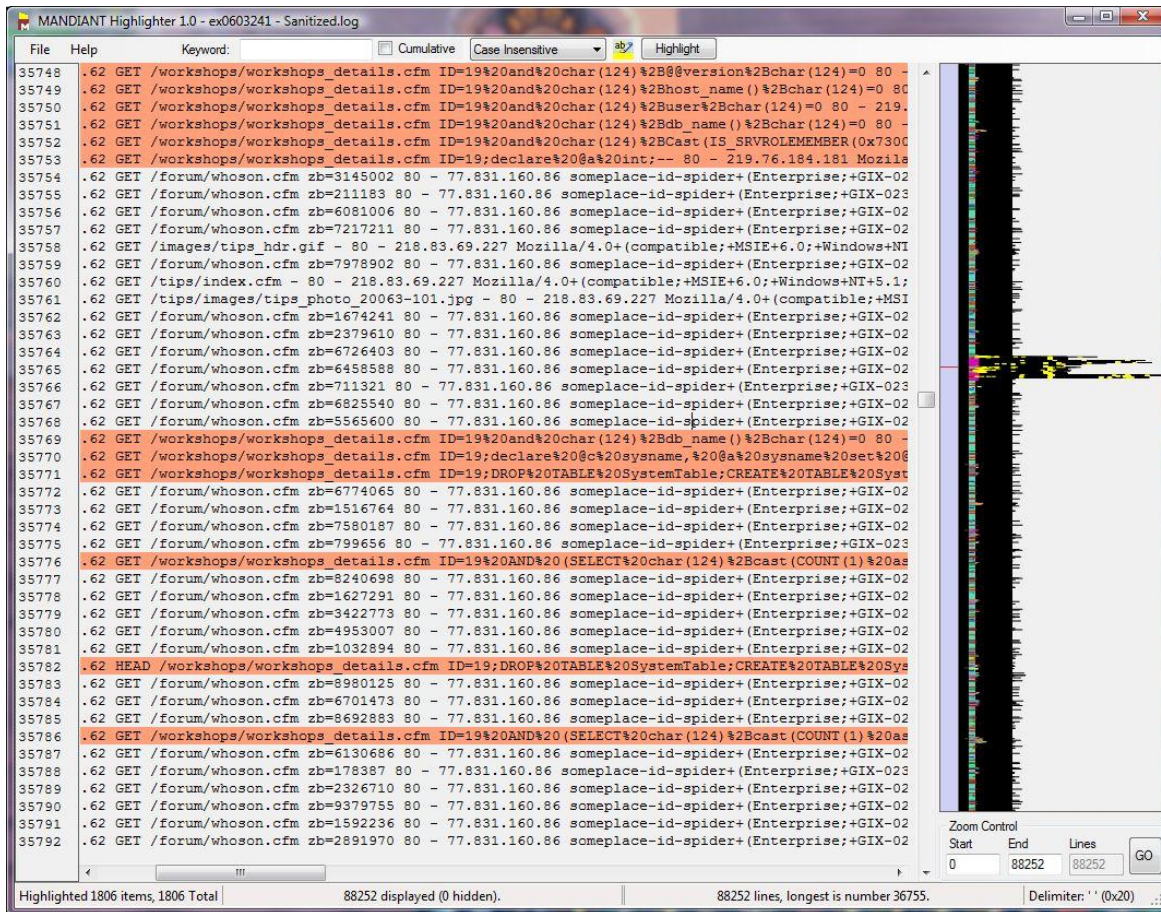
- Used space for visual path



- ▶ Rote mechanical process
 - ▶ Analyst: “Tedious!”

2. Low-level Tasks

- Analysts filter out the “normal”
 - line-by-line
- Seek patterns of familiar abnormalities
 - Previous experience creates personal “hit list”
- Analysts observe data individually, not in connection with whole dataset



The screenshot shows the MANDIANT Highlighter 1.0 interface. The main window displays a log file with 1806 items highlighted in red. The log entries are structured as follows:

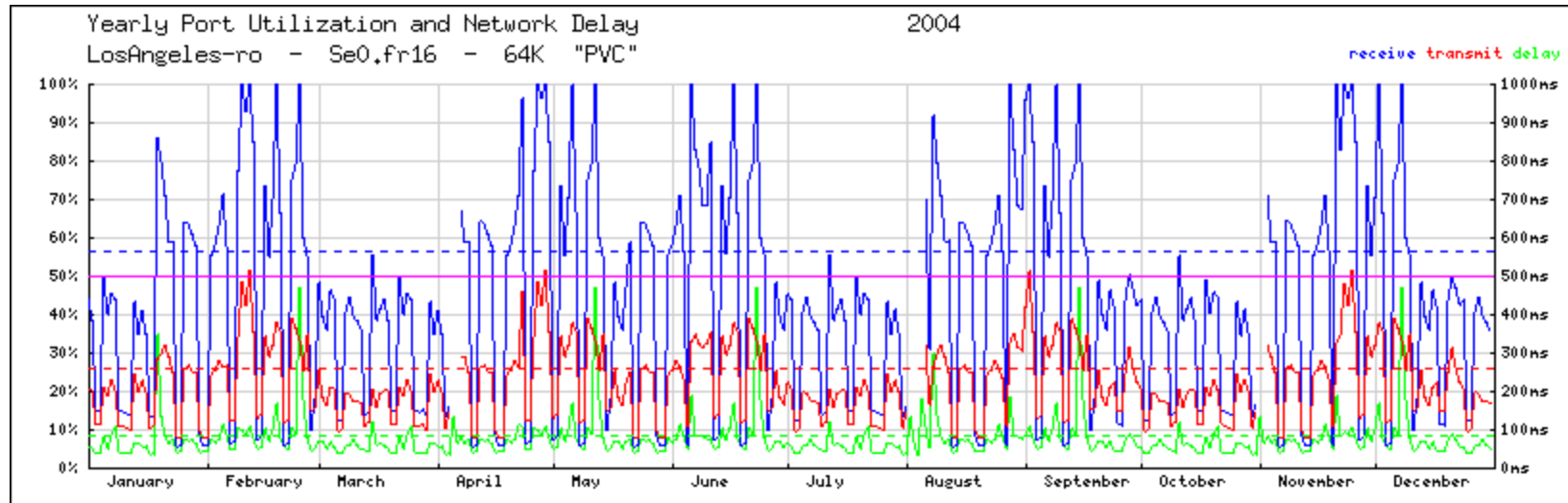
```
35748 .62 GET /workshops/workshops_details.cfm ID=19%20and%20char(124)%2B@@version%2Bchar(124)=0 80 -
35749 .62 GET /workshops/workshops_details.cfm ID=19%20and%20char(124)%2Bhost_name()%2Bchar(124)=0 80 -
35750 .62 GET /workshops/workshops_details.cfm ID=19%20and%20char(124)%2Buser%2Bchar(124)=0 80 - 219.
35751 .62 GET /workshops/workshops_details.cfm ID=19%20and%20char(124)%2Bdb_name()%2Bchar(124)=0 80 -
35752 .62 GET /workshops/workshops_details.cfm ID=19%20and%20char(124)%2BCast(IS_SRVROLEMEMBER(Ox7300
35753 .62 GET /workshops/workshops_details.cfm ID=19;declare%20@a%20int;-- 80 - 219.76.184.181 Mozilla
35754 .62 GET /forum/whoson.cfm zb=3145002 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35755 .62 GET /forum/whoson.cfm zb=211183 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-023
35756 .62 GET /forum/whoson.cfm zb=6081006 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35757 .62 GET /forum/whoson.cfm zb=7217211 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35758 .62 GET /images/tips_hdr.gif - 80 - 218.83.69.227 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT
35759 .62 GET /forum/whoson.cfm zb=7978902 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35760 .62 GET /tips/index.cfm - 80 - 218.83.69.227 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1;
35761 .62 GET /tips/images/tips_photo_20063-101.jpg - 80 - 218.83.69.227 Mozilla/4.0+(compatible;+MSI
35762 .62 GET /forum/whoson.cfm zb=1674241 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35763 .62 GET /forum/whoson.cfm zb=2379610 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35764 .62 GET /forum/whoson.cfm zb=6726403 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35765 .62 GET /forum/whoson.cfm zb=6458588 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35766 .62 GET /forum/whoson.cfm zb=711321 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-023
35767 .62 GET /forum/whoson.cfm zb=6825540 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35768 .62 GET /forum/whoson.cfm zb=5565600 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35769 .62 GET /workshops/workshops_details.cfm ID=19%20and%20char(124)%2Bdb_name()%2Bchar(124)=0 80 -
35770 .62 GET /workshops/workshops_details.cfm ID=19;declare%20@c%20sysname,%20@a%20sysname%20set%20@
35771 .62 GET /workshops/workshops_details.cfm ID=19;DROP%20TABLE%20SystemTable;CREATE%20TABLE%20Sys
35772 .62 GET /forum/whoson.cfm zb=6774065 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35773 .62 GET /forum/whoson.cfm zb=1516764 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35774 .62 GET /forum/whoson.cfm zb=7580187 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35775 .62 GET /forum/whoson.cfm zb=799656 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-023
35776 .62 GET /workshops/workshops_details.cfm ID=19%20AND%20(SELECT%20char(124)%2BCast(COUNT(1)%20as
35777 .62 GET /forum/whoson.cfm zb=8240698 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35778 .62 GET /forum/whoson.cfm zb=1627291 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35779 .62 GET /forum/whoson.cfm zb=3422773 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35780 .62 GET /forum/whoson.cfm zb=4953007 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35781 .62 GET /forum/whoson.cfm zb=1032894 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35782 .62 HEAD /workshops/workshops_details.cfm ID=19;DROP%20TABLE%20SystemTable;CREATE%20TABLE%20Sys
35783 .62 GET /forum/whoson.cfm zb=8980125 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35784 .62 GET /forum/whoson.cfm zb=6701473 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35785 .62 GET /forum/whoson.cfm zb=8692883 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35786 .62 GET /workshops/workshops_details.cfm ID=19%20AND%20(SELECT%20char(124)%2BCast(COUNT(1)%20as
35787 .62 GET /forum/whoson.cfm zb=6130686 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35788 .62 GET /forum/whoson.cfm zb=178387 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-023
35789 .62 GET /forum/whoson.cfm zb=2326710 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35790 .62 GET /forum/whoson.cfm zb=9379755 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35791 .62 GET /forum/whoson.cfm zb=1592236 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
35792 .62 GET /forum/whoson.cfm zb=2891970 80 - 77.831.160.86 someplace-id-spider+(Enterprise;+GIX-02
```

The interface includes a search bar at the top with the keyword "ab?". The status bar at the bottom indicates "Highlighted 1806 items, 1806 Total", "88252 displayed (0 hidden)", "88252 lines, longest is number 36755", and "Delimiter: '' (0x20)".

MANDIANT Highlighter

3. Distrust of Visualizations

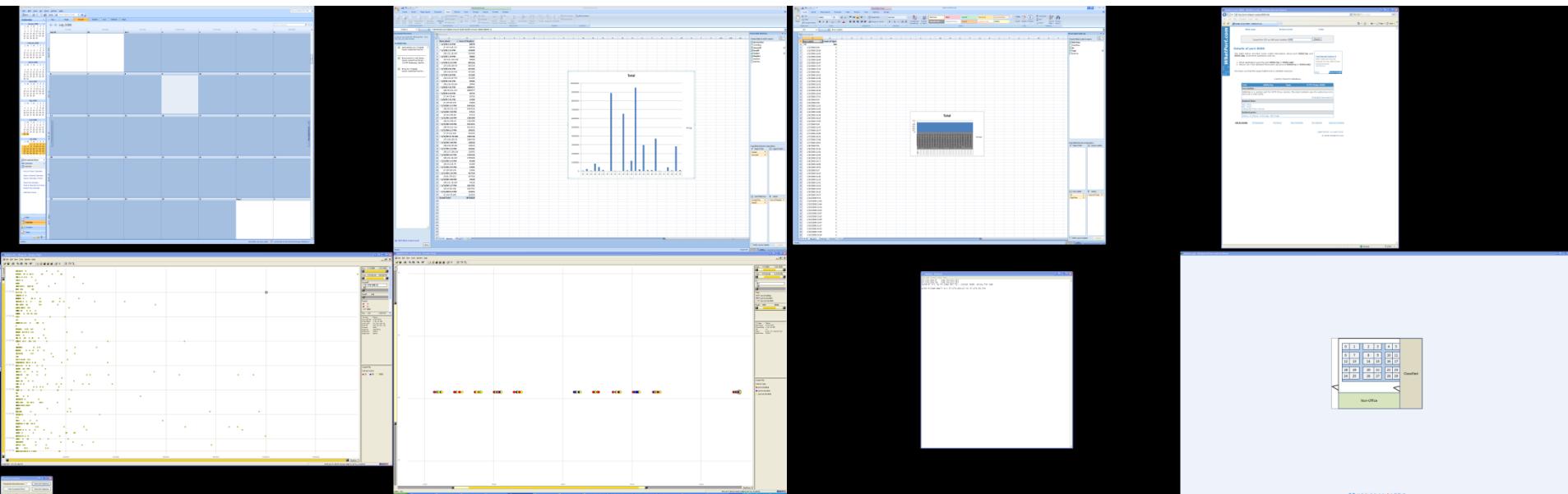
- Analyst: “Visualizations are in the way of the data”



- ▶ Visualizations:
 - ▶ May be too slow
 - ▶ May hide important, small details
- ▶ Analysts can only see, not *manipulate* the data

4. Quest for a “Query”

- Process (“quest”) of analysis results in:
 - Collection of suspicious data
 - Novel process of discovery
 - Formalized “SQL-style” query
- “Query” is the question that finds the answer you have
 - Cumulative result of *interaction* with variety of tools



► *Is this process “querifiable”?*

Intelligence vs. Cyber Analytics

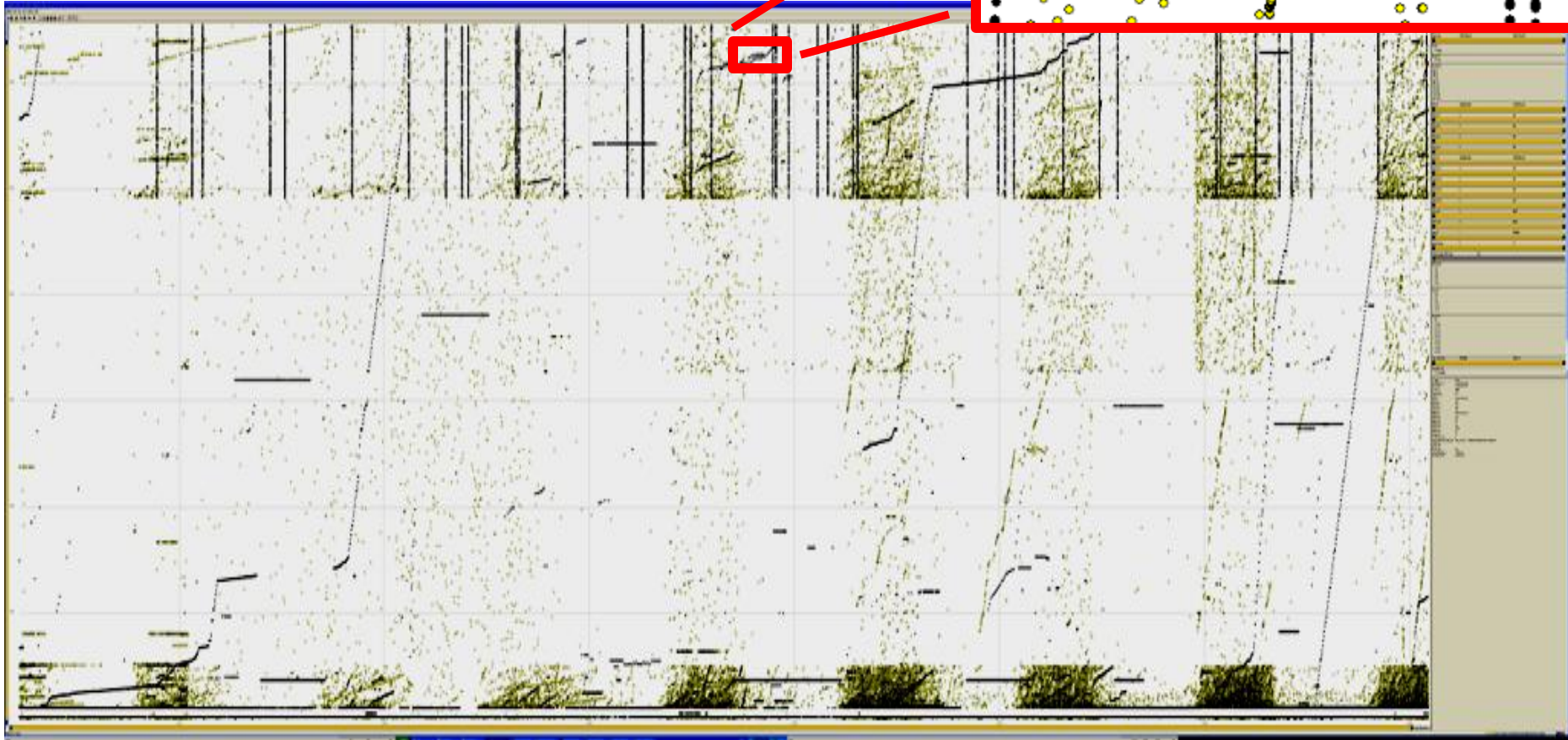
<u>Stegosaurus Scenario</u> (Intelligence Analytics)	<u>Cyber Security Scenario</u> (Cyber Analytics)
Creating a <i>story</i> about the threat. Product = story	Working on task generates process (at times “querifiable”) Product = query
Work done in a <i>visual space</i> . (Sensemaking Process)	Work done <i>within tools</i> . (Tools to Process the Data)
Rely on <i>Visualizations</i> .	Rely on <i>Linux Command Line</i> .
Un-, semi-, and structured data.	<i>Mainly</i> structured data. (packet, etc.)
Interactions <i>organize</i> the information spatially.	Interactions <i>filter</i> the information.
Information takes on personalized meaning (not “Excel” window, but “[x] data over here”)	Information maintains “version of file” or “Excel window” meaning to analyst

Large Display Opportunities: Prototypes

- Multi-scale Visualizations
- De-Aggregate Vital Information (monitoring)
- Support multiple, simultaneous investigation cases
- Provide history and traceability for investigations

Large, High-Resolution Visualization

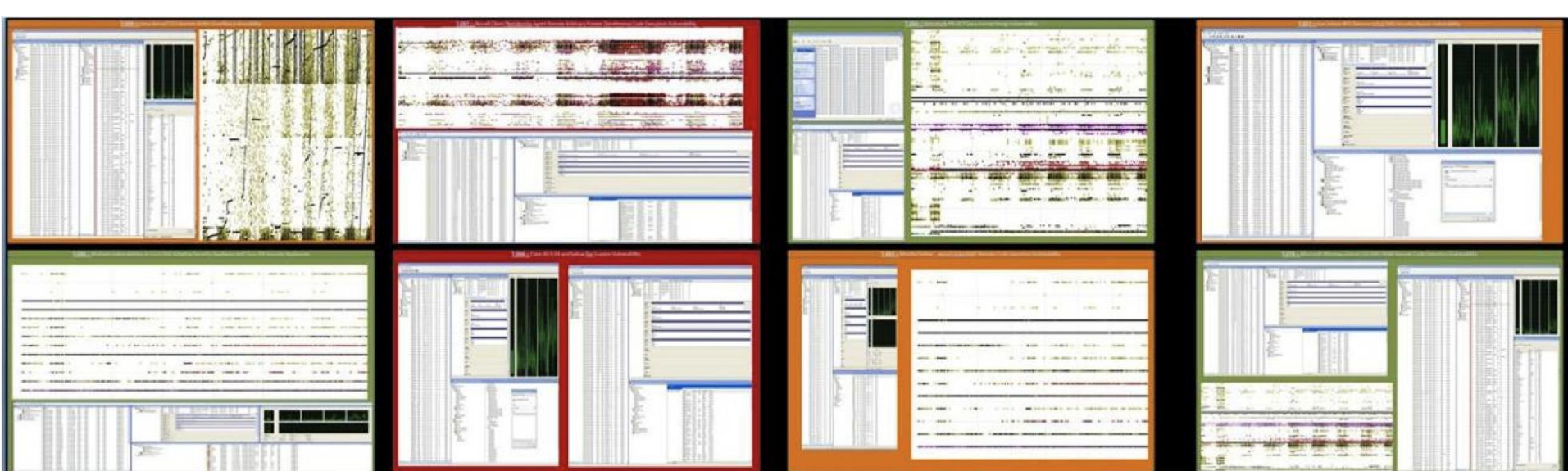
- Visibility of patterns at multiple scales
 - Provides overview *and* detail



De-Aggregate Vital Information

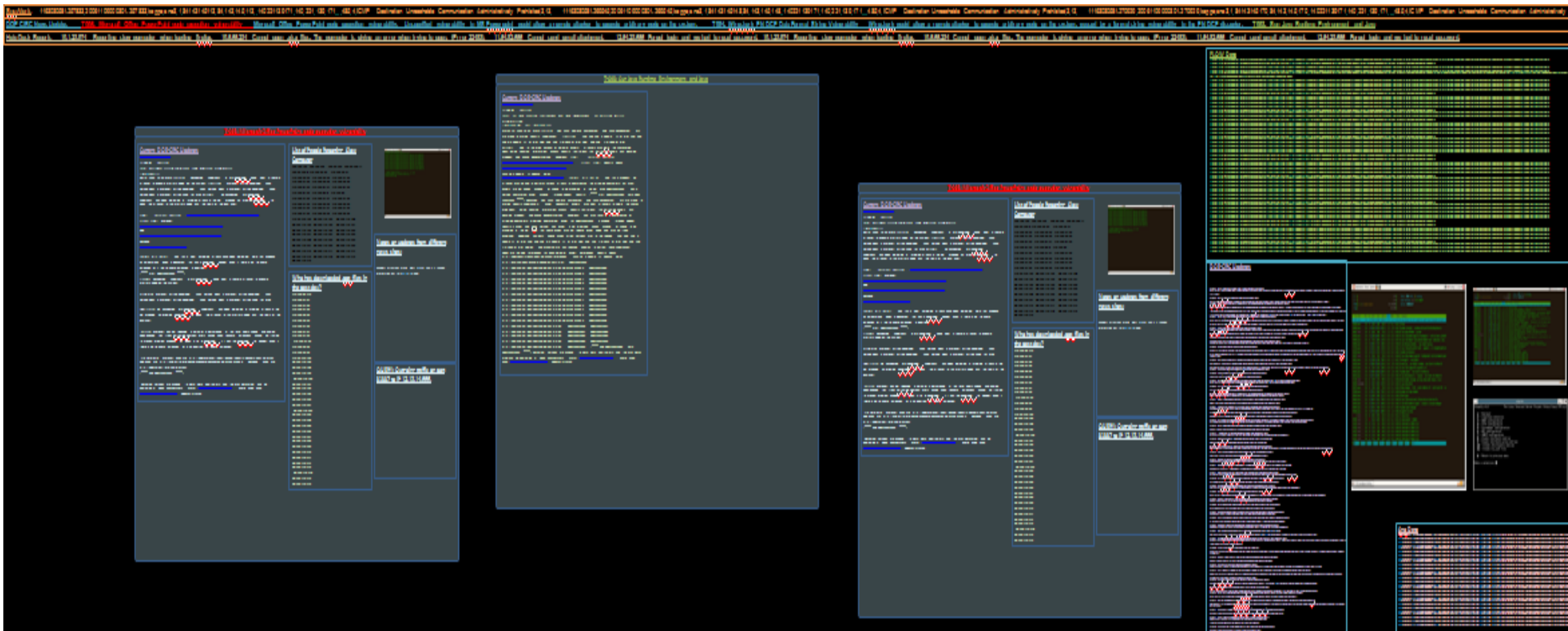


- Provides analyst with situational awareness
 - More upfront information, while maintaining overview
 - Less “interaction junk”



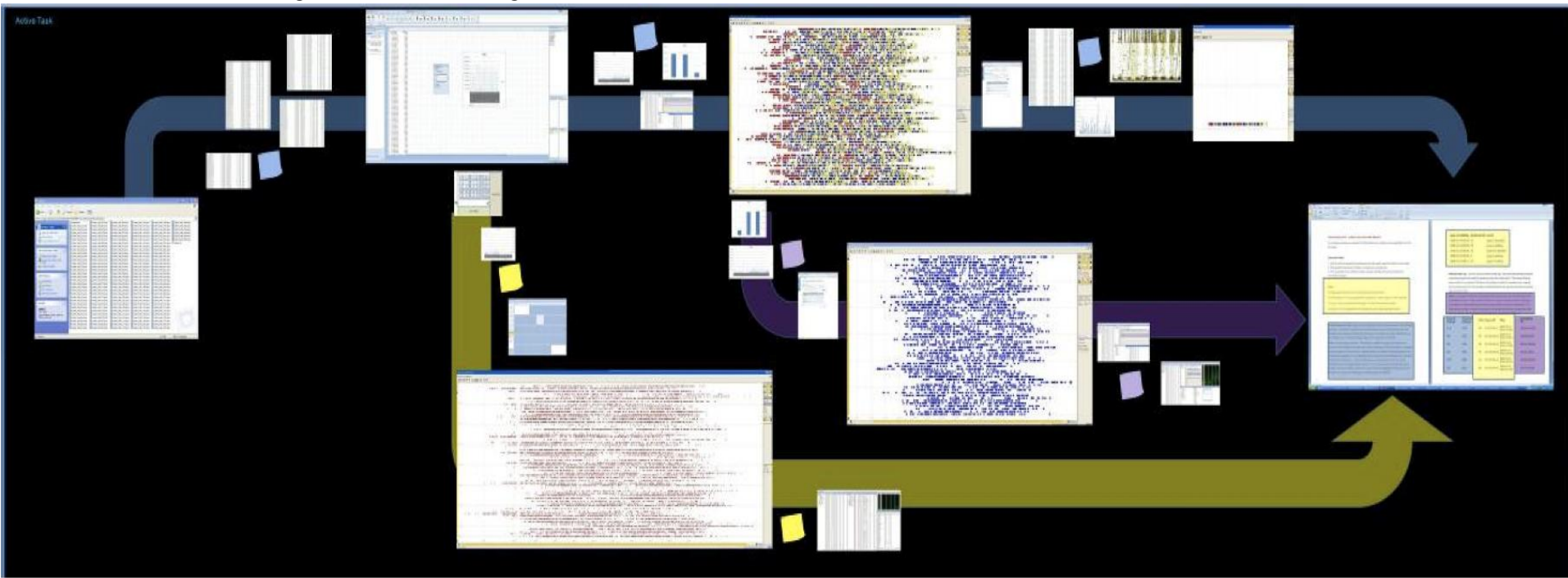
Multiple Simultaneous Cases

- Shows live data
 - Real time updating
- Analyst can set alerts for monitoring
- Enables collaboration by sharing cases

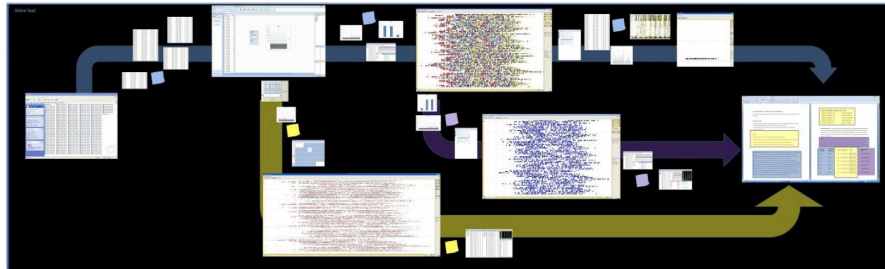


History and Traceability

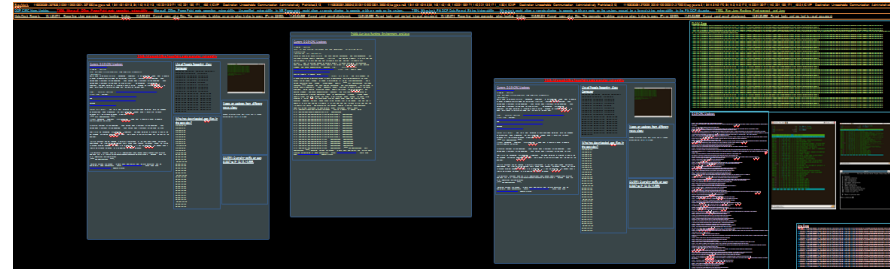
- “History Trees”: concept providing traceability and history of analyst’s workflow



A visualization should be the means for a user to interact and think.



History and Traceability



Multiple, Simultaneous Investigation cases

Large, High-Resolution Visualizations

De-Aggregate Vital Information

