INSTITUTE FOR MODERN INTELLIGENCE
2730 MONACAN ST. SUITE 102 ALEXANDRIA, VA 22314 WWW.IMINTEL.ORG

The Semiotic Apparatus - Fixing Intel and Operationalizing Data 2013

At a Glance

What it is

The Semiotic Apparatus is an Ultra-Large Scale systems infrastructure for data-intensive computing (Intel), and for data-intensive operations (C2).

How it works

- A lightweight ingest system takes all manner of data (traditional intel as well as human, social, cultural, behavioral) "as it lies" and throws it onto the cloud computing "floor"
- An amalgam of massively parallel processes continuously mine, exploit, and enrich data to lift out information and knowledge
- Data managers maintain and cultivate the information topology to optimize collection resources and make data more reliable and useful
- Knowledge engineers integrate and enrich data-models to support ever changing and diverse application and user perspectives
- Analysts explore, associate, and exploit data, information, and processing, creating new insights and solving hard problems
- Commanders and war-fighters employ new semiotic tools and applications to perform comprehensive mission SA, continuous agile assessment, planning, reporting, and execution

What it will achieve

The Semiotic Apparatus breaks the data barriers to put all of our data and computational intelligence assets in the service of the warfighter.

Intelligence is a massively complex endeavor. Some of that complexity is absolutely essential – the panoply of data sources and types and the various ways in which they are employed all contribute to the mission. At some fundamental level they can not be simplified, harmonized, or otherwise constrained because doing so would compromise their power and purpose and most of all their ability to evolve and innovate. Essential complexity may seem scary, but it is a good thing. If we cultivate it, it will bear astonishing fruit.

Unfortunately, there is also a great deal of complexity in intelligence that is accidental – it is bad, unnecessary, and hurts us. The physical, representational, structural, and semantic barriers between data sources, types, domains, and processes fall into this category. In the context of data search, analysis, exploitation and support to operations, they have no upside, they just gum up the works, wasting resources and impeding progress. "Fixing intel," as MG Flynn would have it, and making intelligence operationally relevant in any conflict whether it be force-on-force or counter insurgency, only becomes possible when this accidental complexity is eliminated.

The key innovation underlying the Semiotic Apparatus does just that. It dissolves a prodigious amount of information systems gunk to make possible the creation of not just another whiz-bang capability, but a diverse confederation of capabilities operating in congress upon an amalgamation of disparate data. It is a practical, Ultra-Large Scale systems solution for data unification, persistence, enrichment, and exploitation that accommodates the full diversity of data and semantics, including geospatial and temporal context, in a unified framework without information loss or distortion. It transforms the "data problem" into a rich Intel resource for deep analysis, rapid analytic innovation, and ad hoc mash-ups, while simplifying the flow and control of information. It enables all to see and exploit the full C2 / Intel context while focusing relevant information to those who need it as dynamic events unfold.

The Semiotic Apparatus is the machinery we need for fixing intel and operationalizing data.

Points of Contact

S. Yoakum-Stover, Ph.D.
(202) 746-7705
systover@imintel.org

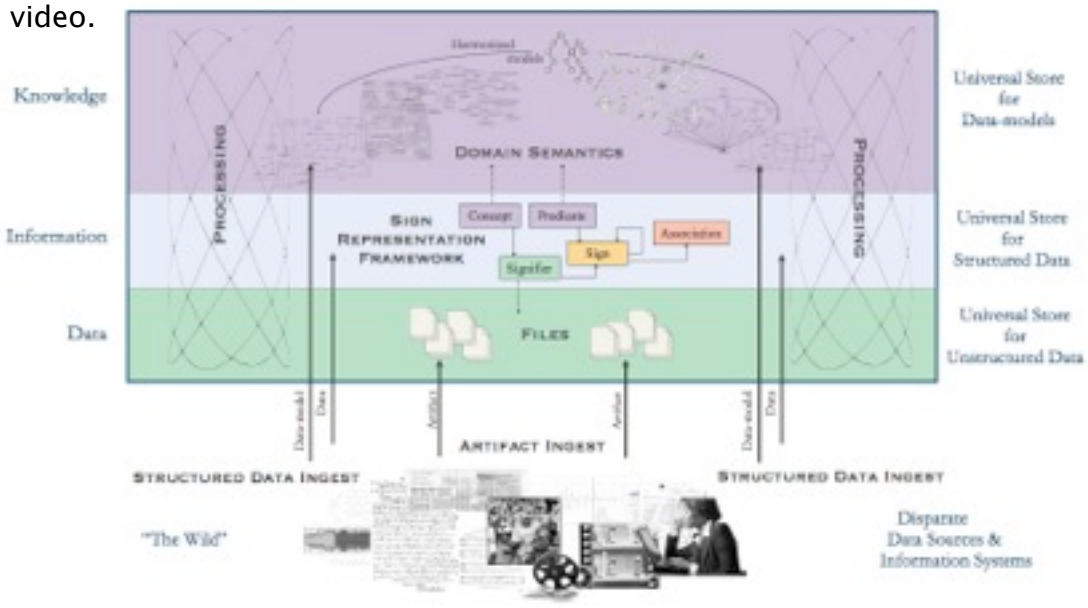
M. Andrew Eick
(202) 746-7706
maeick@missionfocus.com

Thing 5 - The NGA's Cloud

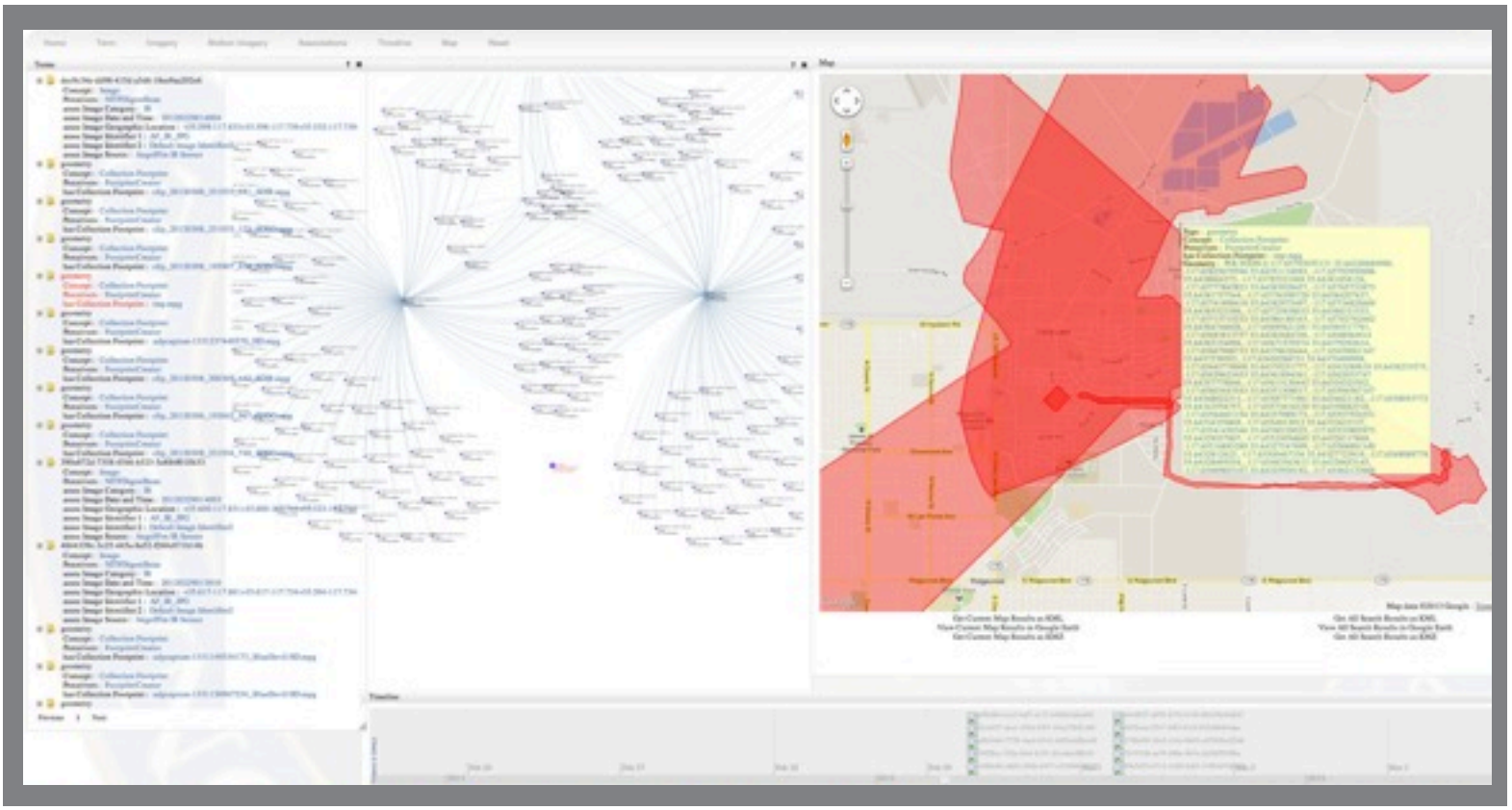
A semiotic **compute & storage** apparatus, built on an open-source **cloud** technology stack, for the storage and processing of diverse data at scale, including geospatial, temporal, human, social, cultural, behavioral, as well as traditional Intel data types and all modalities from documents to streaming video.

Features

- Internet-scale data store
- Massively parallel computation
- Tremendous aggregate bandwidth
- Super powerful knowledge representation technology
- Unified interface to all data, information, knowledge
- Web service access
- Semantic, geospatial, temporal, & contextual disambiguation
- Unified ingest
- Bulk data import / export
- Commodity HW
- Open source SW
- Engineered for PL4



A Web of Data, Information, & Knowledge
Sourced from Disparate Systems
Persisted on a Unified Fabric
Operationalized by a Dynamic Ecosystem of Analytics, Tools, & Applications



Thing 5 - The NGA's Cloud

Scope

- **Any encoding, any modality**
  - Video, images, audio, text, numeric series, ...
- **File standards of any kind**
  - NITF, MP2TS, XML, IRC, email, HTML, ...
- **Structured data & data-models of all kinds**
  - ShapeFiles, spreadsheets, metadata,...
  - Relational, object, hierarchical, graph, key-value, ...
  - Taxonomies, schemas, ontologies, ...
- **Anything with a geometry / time**
  - GIS features, events, ...
- **Whatever processing can be expressed in code**

Benefits

- No barriers to data ingest, its fast, simple, and universal
- All users & all processes can access all the data
- Disparate data are unified and operationalized without information loss or distortion
- Richness and meaning of information from the source is preserved
- Data-model harmonization is more powerfully supported, but not required
- True data integration across domains – connecting the dots – is achieved
- Entirely new kinds of analytics, tools, and applications become possible

Capabilities

Assert, retrieve, delete, tally, query **data, information, knowledge**

Search – **Keyword, semantic, geospatial, temporal**

Extract, characterize, expose – **Make information more discoverable**

Connect – **Assert new associations** (e.g. social network analysis, registration)

Surf – **Follow associations within and across semantic domains**

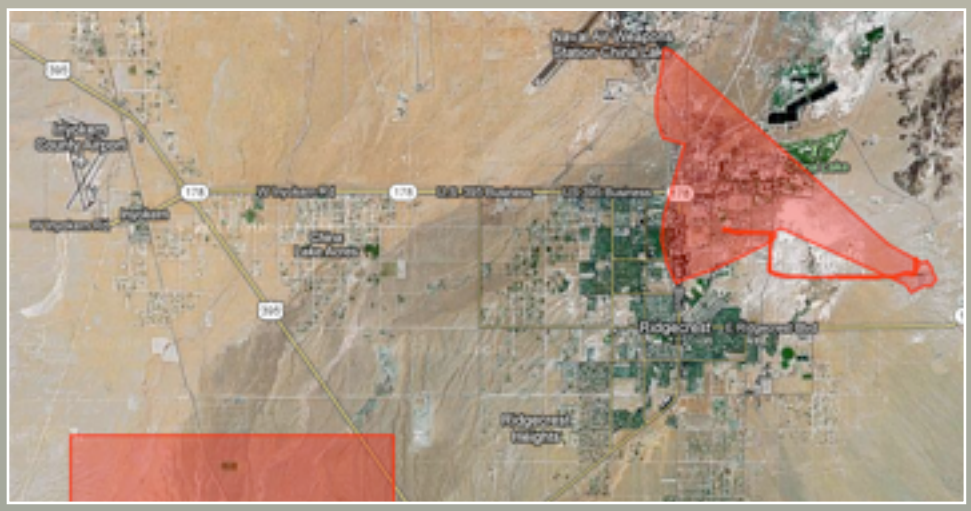
Mine – **Discover and expose new information** (e.g. identification, tracking)

Model – **Play out consequences and explore what if scenarios** (e.g. mobility /visibility analysis)

Manage – **Analyze and cultivate the entire data topology and ecosystem of processing**

Status

- Operational on China Lake development cluster
- PL2 ATO on DDTE network
- Passed FAT on production hardware
- Scheduled for theater deployment 2Q FY13



Mission Focus Points of Contact
M. Andrew Eick (202) 746-7706 maeick@missionfocus.com
S. Yoakum-Stover (202) 746-7705 systover@imintel.org

Government Points of Contact
L. Turner (571) 557-7253 lawrence.l.turner@nga.mil
J. Malapit (571) 557-9083 jeffrey.e.malapit.ctr@nga.mil