SemApp

What is SemApp?

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

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 warfighters employ new semiotic tools and applications to perform comprehensive mission SA, continuous agile assessment, planning, reporting, and execution.

What it will achieve

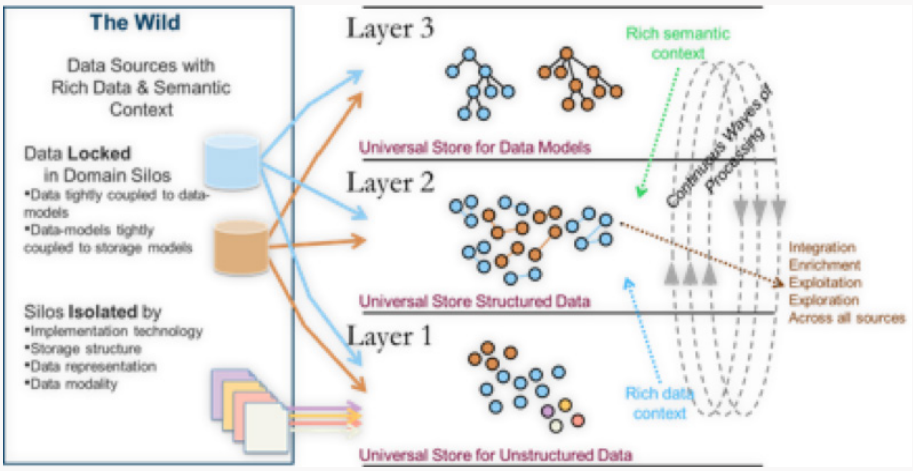
This unification of the data is the first step necessary to achieve advanced analysis envisioned by analysts (such as Structured Observation Management) to solve the toughest problems requiring access to the full diversity of data at scale. SemApp breaks the data barriers to put all of our data and computational intelligence assets in the service of the warfighter.

Why are we doing it?

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-

Figure 1. SemApp provides a universal store to accept all data types without loss or distortion. Continuous waves of processing enrich relationships within the data space with each user interaction, newly added data, or newly developed analytic.



Scope

- Any encoding, any modality
  - Video, images, audio, text, numeric series
- All file standards
  - NITF, MP2TS, XML, IRC, email, HTML
- All structured data and data-models
  - ShapeFiles, spreadsheets, metadata
  - Relational, object, hierarchical, graph, key value
  - Taxonomies, schemas, ontologies
- Anything with a geometry / time (e.g. GIS features, events)
- Whatever processing can be expressed in code

Benefits

- No barriers to data ingest, it's fast, simple, and universal
- All users and all processes can access all the data
- Disparate data are unified and operationalized without information loss or distortion
- Data-modal harmonization is more powerfully supported, but not required
- True data integration across domains - connection the dots - is achieved
- Entirely new kinds of analytics, tool,s and applications become possible
- Analytic products are made of and create a rich web of information

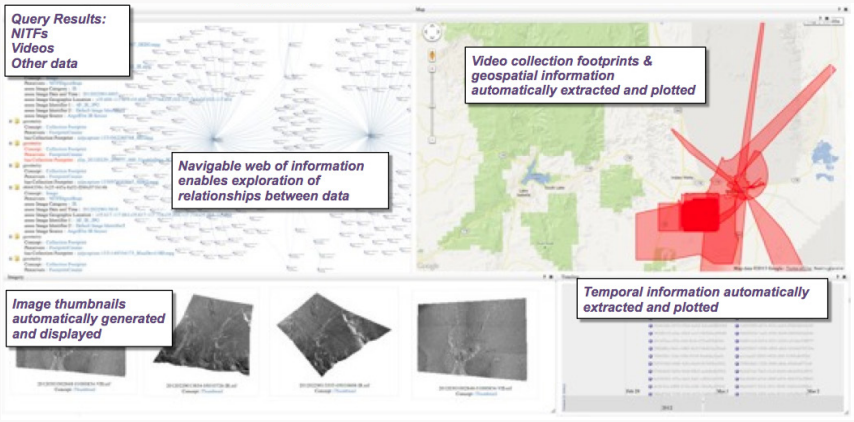


Figure 2. Screen Capture from the result of Geospatial Query around China Lake. The SemApp infrastructure has these relationships connections built in and provides to all applications and users.

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 SemApp 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.

Key Features

- Internet-scale data store
- Capability increasable by simply adding commodity HW nodes
- Engineered to be fault tolerant on commodity HW
- Massively parallel computation
- Tremendous aggregate bandwidth
- Based on semiotic science & knowledge representation technology
- Semantic, geospatial, temporal, & contextual disambiguation
- Unified interface to all data, information, knowledge
- Web service access
- Unified ingest
- Bulk data import / export
- Engineered for PL4

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)
- 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