



SEMAPP OVERVIEW

2013 08 15

S. Yoakum-Stover, Ph.D.
M. Andrew Eick

Institute for Modern Intelligence
& Mission Focus



MOVE IT FORWARD. MAKE IT HAPPEN.



CONTENTS

- **WHAT IS IT**
- **WHY IT'S CALLED THING 5**
- **WHAT PROBLEM DOES IT SOLVE**
- **RELEVANCE TO NGA**
- **BACK UP**

MOVE IT FORWARD. MAKE IT HAPPEN.



WHAT IS IT?



MOVE IT FORWARD. MAKE IT HAPPEN.

SEMAPP



MOVE IT FORWARD. MAKE IT HAPPEN.

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.

SEMIOTICS

The science which studies how meanings are made and how reality is **represented**

The study of **signs** - anything which stands for something else



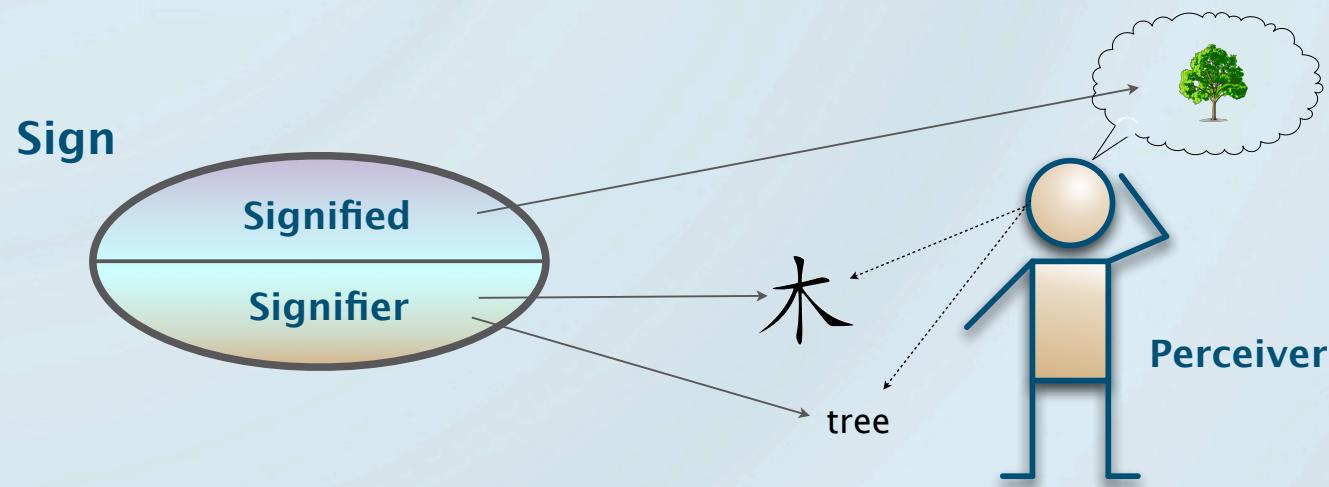
Developed by
Ferdinand de Saussure (1857-1913)



Charles Sanders Peirce (1839 - 1914)

Logic and Linguistics may be classified as branches of semiotics

MOVE IT FORWARD. MAKE IT HAPPEN.

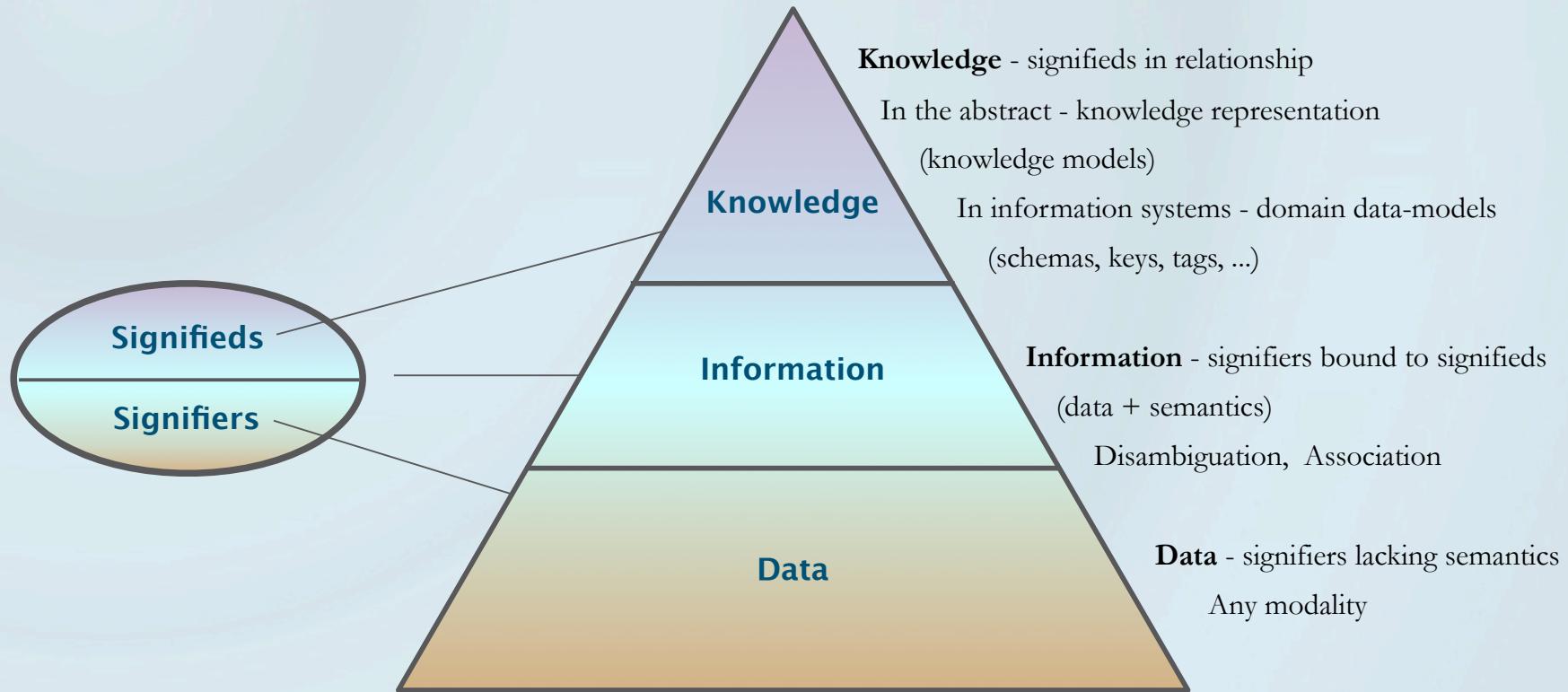


To represent a thing, an information system must capture both signifier and signified



MOVE IT FORWARD. MAKE IT HAPPEN.

SIGN STORE



In typical information system, signifieds are presented as the storage-model (schema)

This approach restricts the scope to 1 perspective (set of meanings)

SemApp represents signifiers and signifieds as equal citizens

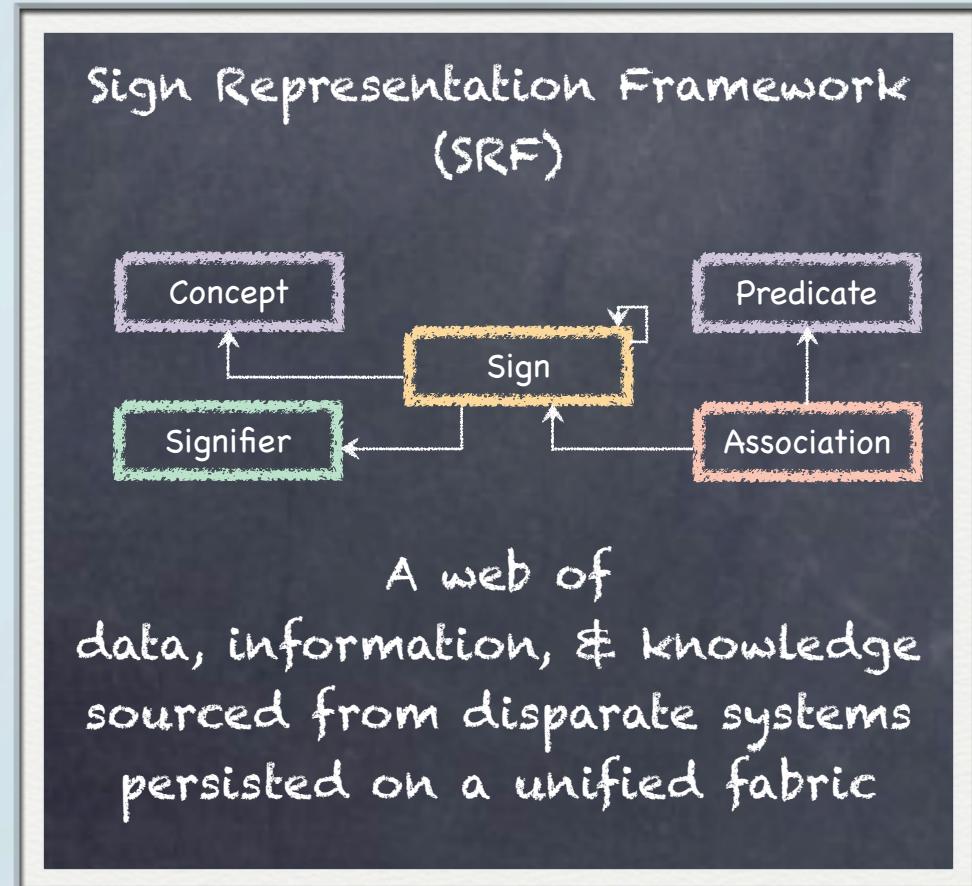
The storage-model is decoupled from the perspective

This approach accommodates any / all perspectives

SIGN REPRESENTATION FRAMEWORK

SIGN STORAGE MODEL FOR ULTRA-LARGE SCALE

- Accommodate all, specify none
 - 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, ...
- Access all content - unified interface
- Ingest anything
- Preserve domain semantics
- Disambiguate semantically, geospatially, temporally, contextually
- Harmonize data-models, or not
- Integrate information
- Build entirely new kinds of tools, analytics, applications





COMPUTE & STORAGE CLOUD

CHARACTERISTICS

- Internet-scale data store
- Massively parallel computation engine
- Tremendous aggregate bandwidth
- Commodity HW
- Open source SW

ACHIEVING SCALE

- Cloud achieves scale
 - Data size & processing performance
 - Marrying economy with technology
- SRF achieves another level of scale
 - Production and content management
 - Faithfully representing meaning and organizing diversity

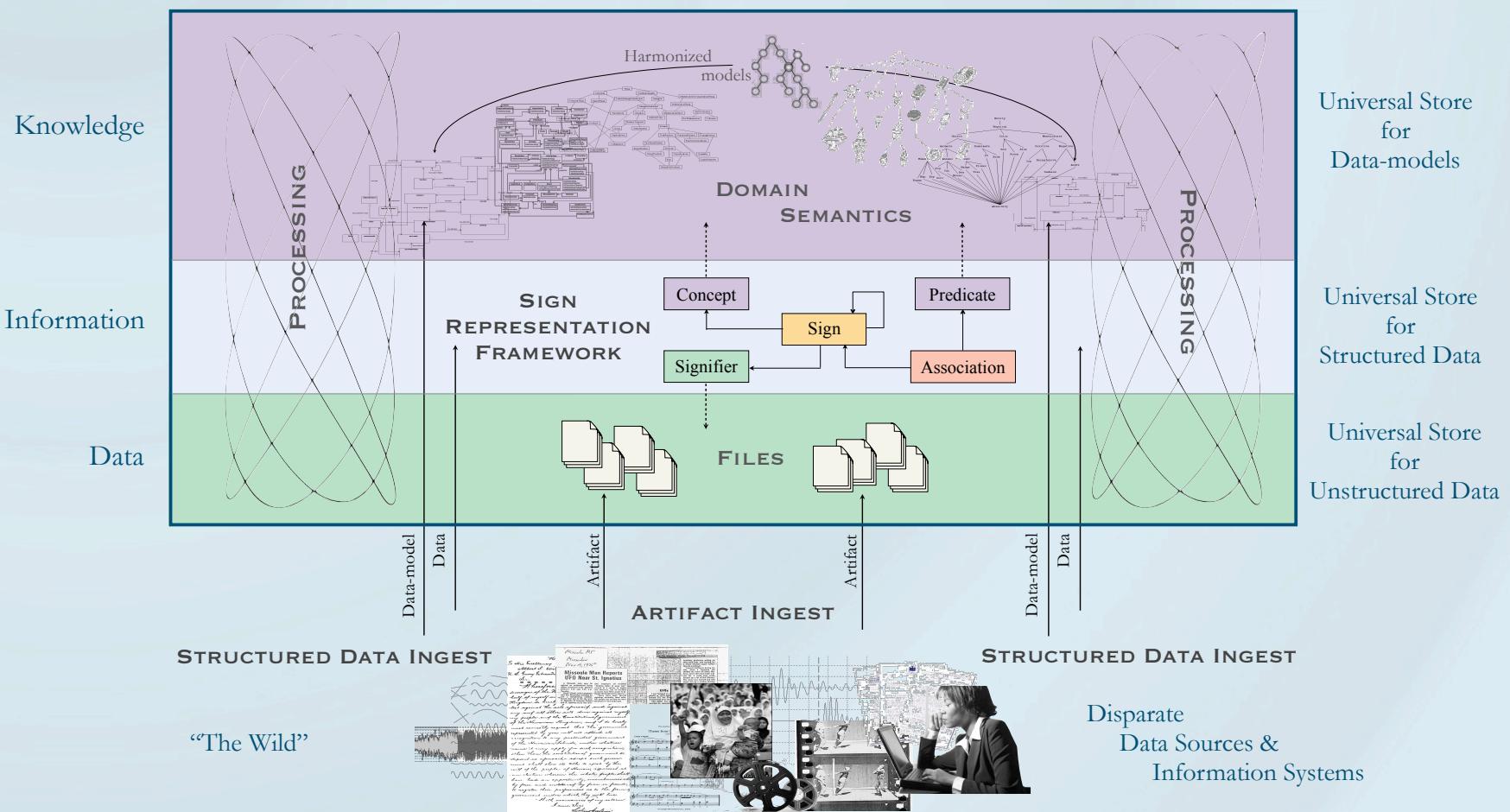
MOVE IT FORWARD. MAKE IT HAPPEN.



SEMIOTIC APPARATUS (SEMAPP)

MOVE IT FORWARD. MAKE IT HAPPEN.

SIGN-SPACE = SEMIOTIC APPARATUS + DATA + PROCESSING





BENEFITS & CAPABILITIES

BENEFITS

- No barriers to data ingest, it's 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 enabled
- 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



WHY'S IT CALLED THING 5?



MOVE IT FORWARD. MAKE IT HAPPEN.



IN THE BEGINNING

CDR KRAFT, SIR... SO WHERE DO WE FIT IN?

I GOT A HANDFUL OF THINGS I WORRY ABOUT

THING 1 IS NVS

THING 2 IS ...

YOU'RE THING 5

YES, SIR :-)



WHAT PROBLEM DOES IT SOLVE?



MOVE IT FORWARD. MAKE IT HAPPEN.



THE “DATA” PROBLEM

OUR INTELLIGENCE ASSETS

DATA, INFORMATION, KNOWLEDGE, ANALYTICS, TOOLS, APPLICATIONS...

ARE FRACTURED

IN PHYSICALLY AND SEMANTICALLY DISPARATE SYSTEMS

OPERATIONS ARE IMPEDED

BY OUR INABILITY TO FULLY

SEARCH, EXPLORE, ENRICH, MANAGE, AND EXPLOIT OUR ASSETS

TO CRACK THE DATA / PROCESSING / GLOBAL SENSE-MAKING PROBLEM

SEMAPP

UNIFIES ALL DATA & SUPPORTS ALL PROCESSING



RELEVANCE TO NGA



MOVE IT FORWARD. MAKE IT HAPPEN.



BOLD VISION WICKED BARRIERS

As long as we have the courage to
“let the data surprise us,”

I am convinced our analysts will lead to new discoveries—and not only by NGA.

We will **enrich the analysis of others**,
especially the analysis performed by the **all source analyst**.

Letitia Long, Director, NGA

NATIONAL GEOSPATIAL-INTELLIGENCE AGENCY
Know the Earth... Show the Way... Understand the World

NGA Has a *Bold* Vision

Putting the Power of GEOINT in Your Hands



NGA Strategy

Provide online, on-demand access to our GEOINT knowledge

Broaden and deepen our analytic expertise to produce new value

“We will continue to deliver to our varied customer set what they need, when they need it, how they need it. But we have to be thinking about the future. We have to be continually pushing ourselves so that we do remain at the forefront.”

— Letitia A. Long, Director, NGA

Approved for Public Release – NGA Case #13-153

MOVE IT FORWARD. MAKE IT HAPPEN.

BARRIERS

- The Data Problem
- The Processing Problem
- The Network Effect Problem

THING 5
BREAKS THE BARRIERS



KNOW THE EARTH

MEANS KNOWING WHAT IS - The Physical & the Human terrain

FOUNDATION DATA - From imagery to custom / thematic maps

- Representing our human reality
- Telling the story of the earth and the dynamic things, natural and man-made, that operate upon it
- Foundation for analytic discourse

MAKING & MANAGING FOUNDATION DATA - Core to NGA's core domain

- Ravenous and impatient data consumer
- Starving and languishing amidst troves of data, information, and knowledge
- Something is wrong

THE DATA PROBLEM - Diversity, ambiguity, scale

- How meaning is represented - knowledge representation and semiotics
- Domain models and schemas built into every information system
- Not just more data, profoundly different - really, Really Big Data
- Inability to access, understand, manage and use our data assets as a coherent whole

THING 5

- Accommodates diversity, reduces ambiguity, and handles scale
- Access, understand, manage and use our data assets as a coherent whole
- Feed the production chain with rich, diverse information
- Custom / thematic maps as linked data using information from any source

MOVE IT FORWARD. MAKE IT HAPPEN.



SHOW THE WAY

MEANS NAVIGATING THAT TERRAIN SAFELY

- Tactical level of capability that a local understanding enables

ANALYTICS

- Modeling, simulation, what-if analyses, ...
- Operationalizing the data
- Conduct of analytic discourse, a man-machine partnership

MAKING & USING ANALYTICS - Essential to NGA's core domain

- Scattered across programs and organizations
- Competing for resources and saddled with overhead
- Something is wrong

THE PROCESSING PROBLEM - The pipeline model

- Data flowing between processes
- Data and processing assets are different at every point in the enterprise
- Bandwidth-limited & complex
- Inability to leverage, understand, manage and apply our analytic assets as a coherent whole

THING 5

- Replaces the pipeline model with a wave model - analytic waves over data at rest
- Leverage, understand, manage and apply our analytic assets as a coherent whole
- All analytics access and enrich a shared ocean of data, information, knowledge
- Simple, robust, efficient, powerful, creative



UNDERSTAND THE WORLD

MEANS KNOWING WHAT IT ALL MEANS

- A strategic level of capability that a global understanding enables

INTERCONNECTED DATA - Interconnectedness at all scales

- Putting the pieces together - a living web of information
- Sharpening the saw - adjusting data and analytic assets in anticipation
- Product of analytic discourse

MAKING & USING INTERCONNECTED DATA - Essential to NGA's core domain

- Shared objective across the IC
- Struggling despite troves of data, information, knowledge, and analytics
- Something is wrong

THE NETWORK EFFECT PROBLEM

- Not compounding (each contribution increases the value of the whole and vice versa)
- No feedback mechanism whereby output (product) enriches the input
- Inability to fully benefit from achievements and function strategically

THING 5

- A rich, living web of information naturally emerges
- Each contribution increases the value of the whole and vice versa
- Products are an enrichment of the web
- Fully implement Structured Observation Management
- Achieve and apply new levels of understanding to achieve mission objectives and make NGA better

MOVE IT FORWARD. MAKE IT HAPPEN.



MOVE IT FORWARD. MAKE IT HAPPEN.

PRODUCTS AS INTERCONNECTED DATA

- ❑ A product consists of
 - Selected elements of information (signs and associations)
 - Linked together (associations)
 - Displayed by an application to suit a particular purpose
- ❑ Examples
 - Notebook
 - Thematic map
 - Social network
 - Geospatial track
- ❑ Information elements can appear in any number of different products
- ❑ Product creation enriches the web of interconnected data
- ❑ Nothing to “re-ingest”
- ❑ Every product makes the information more valuable

INSTITUTE FOR MODERN INTELLIGENCE

AN ANALYST NEEDS A GOOD NOTEBOOK

A NOTEBOOK APPLICATION

- ❑ A tool for performing research and analysis
- ❑ Collect, associate, and annotate sign-space elements
 - Artifacts, signs, associations

SIMPLE APPLICATION DATA-MODEL

- ❑ A book containing pages
- ❑ Each page consists of a linked list of parts
- ❑ Any sign-space element can be represented within a part

CONOPS

- ❑ Author assembles parts into a page
 - Drag - drop / copy - paste / edit - assert
- ❑ This enriches the sign-space with associations
 - No actual document is created
 - Originating author (perceiver) is recorded
- ❑ Multiple authors may collaborate on a page

UNCLASSIFIED // PROPRIETARY // COPYRIGHT © 2013 INSTITUTE FOR MODERN INTELLIGENCE, 501(C)3. ALL RIGHTS RESERVED.

10

INFORMATION WEB GROWS AS A SIDE-EFFECT OF ANALYTIC WORK

UNCLASSIFIED // PROPRIETARY // COPYRIGHT © 2013 INSTITUTE FOR MODERN INTELLIGENCE, 501(C)3. ALL RIGHTS RESERVED.

20



SPACE & TIME

SPACE & TIME

- ❑ The physics that makes NGA and Thing 5 special
- ❑ The bedrock of the IC
- ❑ The unifying threads throughout our data

THING 5 CORE

- ❑ Semantic disambiguation
 - Binding of a element of **data** (the signifier) to a **concept** (the signified)
 - Forms a Sign - the atom of semantics in Thing 5
 - e.g. [systover@imintel.org, Sender]
- ❑ Geospatial disambiguation
 - A Sign that also has a **geometry** (representing geospatial coordinates) e.g. [Nile, River, 30°10'N 031°06'E]
 - Isomorphic to a GIS feature
 - Applies to associations as well
- ❑ Temporal disambiguation
 - A sign that also has a **period** (representing temporal coordinates) e.g. [Wall Street bombing, Terrorist Act, 19200916]
 - Isomorphic to an event
 - Applies to associations as well

POWER

- ❑ Native GEOINT support
- ❑ Ideal platform for ABI
- ❑ Search, explore, discover, enrich, manage, exploit these unifying threads throughout our data
- ❑ IC solution for GEOINT storage and processing

MOVE IT FORWARD. MAKE IT HAPPEN.



BACK UP



MOVE IT FORWARD. MAKE IT HAPPEN.



DATA-MODELS & THE TROUBLE WITH DATA



MOVE IT FORWARD. MAKE IT HAPPEN.



MOVE IT FORWARD. MAKE IT HAPPEN.

DATA-MODELS

SPECIFY

- Vocabulary, Structure, Semantics, and Constraints

BUILT FOR A PURPOSE

- Make a particular analytic, tool, application work efficiently
- Entail a particular perspective on the data
- They're all different for a reason

PROVIDE THE INTERFACE TO AND PERSPECTIVE ON THE DATA

- All interactions with data are mediated by a data-model
 - We go thru the data-model to ask questions about the data
 - What is the **name** of the **Student** with **sid** 53666?
 - How many **Students** have **age** = 18?
 - What is average **gpa** of **Students**?
 - We can not directly question the data
 - What is "Jones"?
 - Is there anything on "Guldu" ?
 - What is "53831" related to?
 - Indexes are no exception
 - Every index reflects a hidden data-model
 - The "things" being indexed
 - e.g. Index on **Students name**

Students

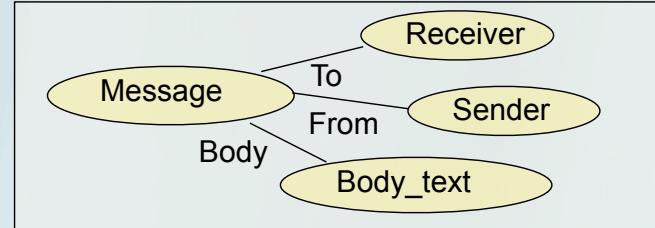
| sid | name | login | age | gpa |
|-------|---------|---------------|-----|-----|
| 53666 | Jones | jones@cs | 18 | 3.4 |
| 53688 | Smith | smith@ee | 18 | 3.2 |
| 53650 | Smith | smith@math | 19 | 3.8 |
| 53831 | Madayan | madayan@music | 11 | 1.8 |
| 53832 | Guldu | guldu@music | 12 | 2.0 |



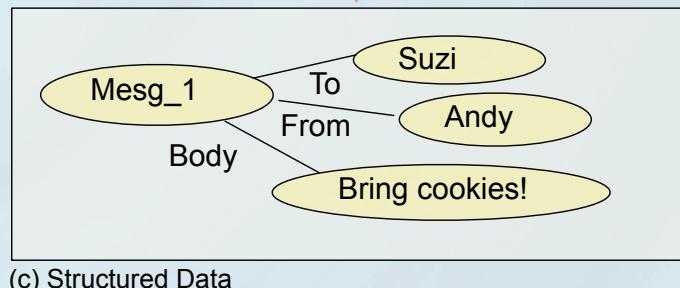
DATA-MODEL IN ACTION (RDB-STYLE)

07/04/07
Suzi,
Bring cookies!
Andy

(a) Unstructured Data



(b) Data-model



(c) Structured Data

The data-model is imposed on the data store (schema)

The data values are frozen beneath it (e.g. We can't say new things about Andy)

Once loaded, it's painful to modify the storage structure

Introducing data from other sources is hard as data-models must conform to the storage structure

Once you do that however, the data is “integrated” (but according to only 1 perspective)

| Message | Receiver | Sender | Body_text |
|---------|----------|--------|----------------|
| Msg_1 | Suzi | Andy | Bring cookies! |
| ... | | | |

(d) Typical database structure

Information is typically lost / distorted in the process

THE PROBLEM WITH DATA



MOVE IT FORWARD. MAKE IT HAPPEN.

is...

It Has to Go Somewhere and often goes Lots of Places

Then it Tends to

Stick where it Lands and take the Shape of its Container

In other words,

The Data-Model gets

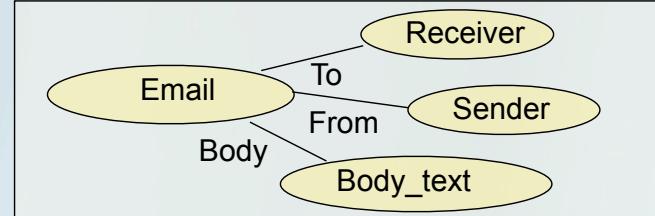
Imposed on the Data Store and the Data is then Frozen Into it

So Don't do That

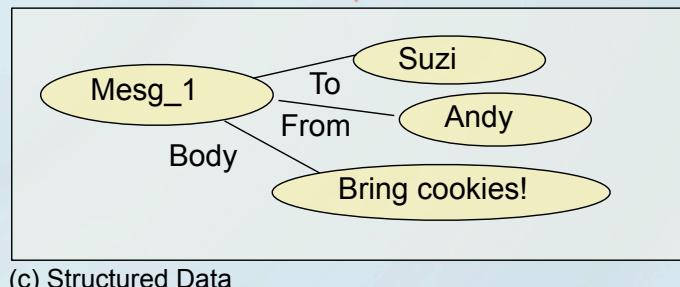
DATA-MODEL IN ACTION (CLOUD-STYLE)

07/04/07
Suzi,
Bring cookies!
Andy

(a) Unstructured Data



(b) Data-model



(c) Structured Data

The data-model is still there, but it's mingled with the data

We still need to know what it is to access our data

We also need to know how data in the value field is structured

| Row | Family | Qualifier | TimeStamp | Value |
|--------|--------|-----------|-------------|----------------------|
| Mesg_1 | Email | Sender | 20090224123 | Andy |
| Mesg_1 | Email | Body_text | 20090224125 | Suzi, Bring cookies! |

(d) Typical CloudDB structure (e.g. BigTable)

Introducing data from other sources is easy

Keeping track their data-models (interfaces) is hard

The data is not integrated

THE PROBLEM WITH CLOUD



MOVE IT FORWARD. MAKE IT HAPPEN.

is...

The Data Structure has to **GO SOMEWHERE**

AND CAN GO

almost Anywhere

So

Data flows Easily **INTO** its Container

BUT LOSES ITS SHAPE

In other words,

The Data-Model gets

Mixed with the Data in a Specific but **Arbitrary** way

So Don't do That



DATA INTEGRATION CONUNDRUM

SITUATION WE HAVE:

A data-model Provides

The Interface to the data AND the Semantics of the data

For data to be integrated it must have a Unified interface

Unified Interface ⇒

A single data-model

A single meaning / perspective

But there is no single right way to represent all Knowledge ⇒

There can be no single data-model for all intel

SITUATION WE WANT:

A unified interface to All the data

that

Does NOT impose a specific Data-model

THE SOLUTION



MOVE IT FORWARD. MAKE IT HAPPEN.

CONSIDER DATA-MODELS FROM A

Higher Level of Abstraction

Distill from there a

Minimal set of Elements Sufficient to Capture
Any data-Model

and then

Build Storage Model based On that That



WHO ARE WE?



MOVE IT FORWARD. MAKE IT HAPPEN.



WHO WE ARE

Institute for Modern Intelligence

IMI aims to change the world!
by developing
an Ultra-Large Scale systems infrastructure
for data-intensive computing & Data-intensive operations

Our mission:

Develop the Science, Practice, and Governance of Modern Intelligence

We are a non-profit, 501(c)3 organization
whose purpose is to fulfill our mission.

S. Yoakum-Stover & M. A. Eick are the IMI founders

Mission Focus

Mission Focus aims to build clean code that works for its customers.
We are an agile software development shop that takes
domain design and development as seriously as
system design and development.

Our tag line is:

Move it Forward, Make it Happen

We are a small, for-profit organization
whose purpose is to fulfill its customers' missions.
Our profits help to support the IMI.

M. A. Eick & S. Yoakum-Stover are CEO and Chief Scientist respectively

Core Domain

Mission Focus and IMI work mostly in the intelligence domain with DoD and IC customers / partners.
Our core domain addresses the storage, processing, and utilization of data in the context of immense volume and diversity.

We are experts in cloud computing and storage technology.
We invented the Sign Representation Framework which underpins a game-changing approach to data unification.

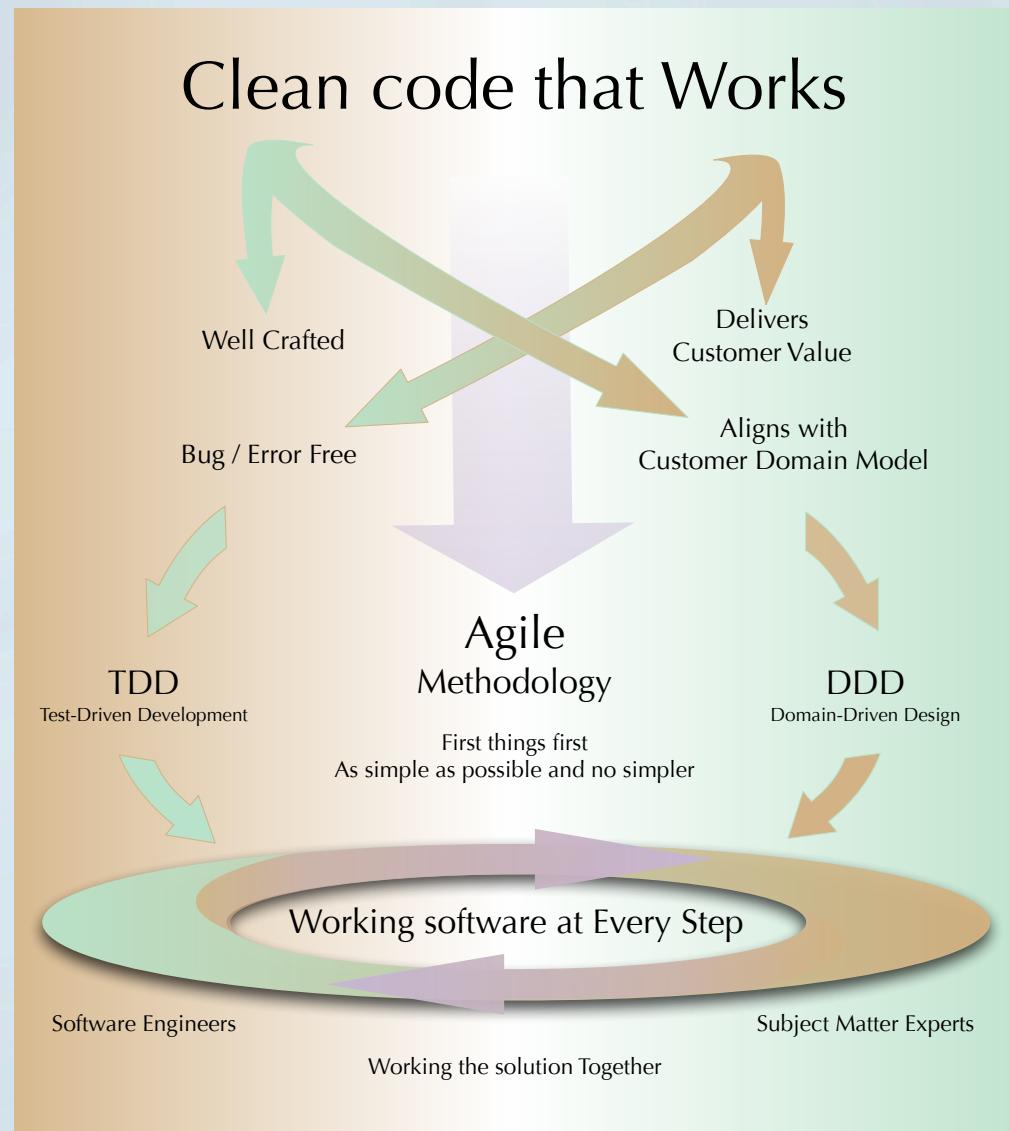
We pride ourselves on our disciplined engineering practices and distinguish ourselves by our ability to continually learn and innovate.

The work we do is meaningful and intentional and is wrapped with our integrity.

Put simply,
We just think harder and work better than the rest.

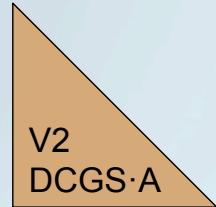
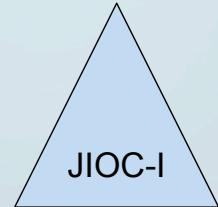


HOW WE WORK





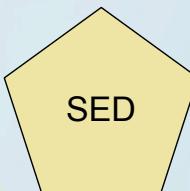
THING 5 HERITAGE



- Dr. Suzi Object Sciences @ I2WD
- JIOC-I V2 integration
- Universal store for documents meets UFEIEO
- We need a universal store for structured data



- Andy SSS / TexelTek under Dr. Heath @ NSA
- GeoBoost – thin client Geo visualization
- Ozone widget framework
- Cloudbase analytics under Dr. Dowd



- Dr. Suzi PIPS @ I2WD Tech Dir
- Andy SSS / TexelTek
- Cross-domain MLS Oracle
- Data Description Framework
- The data integration problem



- Dr. Suzi IMI @ I2WD SETA Tech Dir
- Andy MissionFocus
- Compute & storage cloud aspiration
- Legacy integration implementation



- Dr. Suzi IMI & MissionFocus
- Andy MissionFocus & IMI
- GEOINT Compute & storage cloud implementation
- Sign Representation Framework
- PL4 design
- Marry cloud technology with semiotic science



MOVE IT FORWARD. MAKE IT HAPPEN.

Scale changes everything

Walk with your head in the clouds
and your feet in the dirt

Everything changes when you
deploy to a production
environment

The ground state of
data is not a pure
crystalline form

Everything changes when
you start writing the code

There's More
In_all_of_it_Together
Than we are getting or
could possibly get out of
All of the Parts

Don't sell yourself short (the rest of the world will do it for you soon enough)

Our universe of data & processing is an
Ultra-Large Scale system

New science waits to be discovered

Simple is hard

Diversity gives Intel its richness

Every system should
have the ability to
export its data

Our data & processing assets
belong to the nation
(not the system or the contractor)

What you don't have
in your head, you gotta
have in your feet

The sink is not a
repository for dirty dishes



Be happy, write code!

Godspeed