

Surveillance, Events and the Semantic Web

Environmental Protection Agency's Environmental Information Symposium

Phoenix, Ariz., December 10-12, 2008

Dr. Nancy Grady (GradyN@saic.com)

Overview



- **Surveillance**
- Event models
- Emerging Web technologies
- Putting it all together

Surveillance



Mining for actionable intelligence

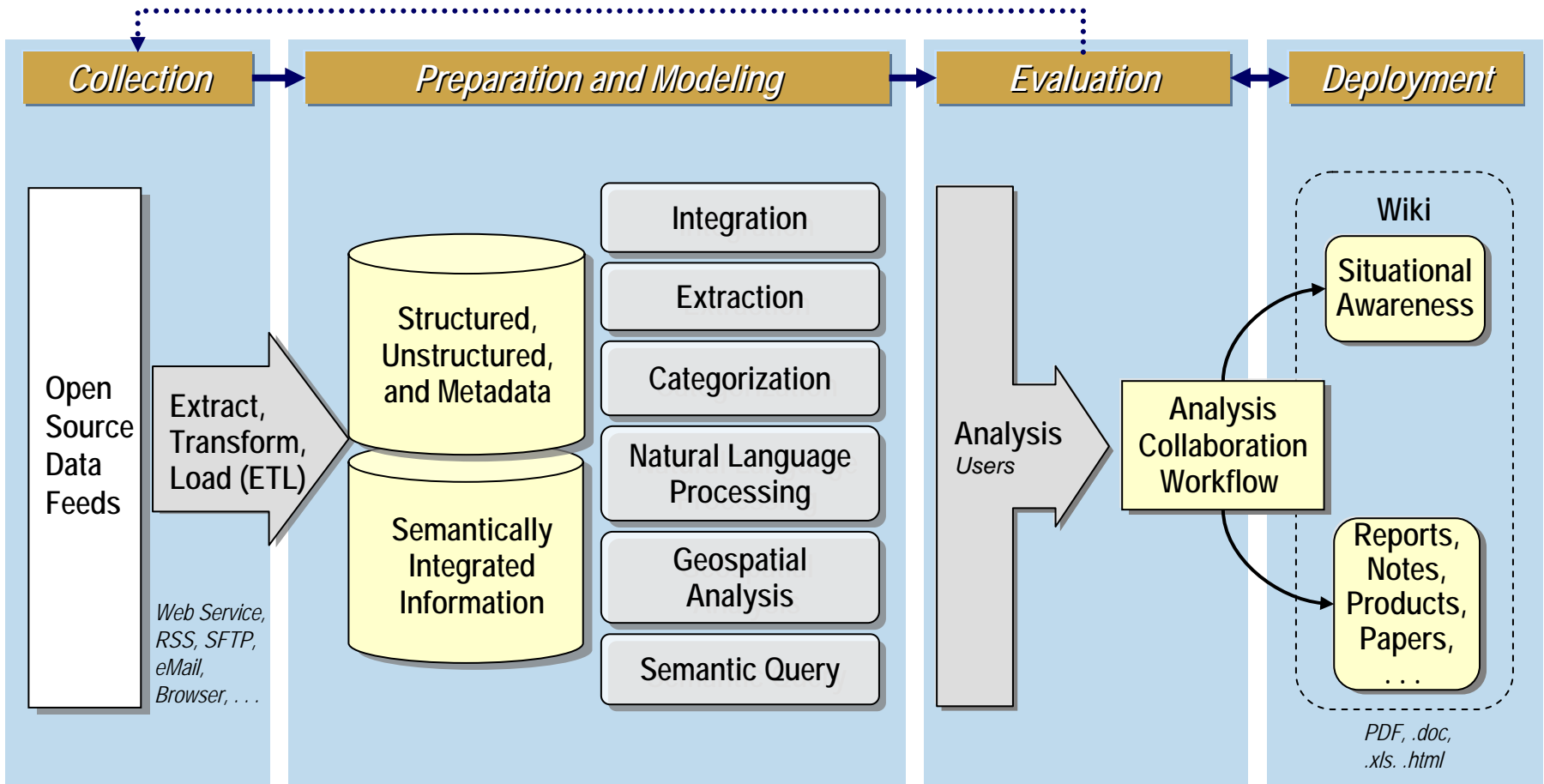
- Traditional surveillance
 - Look for anomalous activity in sensor data
 - Fuse primary or secondary data sources
 - Federated datasets
- New surveillance for “events”
 - Harvest Web-based information
 - Exchange event data with other agencies

Data Fusion for Surveillance



- Structured
 - Standardize
 - Semantic integration through ontologies
- Un-structured
 - Categorization
 - De-duplication
 - Entity extraction
- Un-structured into structured
 - Targeted event extraction (natural language processing)

Conceptual Architecture



Overview



- Surveillance
- **Event models**
- Emerging Web technologies
- Putting it all together

Events Are the Output of Surveillance



- Who
- What
- Where
- When
- How
- How many
- Why

Exchanging Events



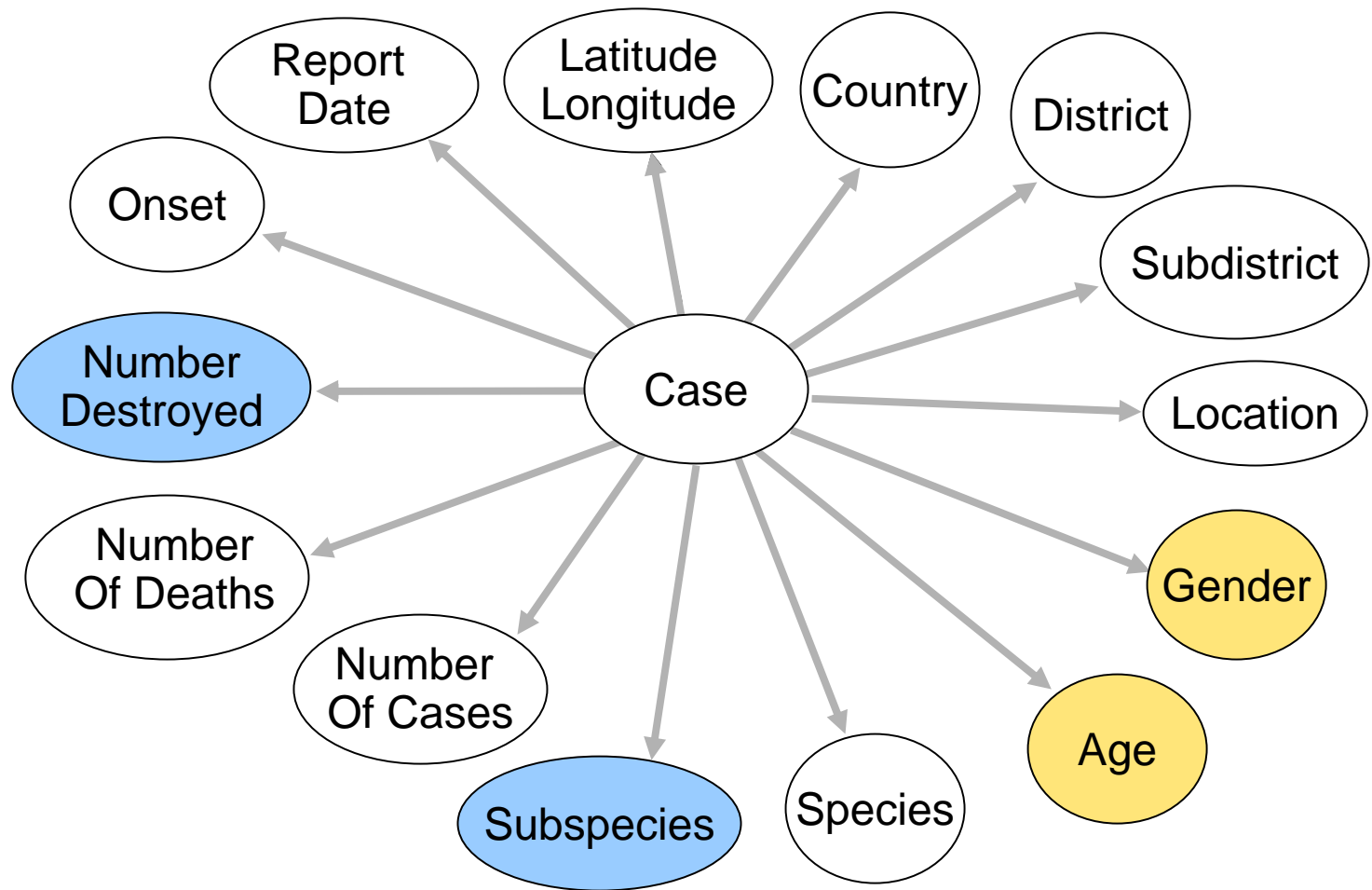
- *Situational awareness* is communicating significant events
- Easier to collaborate around events than primary/secondary data
 - Privacy
 - Bandwidth
 - “Local” expertise on data
 - Existing surveillance systems tuned to data
- Open source (or intelligence) will be about events

Event Representation



- Prose
- Spreadsheets
- Relational database
- Relationship database (ontology)
- Semantic Web

Spreadsheet Model

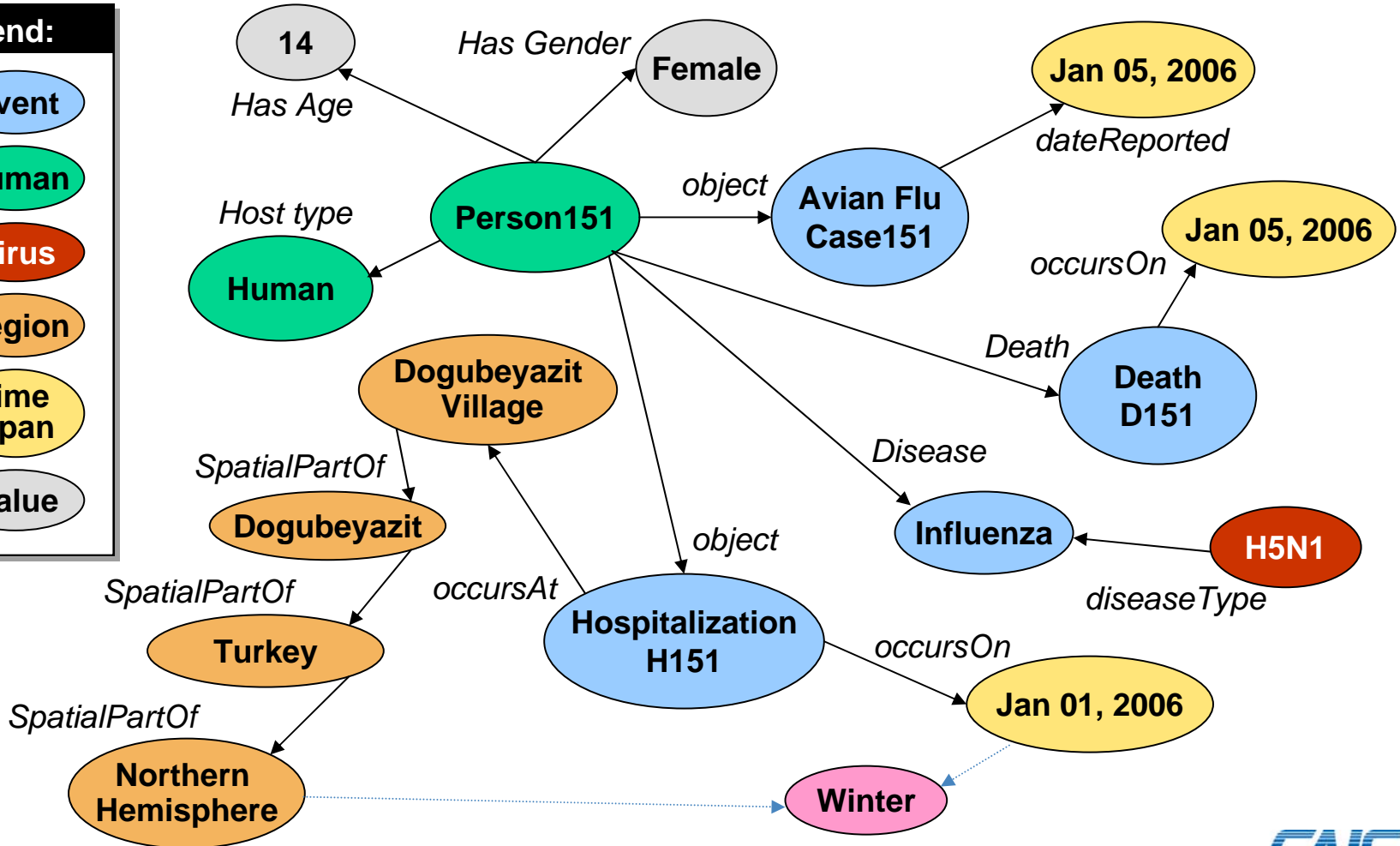
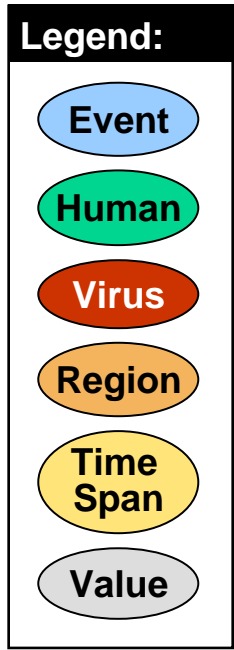


Relational Database Event Model



- Header
 - domain, agent type, dates, workflow
- Host (and symptoms)
- Agent,disease/vector/reservoir
- Location (and landforms)
- Sources (include veracity)
- Measurements

Relationship (Ontological) Model



Semantic Model



Subject-Verb-Predicate Markup

- Person hasAge <value>
- Person hasDisease <disease name>
- Person hasResidence <region>

Comparisons



- Prose is difficult to track and query
- Spreadsheets are easy but limited
- Relational databases are more familiar to developers and easier to aggregate
- Ontologies are easier for queries using context and offer better scalability to many datasets
- Semantic Web better for interoperability and flexibility, not as good for contextual queries, good for tagging within prose

Maturing an Event Model



- Source descriptions
 - Veracity
- Analytical pedigree
 - Preparation and analytical techniques
- Community standards
 - Codes and representations
- Spatial extend descriptions

Event Analytics Will Require



- De-duplication
- Overlap or “collision” detection
- Temporal changes
- Historical trending
- Correlation with activity “scores”

Overview



- Surveillance
- Event models
- **Emerging Web technologies**
- Putting it all together

Web (1.0, 2.0, 3.0)



- **Content**

- 1) Brochureware
- 2) Social networking
- 3) **Semantic Web** ... data.

- **Search**

- 1) PageRank prioritization
- 2) Influencers
- 3) **Semantic searches**

*HTML = Extensible HyperText Markup Language

**CSS = Cascading Style Sheets

***XML = eXtended Markup Language

- **Software**

- 1) HTML* and CSS**, XML***
- 2) **Wikis**, blogs, tag clouds
- 3) **Services** for data exchange

- **Hardware**

- 1) Servers
- 2) Server farms
- 3) **Cloud computing**

Incident



page discussion edit history

from SEtoSE

Sungei Kadut 23 Sep 08

Outbreak Sungei Kadut Chikungunya **Super Event** Chikungunya Singapore
Number of Cases 4 **Description**
Onset Start Date 2008/09/18 **Onset End Date**

Hospitalization
Residence Sungei Kadut, Singapore **Exposure Origin**
Work **Travel**
Web Source
Page Source Singapore_Ministry_of_Health_Update_2008_September_30

4 new cases in the vicinity of Sungei Kadut developed symptoms between 18 Sep and 23 Sep 08. This brings the total number of cases in the cluster to 25.

Facts about Sungei Kadut 23 Sep 08

MemberOfOutbreak	Sungei Kadut Chikungunya + 🔍
MemberOfSuperEvent	Chikungunya Singapore + 🔍
NumberOfCases	4 + 🔍
OnsetStartDate	18 September 2008 + 🔍
ResidenceCountry	Singapore + 🔍
ResidenceLocation	Sungei Kadut + 🔍
SourcePage	Singapore Ministry of Health Update 2008 September 30 + 🔍

Category: IDIncident

Log in / create account

- infectious disease events
 - Add an event
 - List Outbreaks
 - List Incidents
- objects
 - People
 - User groups
 - Organizations
 - Artifacts
 - Topics
- corporateevents
 - Proposals
 - Projects
 - IRADs
 - Conferences
- site
 - Category Tree
 - Categories
 - Properties
 - Types
- navigation
 - Main Page
 - Community portal
 - Current events
 - Recent changes
 - Random page
 - Help

Semantic Query Over Incidents



[Log in / create account](#)

page discussion edit history

Sungei Kadut Chikungunya

from SEtoSE

infectious disease events

- Add an event
- List Outbreaks
- List Incidents

objects

- People
- User groups
- Organizations
- Artifacts
- Topics

corporateevents

- Proposals
- Projects
- IRADs
- Conferences

site

- Category Tree
- Categories
- Properties
- Types

navigation

- Main Page
- Community portal
- Current events
- Recent changes
- Random page
- Help

Incidents in this Outbreak

Incidents	NumberOfCases	OnsetStartDate	OnsetEndDate
Jalan Jendela 18 Sep 2008	1	18 September 2008	
Sungei Kadut 23 Sep 08	4	18 September 2008	
Sungei Kadut 11 Sep 08	1	11 September 2008	
Sungei Kadut 7 Sep 08	3	7 September 2008	
Sungei Kadut 27 Aug 08	6	21 August 2008	
Sungei Kadut 20 Aug 08	1	20 August 2008	
Sungei Kadut 18 Aug 08	1	18 August 2008	
Sungei Kadut 14 Aug 08	1	14 August 2008	
Sungei Kadut 8 Sep 08	3	10 August 2008	
Sungei Kadut 9 Aug 08	1	9 August 2008	
Sungei Kadut 4 Aug 08	1	4 August 2008	
Sungei Kadut 3Aug 08	1	3 August 2008	

Facts about Sungei Kadut Chikungunya

- End date: 23 September 2008
- FromCountry: Singapore
- HasDisease: Chikungunya
- MemberOfSuperEvent: Chikungunya Singapore
- NumberOfCases: 25
- Start date: 3 August 2008

Category: IDOutbreak

Sungei Kadut Chikungunya

Super Event: Chikungunya Singapore

Number of Cases: 25

Start Date: 2008/08/03

End Date: 2008/09/23

Country: Singapore

Disease: Chikungunya

Web Source

Page Source

Wiki for Collaboration and Semantic Tagging



from SEtoSE

infectious disease events

- Add an event
- List Outbreaks
- List Incidents

objects

- People
- User groups
- Organizations
- Artifacts
- Topics

corporate events

- Proposals
- Projects
- IRADs
- Conferences

site

- Category Tree
- Categories
- Properties
- Types

navigation

- Main Page
- Community portal
- Current events
- Recent changes
- Random page
- Help

search

Go Search

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version

page discussion edit history

Chikungunya Outbreak

Chikungunya fever is a viral disease transmitted to humans by the bite of infected mosquitoes. Chikungunya virus was first isolated from the blood of a febrile patient in Tanzania in 1953, and has since been cited as the cause of numerous human epidemics in many areas of Africa and Asia, and most recently in a limited area of Europe.

See also [Chikungunya Reference Information](#)

In the News

[Google News Timeline](#)

For latest news scroll down, or click on right hand column of timeline graphic. Timeline graphic may take several seconds to load and display.

World Map

Click here for a larger version.

Red dots within countries indicate traveler cases.

Chikungunya Virus
SciDevNet.com @2007
Creative Commons
Attribution 2.0 Licence

Log in / create account

What Is Cloud Computing?



Definition:

It is a model of computing where a pool of massively scalable, IT-related capabilities are provided “as a service” across the Internet, in an on-demand, pay-per-use manner

Cloud Services



- SaaS – “Software as a Service” but from Cloud
 - Google Apps™, Apple MobileMeSM,
- PaaS – (application) Platform as a Service
 - Amazon® Machine Instance (AMI)
- Storage as a Service (no acronym)
 - Amazon S3™, Amazon SimpleDB™, Google Base™
- IaaS – Infrastructure as a Service
 - Amazon Elastic Compute Cloud (Amazon EC2®), VMWare vCloud™

Google Apps and Google Base are trademarks of Google, Inc. in the United States and/or other countries. Apple MobileMe is a service mark of Apple, Inc. in the United States and/or other countries. Amazon, Amazon S3, Amazon SimpleDB, and Amazon EC2 are registered trademarks or trademarks of Amazon Technologies, Inc. in the United States and/or other countries. VMWare vCloud is a trademark of VMWare, Inc. in the United States and/or other countries.

Cloud Features



- Highly available, usually 99 percent or greater
- Utility per use billing
- Service credit for error/outage
- No long-term commitment required
- Large pool of reliable resources
- Easily and highly scalable
- Dynamic load balancing
- Aggregated and scalable virtual servers
- Multiple Internet providers/gateways
- Multi metadata center and offsite redundancy

Cloud Benefits



- Economies of scale will cross all markets
- Opens up large-scale enterprise services to small businesses that could not afford their own enterprise
- So cheap, large business can not ignore it
- Opens new low-risk experiment and prototype area to businesses. So cheap you can try then justify
- Allows business to handle surge storage and computing problems
- Allows IT to integrate faster and be more responsive to its customers

Mobile Platforms



- Smart phones
- Consumer readers
- Business readers

Overview



- Surveillance
- Event models
- Emerging Web technologies
- **Putting it all together**

Information Needs



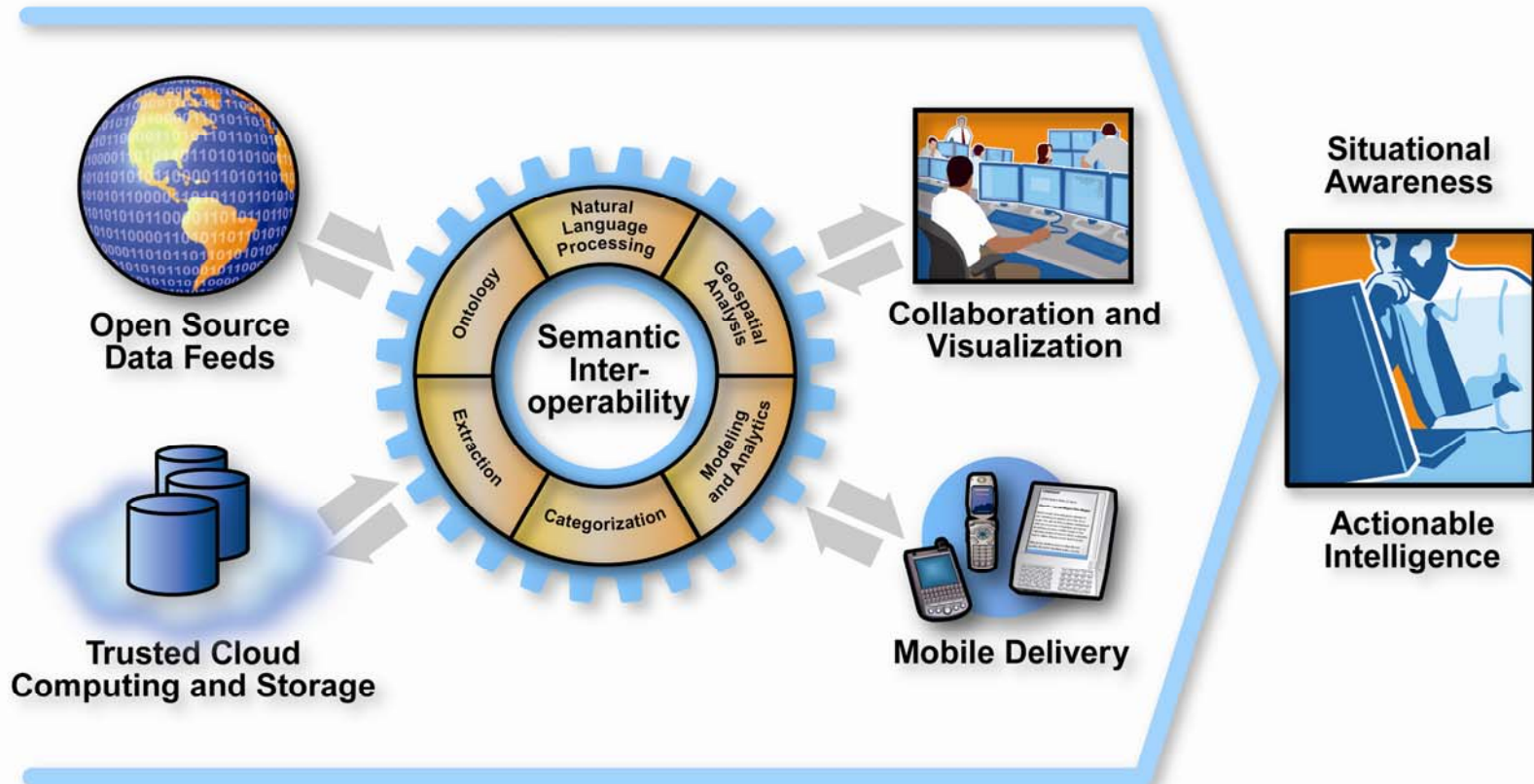
- Understanding what external information exists (about your organization and domain)
- Aggregating and analyzing against internal information
- Aggregating, integrating and analyzing all external information that directly impacts an internal project
- Identifying, aggregating and integrating regulated documents as part of a large-scale document management and migration system

Analyst Needs to Harvest Open Source Information



- Search engines are not enough
- Filter and triage information for analysts
- Integrate internal and external data
- Store analysis and vetted results
- Collaborative environment for knowledge workers
- Tracking of events
- Situational awareness reporting
- Dissemination to a mobile workforce

SAIC Open Source Analytic Framework (OSAF)



Correspondence: gradyn@saic.com