

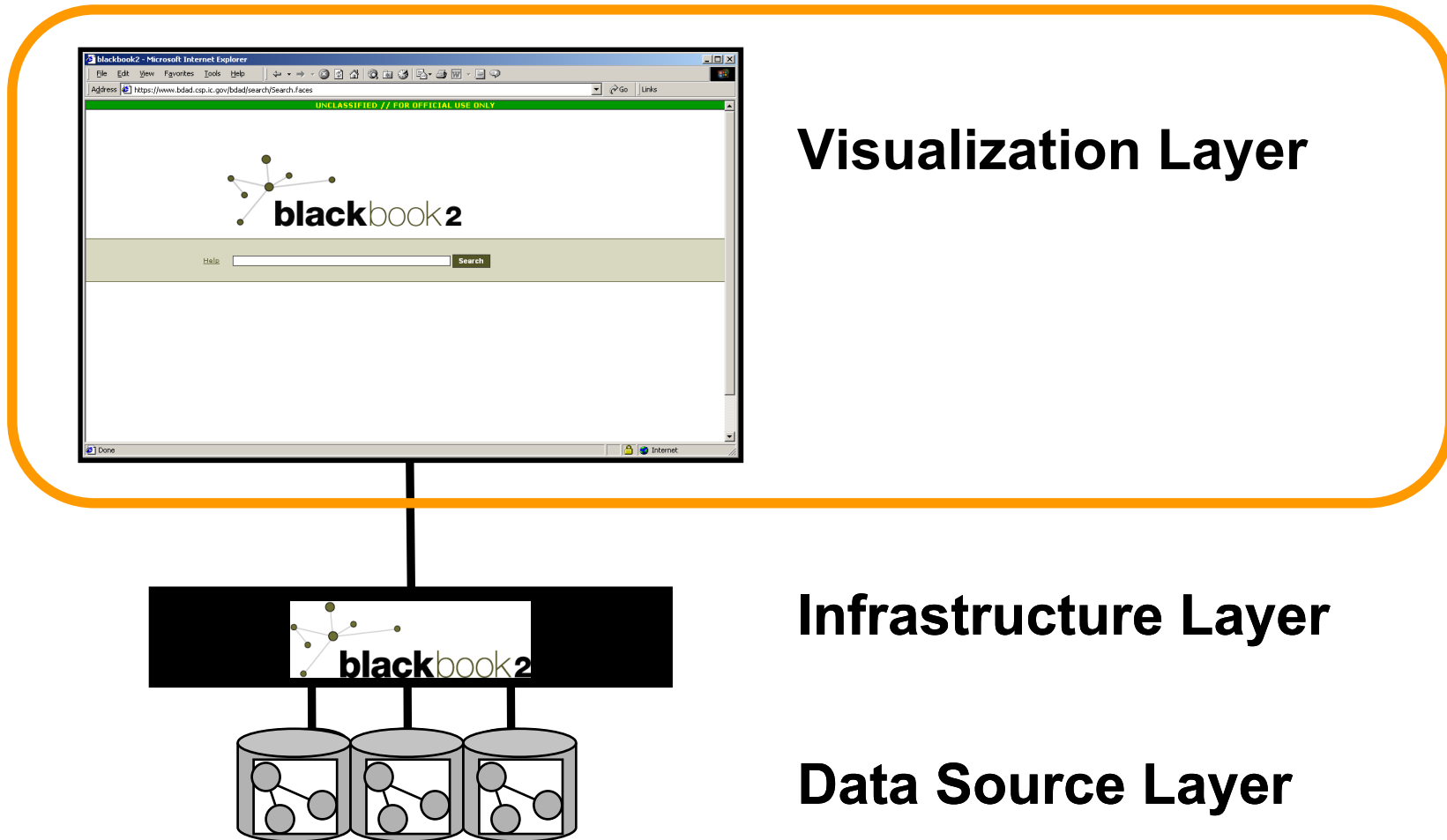


# blackbook2

# Overview

- Blackbook2 is a J2EE server-based data integration framework
- Relies on open standards to promote robustness and interoperability
  - JENA, JUNG, Lucene, JAAS, D2RQ
- Based on semantic web technologies
  - RDF, RDF Schema, OWL, SPARQL
  - Vocabulary agnostic
- Provides a default web application interface, SOAP and RESTful interfaces
- Blackbook2 is PL3 Appendix E certified (PL3+)

# Architecture

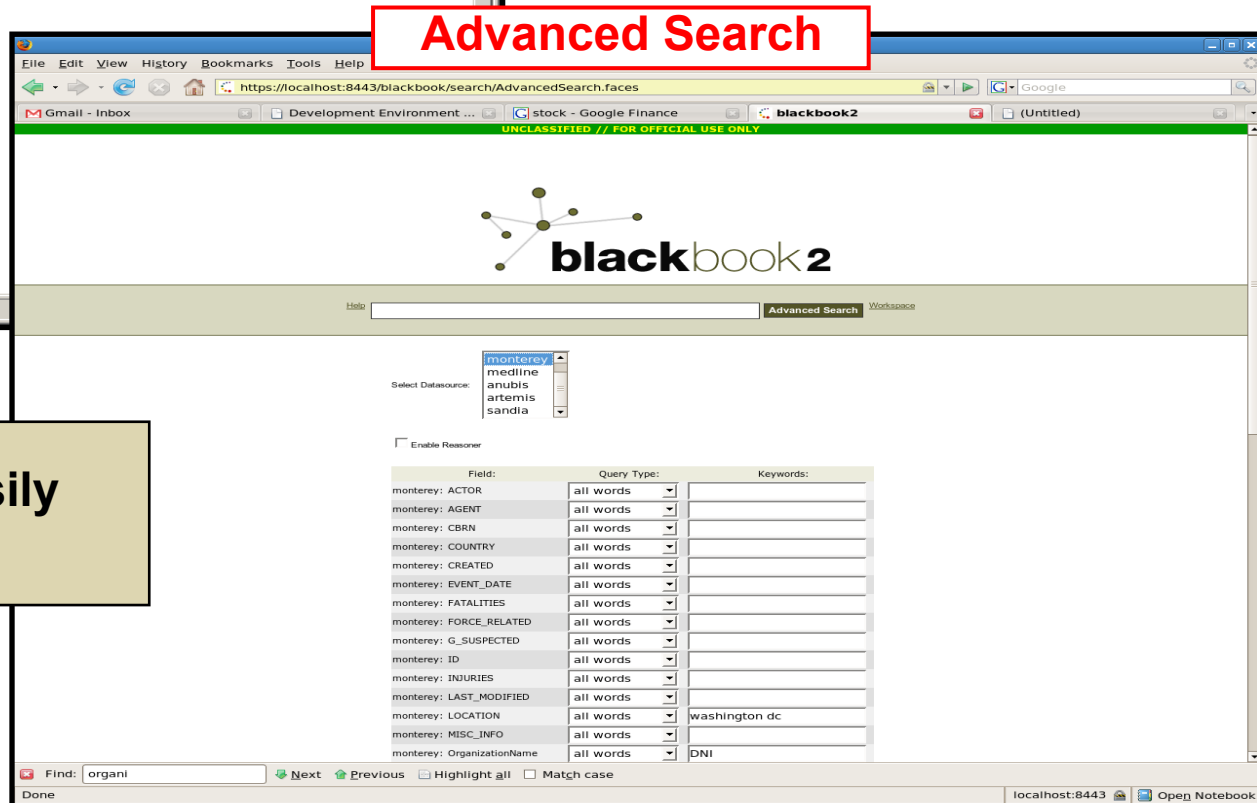
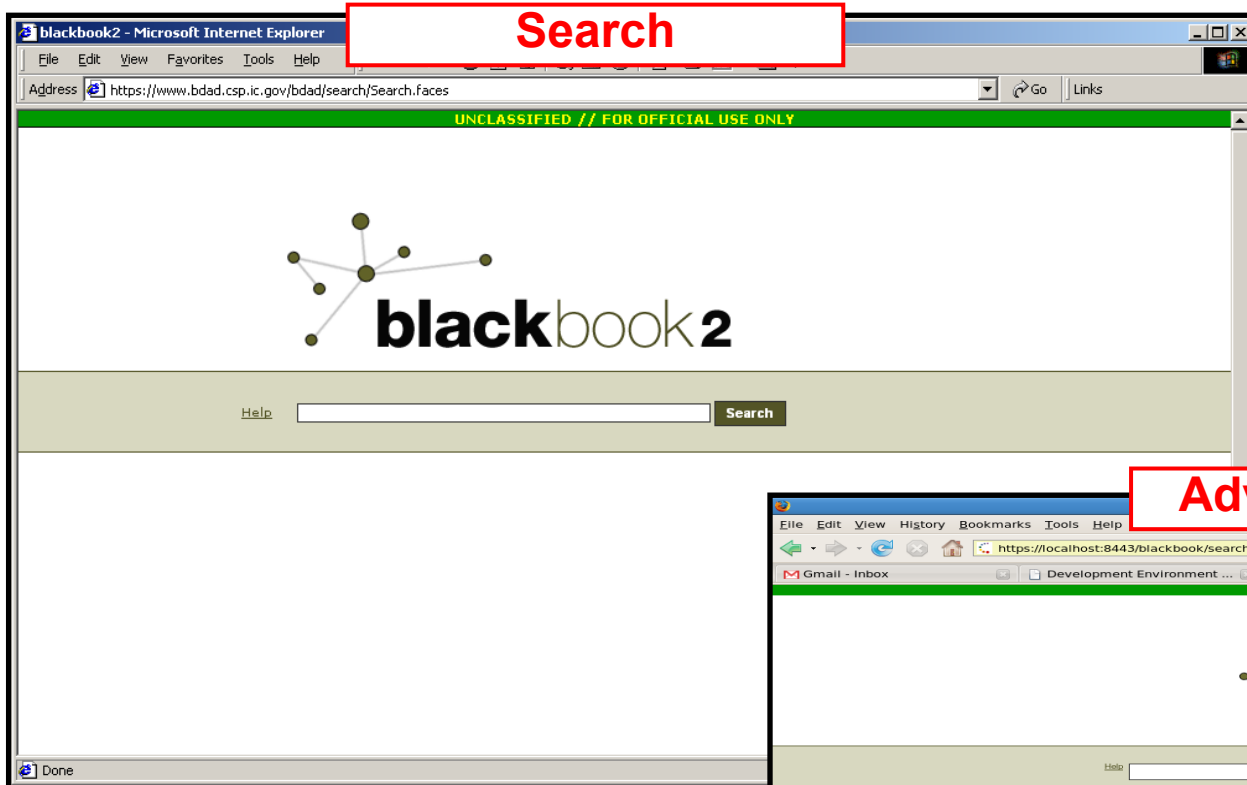


**Visualization Layer**

**Infrastructure Layer**

**Data Source Layer**

# User Interface



A front-end "Google-like" user interface allows analysts to easily perform keyword and attribute based searches.

# User Interface

**Google-like Results**

The screenshot shows the blackbook2 web interface in Mozilla Firefox. The search term 'anthrax' is entered in the search bar. The results are displayed in a list format, with each item having a 'Detail' link. The results include:

- Aum Shinrikyo: Once and Future Threat?** Aum Shinrikyo began its public campaign of terror on June 27, 1994. On that Monday in Matsumoto, a city of 300,000 population 322 kilometers northwest...
- A Poisonous Plot** Watching the police officers come and go, some of them in protective white suits and masks, and seeing the long hours they spent in the top-floor apar...
- The missing pieces** Al-Qaeda commander Abu Mussab al-Zarqawi needed treatment for a shattered leg that was injured, apparently, during the American bombing raids on Afgha...
- La Victoria** Suspected National Liberation Army (ELN) guerrillas killed one policeman and injured another policeman and a...
- Sitra** Arsonists set fire to a store in Sitra, killing a Bangladeshi and injuring another. Shia extremists are suspected.
- Aikimbayev** Aikimbayev
- Atshabar** Atshabar

**Network**

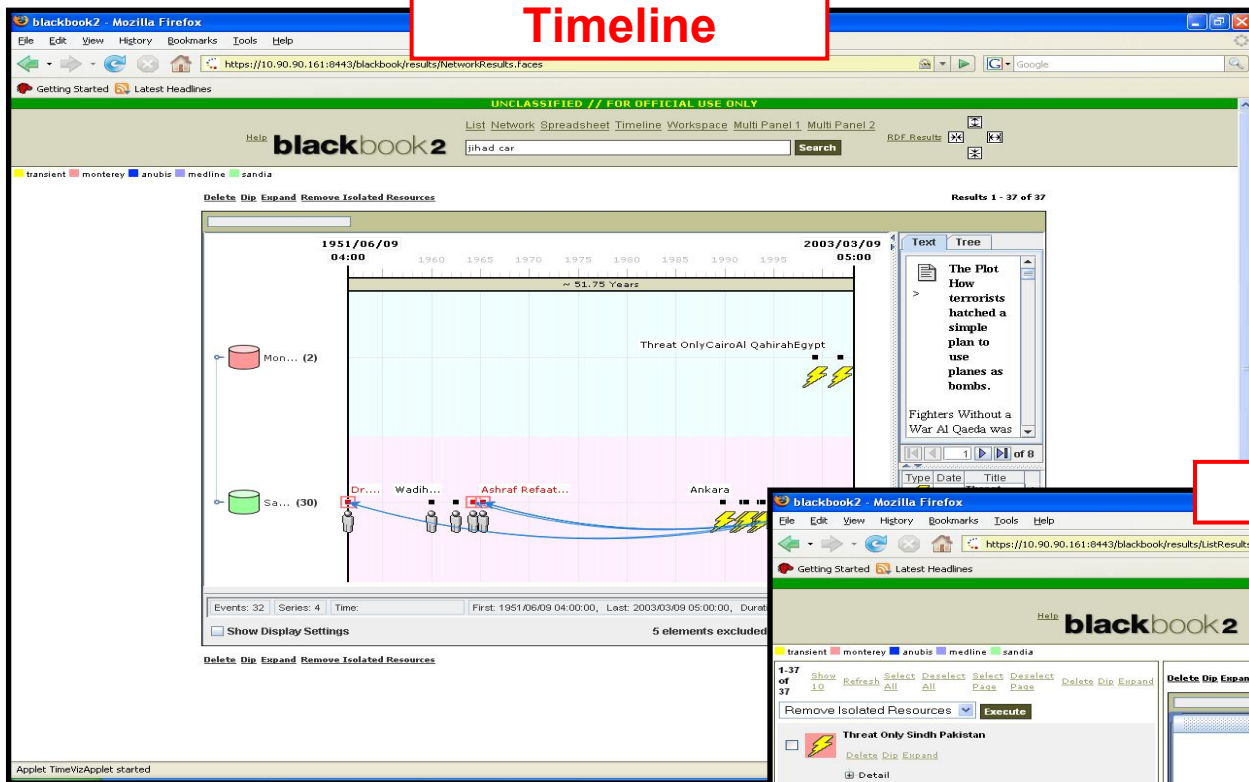
The screenshot shows the blackbook2 web interface in Mozilla Firefox. The search term 'jihad car' is entered in the search bar. The results are displayed in a network diagram format, showing relationships between various entities. The entities include:

- Ankara
- Cairo
- Wadh El Hage
- Background: Abu Omar
- Arabic Veterans of the Afghan War
- Does Bin Laden pose a Threat to Israel?
- Terrorism against Islamic Organizations and Groups
- The 'Afghan Alumni'
- MAK - Services Office International Islamic Front for Jihad Against the Jews and Crusaders
- Dr. al Ayman Mohammed Rabih Zawahir
- Hamball
- Ashraf Refaat Nabith Henin
- The CEO of al-Oaeda: Khaled Sheikh Mohammed
- Bold Tracks of Terrorism's Mastermind
- THE MAN BEHIND BIN LADEN: How an Egyptian doctor became a master of terror.
- The Plot How terrorists hatched a simple plan to use planes as bombs.
- attention turns to the other prime suspect
- The Shoe Bomber's World
- The Demise of Radical Islam in Turkey
- OSIS man to al-Qaeda's Master
- Ottawa
- ISIS

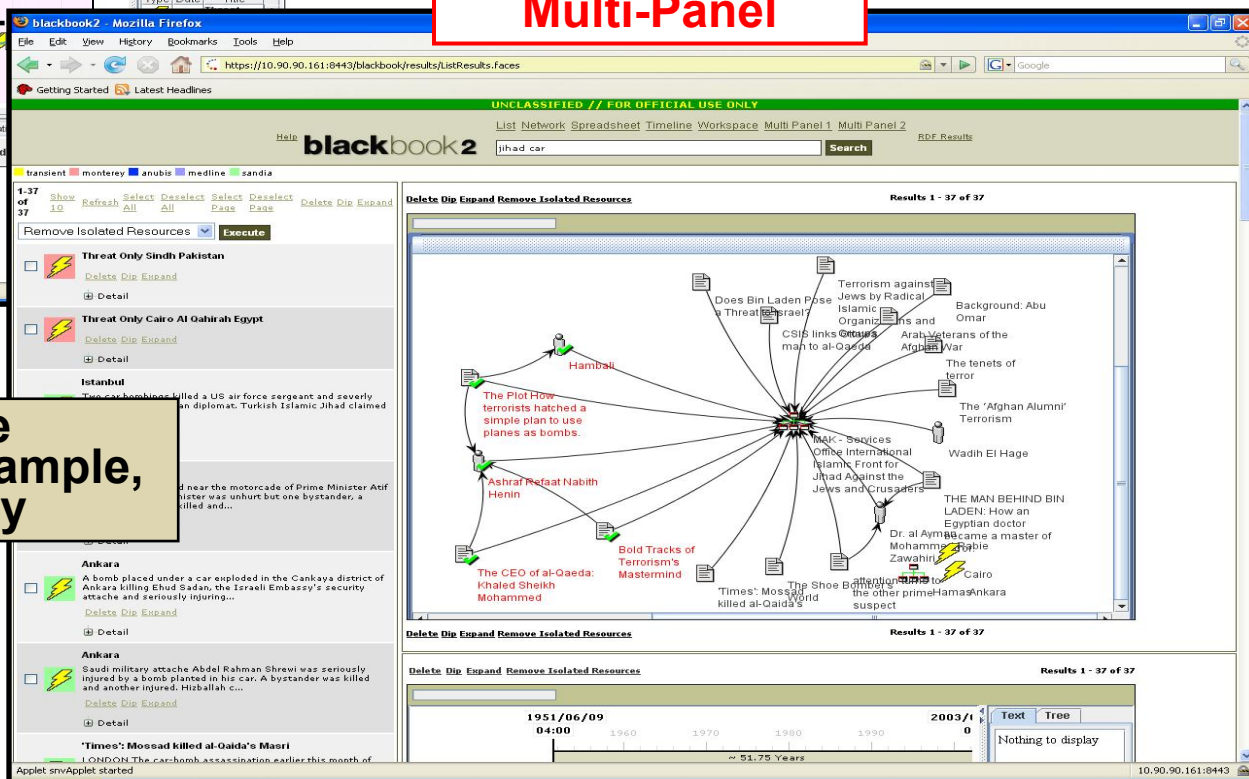
Different ways to view the same information. "Network", for example, displays entities of different types and their relationships to other entities.

# User Interface

**Timeline**

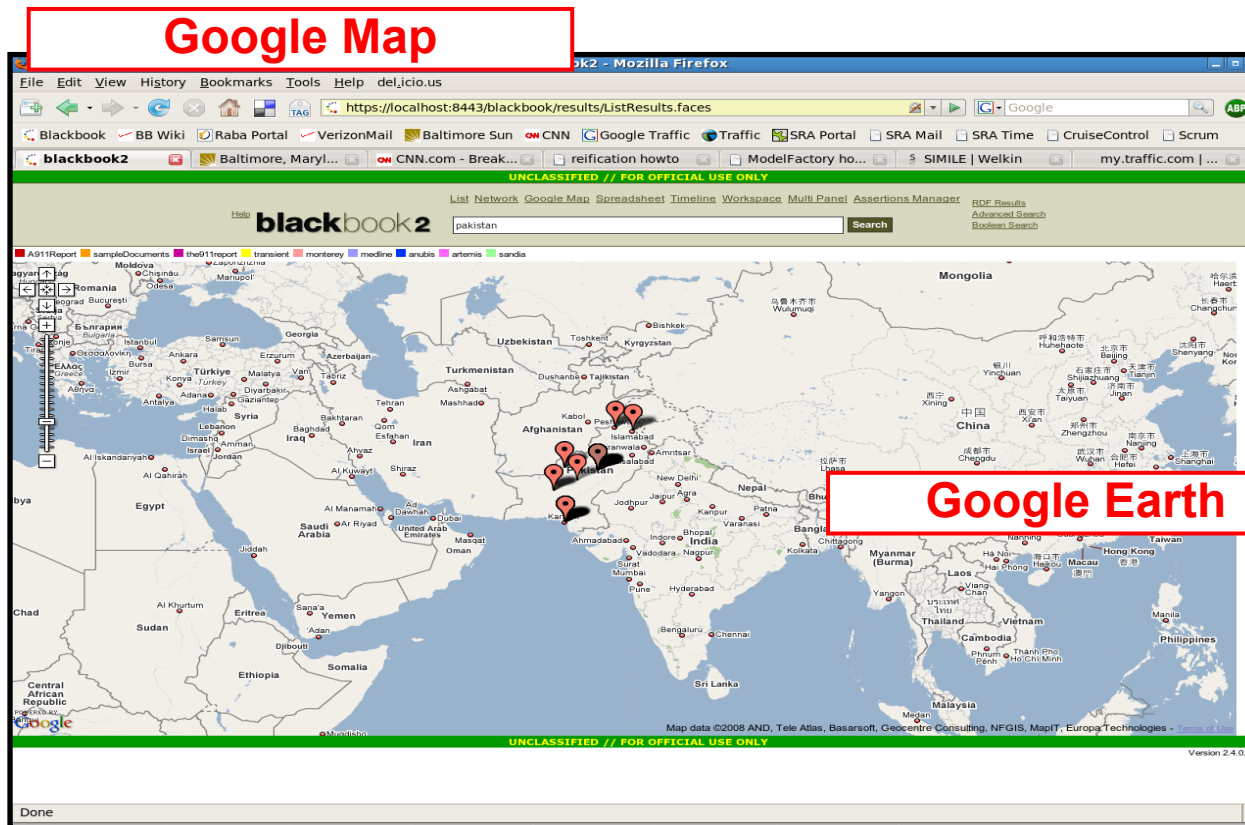


**Multi-Panel**



Different ways to view the same information. "Timeline", for example, displays entities chronologically

# User Interface



**Allows analysts to visualize geospatial content using Google-map and Google Earth.**

# User Interface

## Analyst Notebook

The Analyst Notebook interface displays a complex network diagram with numerous nodes and connecting lines, overlaid on a web browser window. The browser address bar shows the URL: <http://www.i2inc.com/public/sides/viewBigger.pl>. The interface includes a menu bar (File, Edit, View, Insert, Format, Tools, Analysis, Data, Window, Help) and a toolbar with various icons. The network diagram consists of many nodes, some represented by icons like houses and cars, connected by lines. A red box highlights a specific area of the network. The bottom of the window shows the text "i2 : Investigative Analysis Software".

## Mediawiki

The Mediawiki interface shows a page titled "Anthrax Events" with a table of contents and a section on "Operational Issues". The page content includes:

**Operational Issues**

- 1.1 The Value of Pre-Existing Relationships
- 1.2 Surge Capacity
- 1.3 Incident Command System vs. Bureaucratic Paradigm
- 1.4 Leadership and "Authority"

**The Value of Pre-Existing Relationships**

In the setting where information was either flowing rapidly and without control, or was not flowing at all, where command and control structures were changing rapidly as more and more agencies responded, and where there was a void of leadership structure, the vital need for credible and reliable information was essential. The value of pre-existing and often informal relationships and communication structures cannot be stressed enough. Information was exchanged frequently through midlevel communication chains that had been well vetted by planning and exercising together in the pre-event stages. These pre-existing channels served to bypass the obstacles of the time-consuming bureaucratic review processes and to bypass the need for instant reliability and credibility checks that go into the introductory phases of meetings or telephone calls. We spoke to people we knew as trusted colleagues in channels that were well rehearsed to get the information and get it quickly.

The diagram illustrates relationships between individuals (Derring, Fredericks, Smith, Evans, Atta, Jones, Simpson) and locations (capitol, anthrax, Washington). Arrows indicate connections between these entities.

This is not to say, however, that these informal relationships should usurp formal hierarchy but should be integrated into a command and control structure that relies on established lines of authority with well-defined roles and responsibilities at all levels. The exercise and use of this command structure by all levels will minimize the need to rely on informal communication structures.

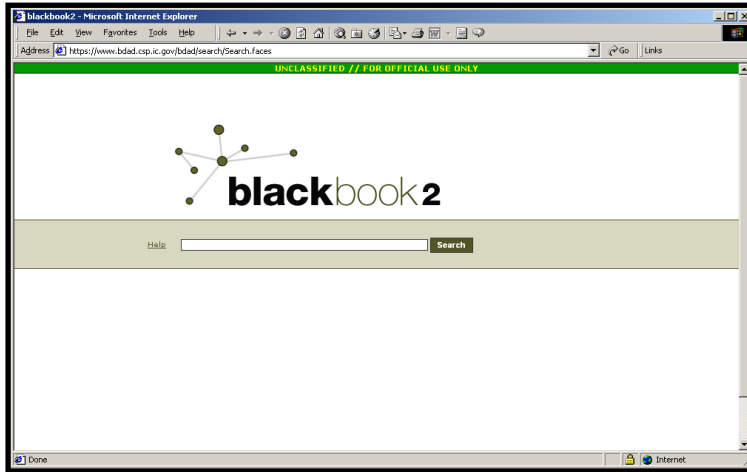
**Surge Capacity**

In Maryland, response by public health was possible through the utilization of personnel and resources diverted from other programs within the state and local health departments. A small cadre of personnel was quickly trained, advised about new roles and responsibilities, including, for many, altered work schedules, and put to work on response-related activities within 1-2 hours, all at the expense of their usual work projects. While a response was rapidly mounted, under these conditions, it would not have been sustainable for long. The Department of Health and Mental Hygiene was able to maintain such an effort for roughly 1-2 weeks before significant strains were felt in those programs from which personnel were drawn. Development of a robust workforce by increasing absolute numbers, as well as cross training of personnel and expanding that cadre from which to draw is essential for future surge capacity. The services provided by the department of health include, in Maryland, the state laboratory. In the case of laboratory, surge space as well as personnel surge capacity became a critical issue because of the demands for specialized laboratory physical plant requirements.

Object	Type	Properties
Smith	Person	firstname: Joe lastname: Smith location: 18th Street NW date: 2001-09-16
Jones	Person	firstname: Sam lastname: Smith



# Architecture



**Visualization Layer**



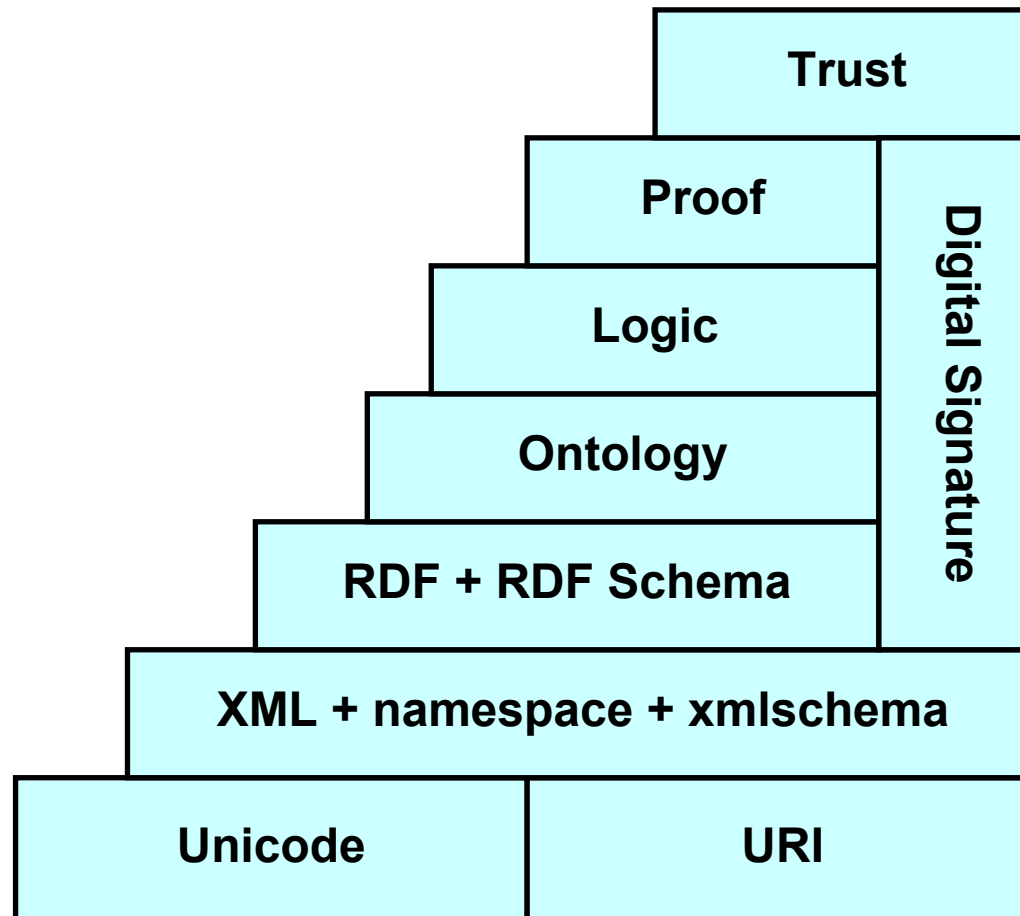
**Infrastructure Layer**



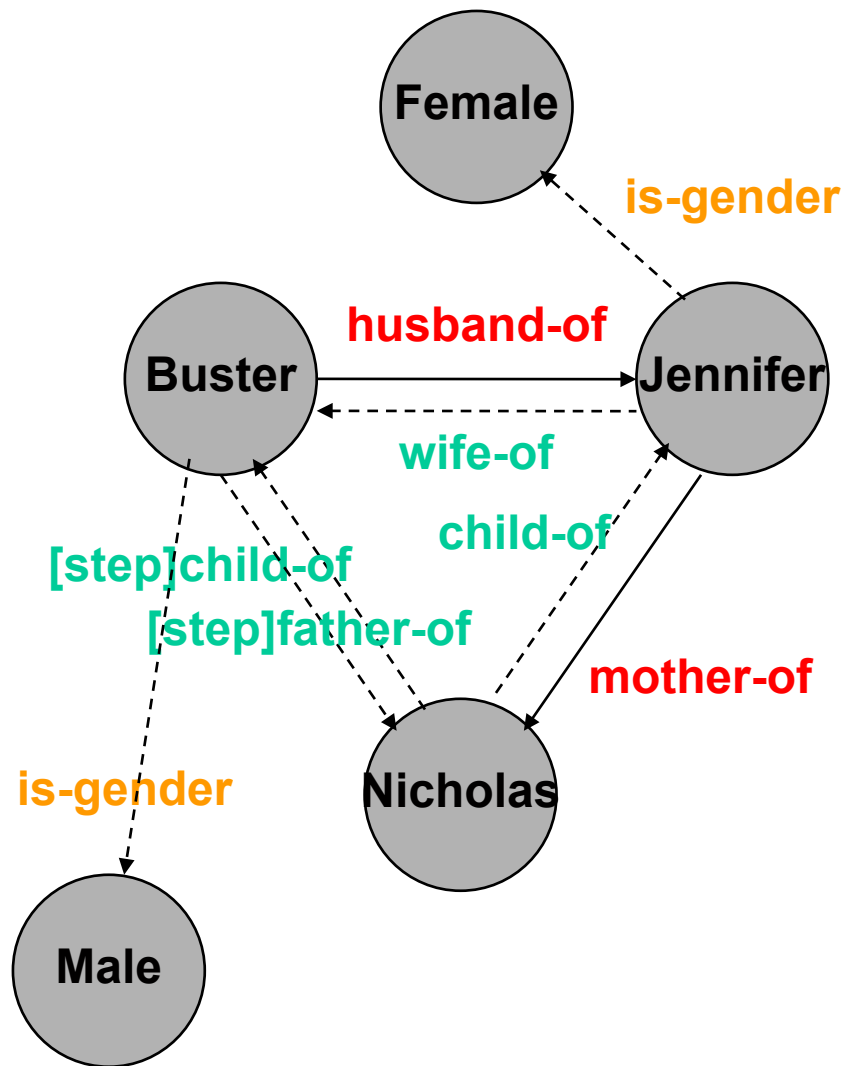
**Data Source Layer**

# Semantic Web

- The Semantic Web is the next generation of the current web in which computers can interpret the meaning of the web content because of explicit semantics provided in markup.



# Example 1: Inference



An analyst creates:

- 1) Entity "Buster"
- 2) Entity "Jennifer"
- 3) Entity "Nicholas"

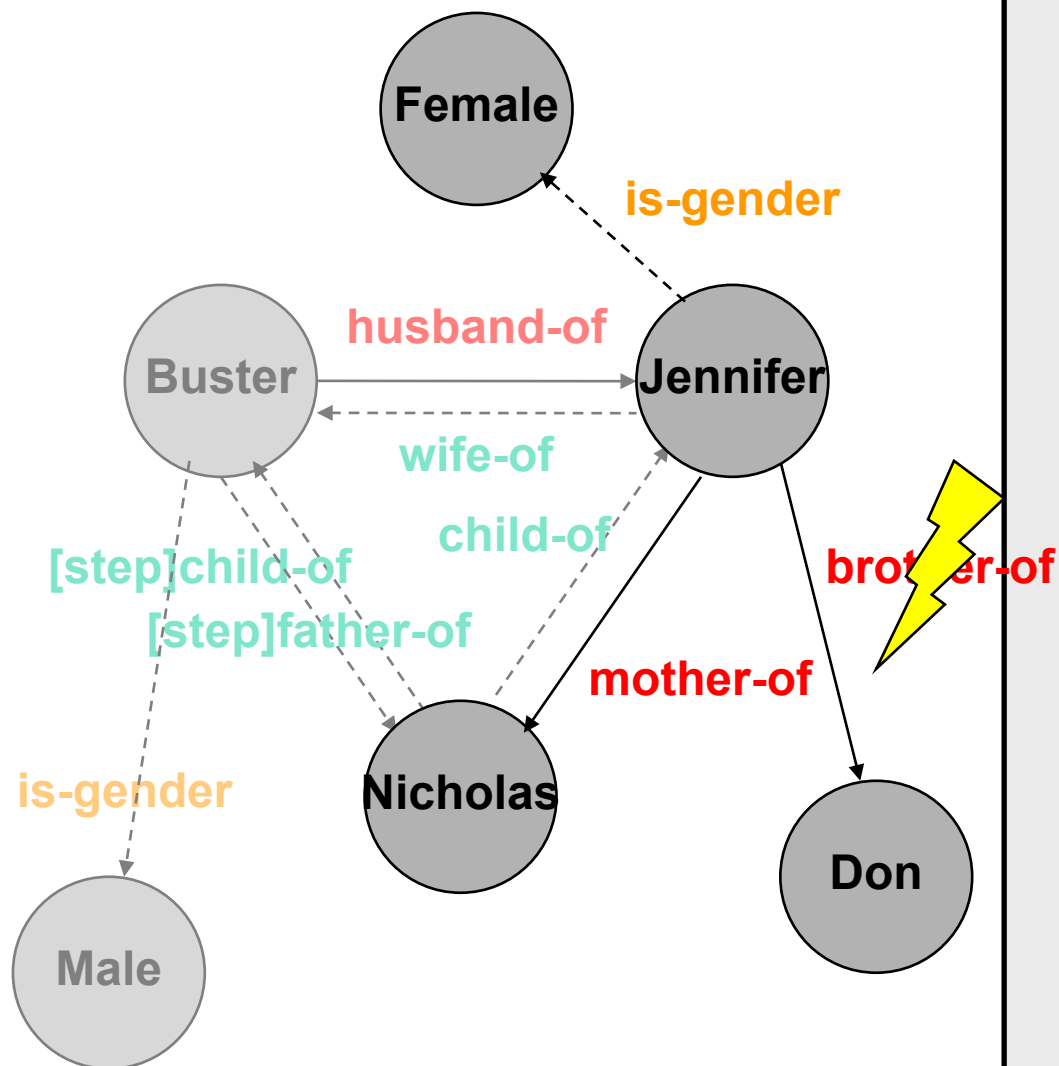
An analyst makes the assertion:

- 4) "Buster **husband-of** Jennifer"
- 5) "Jennifer **mother-of** Nicholas"

Blackbook system can infer:

- 6) "Jennifer **wife-of** Buster"
- 7) "Nicholas **child-of** Jennifer"
- 8) "Buster **[step]father-of** Nicholas"
- 9) "Nicholas **[step]child-of** Buster"
- 10) "Buster **is-gender** Male"
- 11) "Jennifer **is-gender** Female"

# Example 2: Invalid Logic Assertion



An analyst creates:

1) Entity "Don"

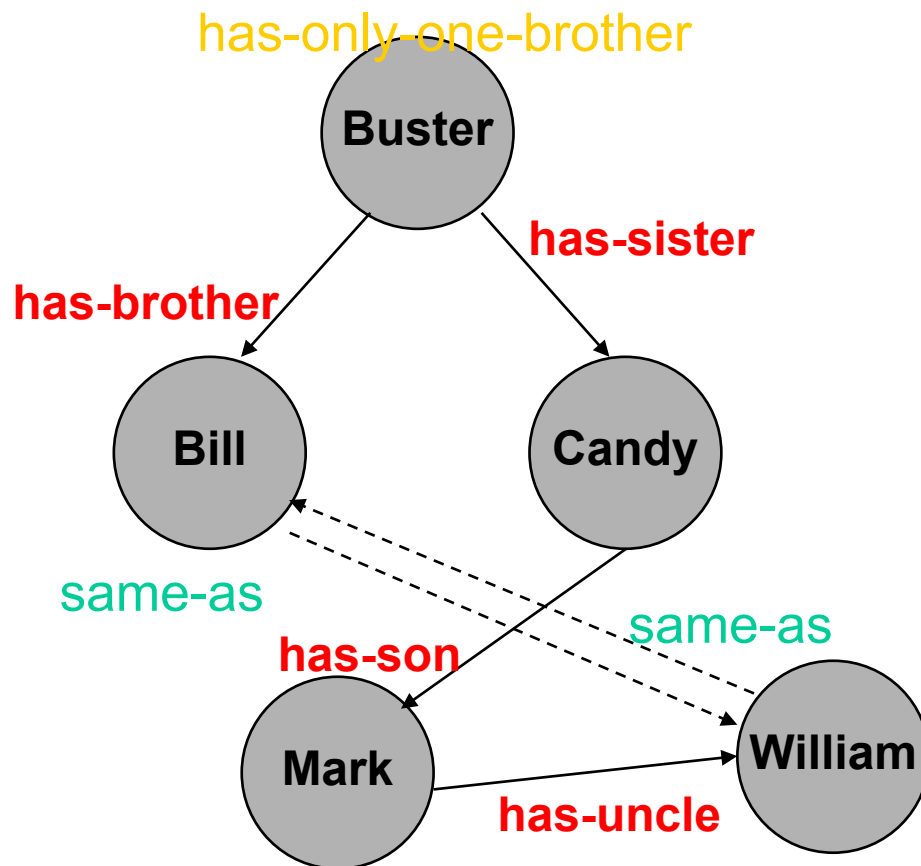
An analyst makes the assertion:

2) "Jennifer **brother-of** Don"

Blackbook system can infer:

3) Invalid Assertion  
(Gender conflict)

# Example 3: Constraints & same-as



An analyst makes the assertion:

- 1) "Buster **has-brother** Bill`
- 2) "Buster **has-sister** Candy`
- 3) "Candy **has-son** Mark`
- 4) "Mark **has-uncle** William`

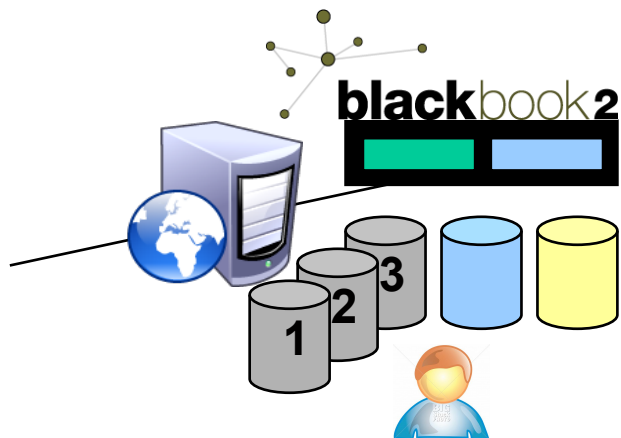
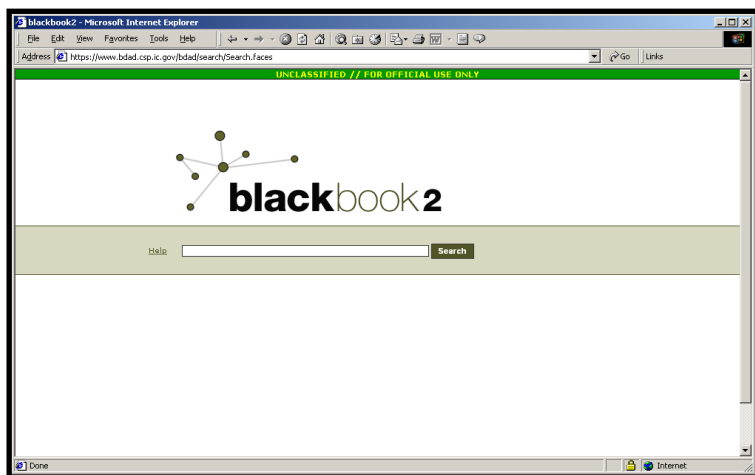
An analyst applies the constraint:

- 5) "Buster **has-only-one-brother** Bill`

Blackbook system can infer:

- 6) "William **same-as** Bill`
- 7) "Bill **same-as** William`

# Algorithms, Security, AKB



Algorithm plug-ins can be added



Security PL3+ / User Credentials



CASPORT  
Scattered Castles  
DIAS

Analysts can store assertions into an Analytic Knowledge Base (AKB)



# User Interface

## Workflow

The screenshot shows the 'blackbook2' web application interface. At the top, there's a navigation bar with 'blackbook2' and various menu options. Below that, there are several panels:

- Algorithms:** A list of available algorithms: Dip, Expand, Jena Keyword, Lucene Keyword, and Materialize.
- Process Diagram:** A visual flowchart showing a sequence of steps: '1. Lucene Keyword' connects to '2. Materialize', which then connects to '3. Dip'. There are 'Refresh' and 'Clear' buttons above the diagram.
- Process Flow:** A table for configuring the process flow. It has columns for 'States', 'To States', and 'Additional Criteria'. It shows steps like '0. Expand', '1. Lucene Keyword', '2. Materialize', and '3. Dip' with various configuration options like 'DataAccess' and 'fork'.
- Populate:** A button to populate the process diagram.

## Yahoo Pipes

The screenshot shows the 'pipes' web application interface. It features a grid of pipes (tasks) connected in a sequence:

- Sources:** A list of available sources on the left sidebar, including 'Fetch CSV', 'Feed Auto-Discover', 'Fetch Data', 'Fetch Page', 'Fetch Site Feed', 'Flickr', 'Google Base', 'Item Builder', 'Yahoo! Local', and 'Yahoo! Search'.
- User inputs:** A section below the sources for user inputs.
- Pipes:** A grid of pipes connected in a sequence: 'Where (location)' feeds into 'URL Builder', which feeds into 'Fetch Feed'.

“Workflow` allow analysts to define the order of tasks, configure algorithm parameters, and batch processes concurrently

# User Interface

**Workspace**

Workspace (1)

- Workflow Process Definitions
  - anoibp
    - anoibp (1)
    - anthraxChemQuery (1)
      - 2434536128618620
      - chrisRock
      - habob
      - helloJoe
    - jimmyLewis (1)
    - lancingMichigan
    - lancingMichigan (1)
    - larou
    - MyNewProcess
    - MyNewProcess
    - MyNewProcess
    - MyNewProcess
    - tonyBraxton
    - xcarter
    - xcarter (1)

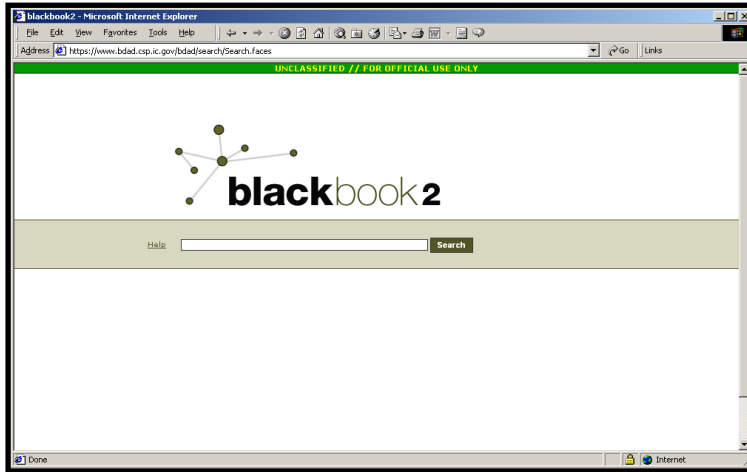
Instance	Start Date	Current Node	End Date
2434536128618620	May 15, 2007	finished	May 15, 2007 <a href="#">Results available</a>

Find: package Find Next Find Previous Highlight all Match case

“Workflow` and “Workspace` allow analysts to define the order of tasks, store them in private folders and/or share them publicly with colleagues.



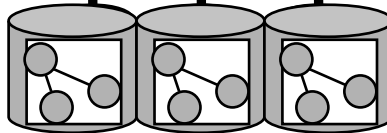
# Architecture



**Visualization Layer**



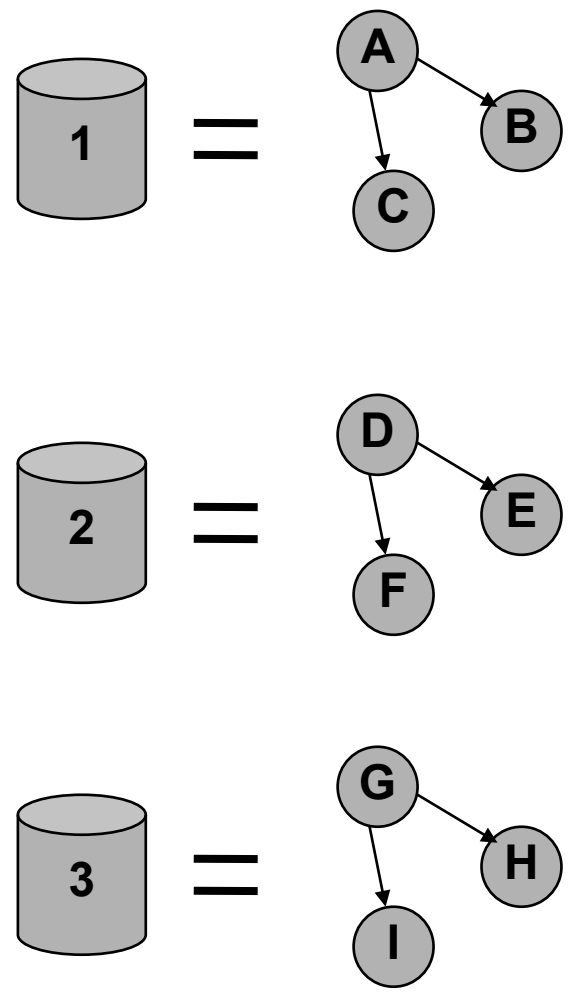
**Infrastructure Layer**



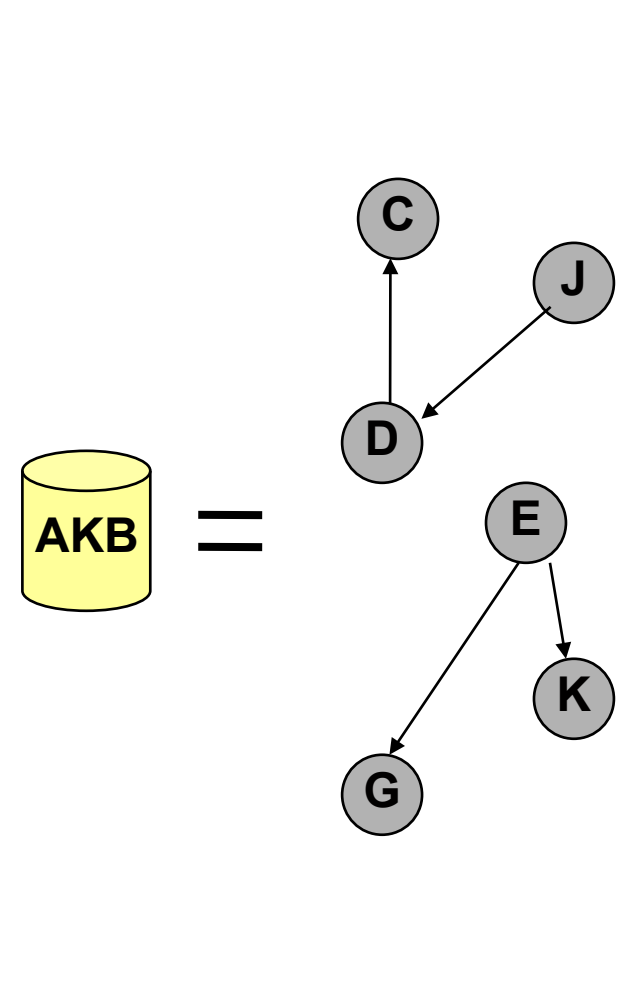
**Data Source Layer**

# Composite Knowledge

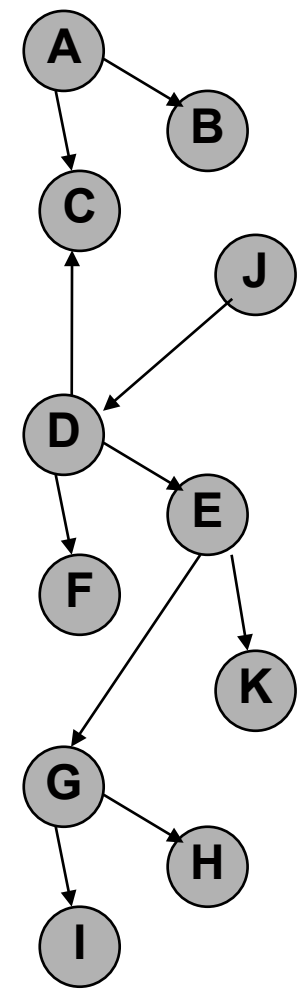
Original Datasource



Analyst Knowledge Base

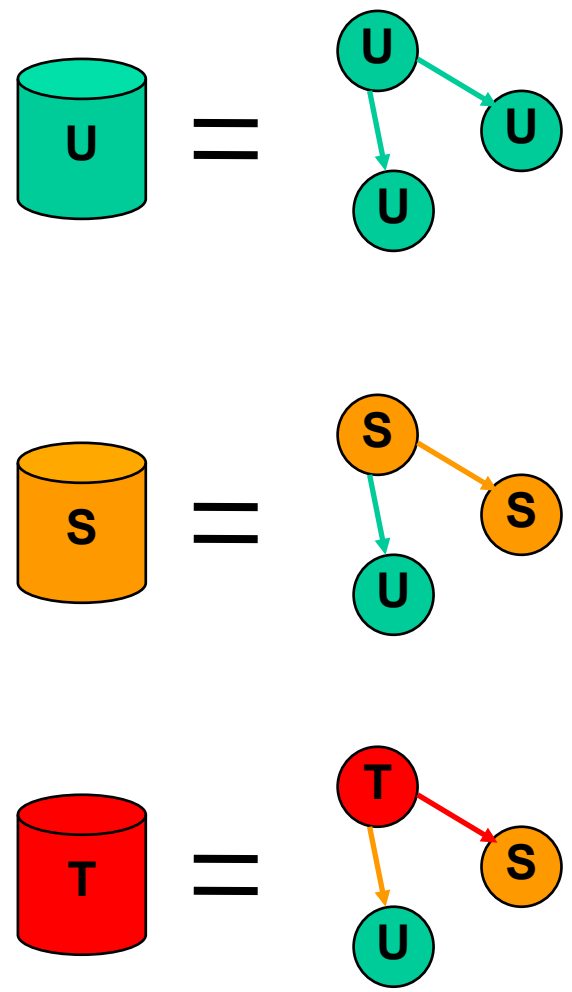


Composite Knowledge

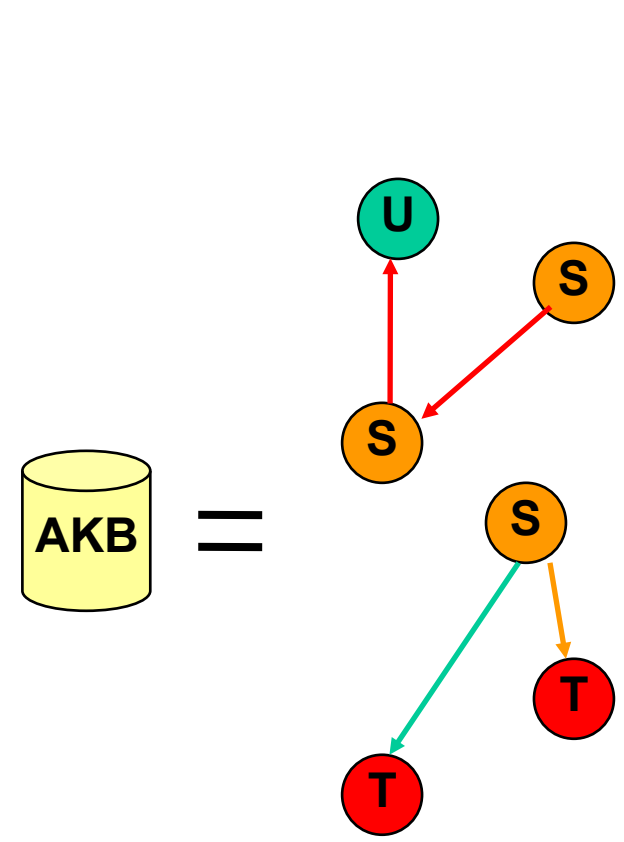


# Composite Knowledge with Security

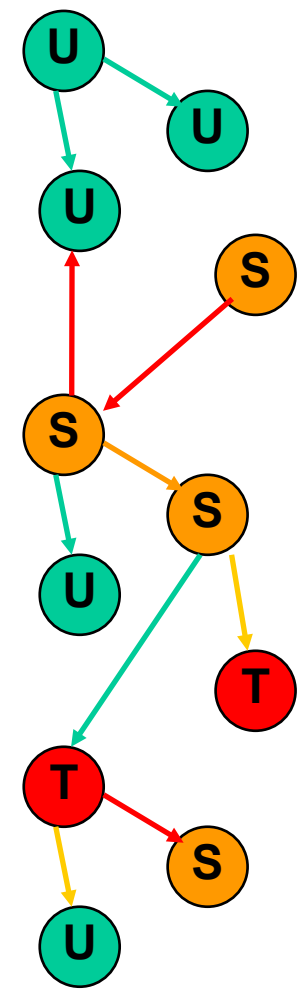
Original Datasource






Analyst Knowledge Base

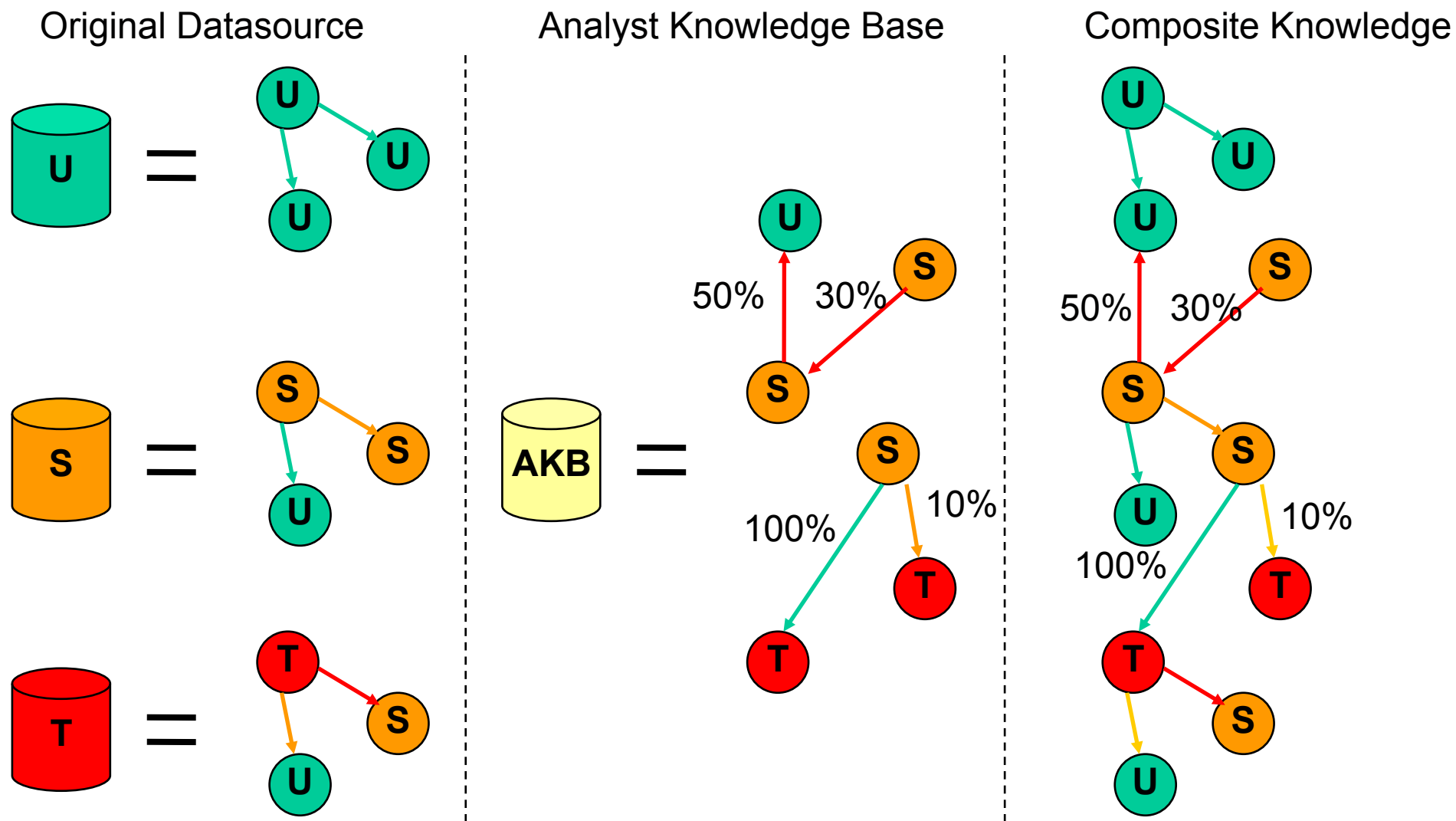


Composite Knowledge



	Unclassified
	Secret
	Top Secret

# Composite Knowledge with Confidence



# User Interface

**Relationship Manager**

Allows analysts to specify the relationship between two or more entities

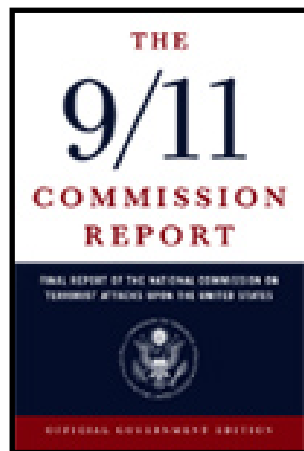
**Entity Manager**

Allows analysts to create entities of different types, and modify attributes

**Ontology Import**

Allows analysts to upload their own ontology

# Unstructured/Structured to RDF



## Unstructured

### THE ATTACK LOOMS

#### ALSO IN CALIFORNIA

described the Southeast Asia travels of Nawaf al Hazmi, Khalid al Mihdhar, and others in January 2000 on the first part of the "planes plot." In that chapter we also described how Mihdhar was spotted in Los Angeles in early in January 2000, along with associates who were not known to the FBI. Mihdhar was lost to sight when the group passed through Los Angeles on January 15, Hazmi and Mihdhar arrived in Los Angeles. They stayed in Los Angeles for a few weeks there before moving on to San Diego.<sup>2</sup>

#### Los Angeles

Mihdhar came to California, we do not know for certain. Khalid al Mihdhar (KSM), the organizer of the planes operation, explains that Los Angeles was a convenient point of entry from Asia and had the added benefit of being far away from the intended target area.<sup>3</sup>

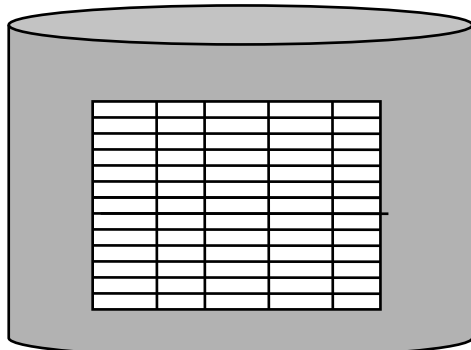
Mihdhar and Hazmi were ill-prepared for a mission in the United States. Their motivation for this plot were their devotion to Usama Bin Ladin, their determination to get valid U.S. visas. Neither had spent any time in the United States, and neither spoke much, if any, English.<sup>3</sup>

It is plausible that they or KSM would have tried to identify, in advance, a friendly contact for them in the United States. In detention, KSM denies that al Qaeda had any agents in Southern California. We do not credit this denial.<sup>4</sup> We believe it is unlikely that Hazmi and Mihdhar—neither of whom, in contrast to the Hamburg group, had any prior exposure to life in the West—would have come to the United States without arranging to receive assistance from one or more individuals informed in advance of their arrival.<sup>5</sup>

KSM says that though he told others involved in the conspiracy to stay away from mosques and to avoid establishing personal contacts, he made an

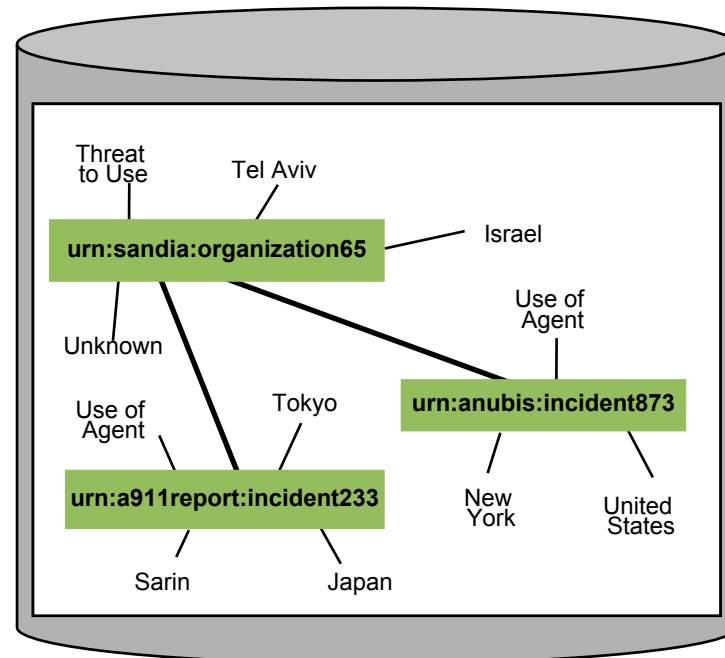
Translation

## Structured



RDMS/XML

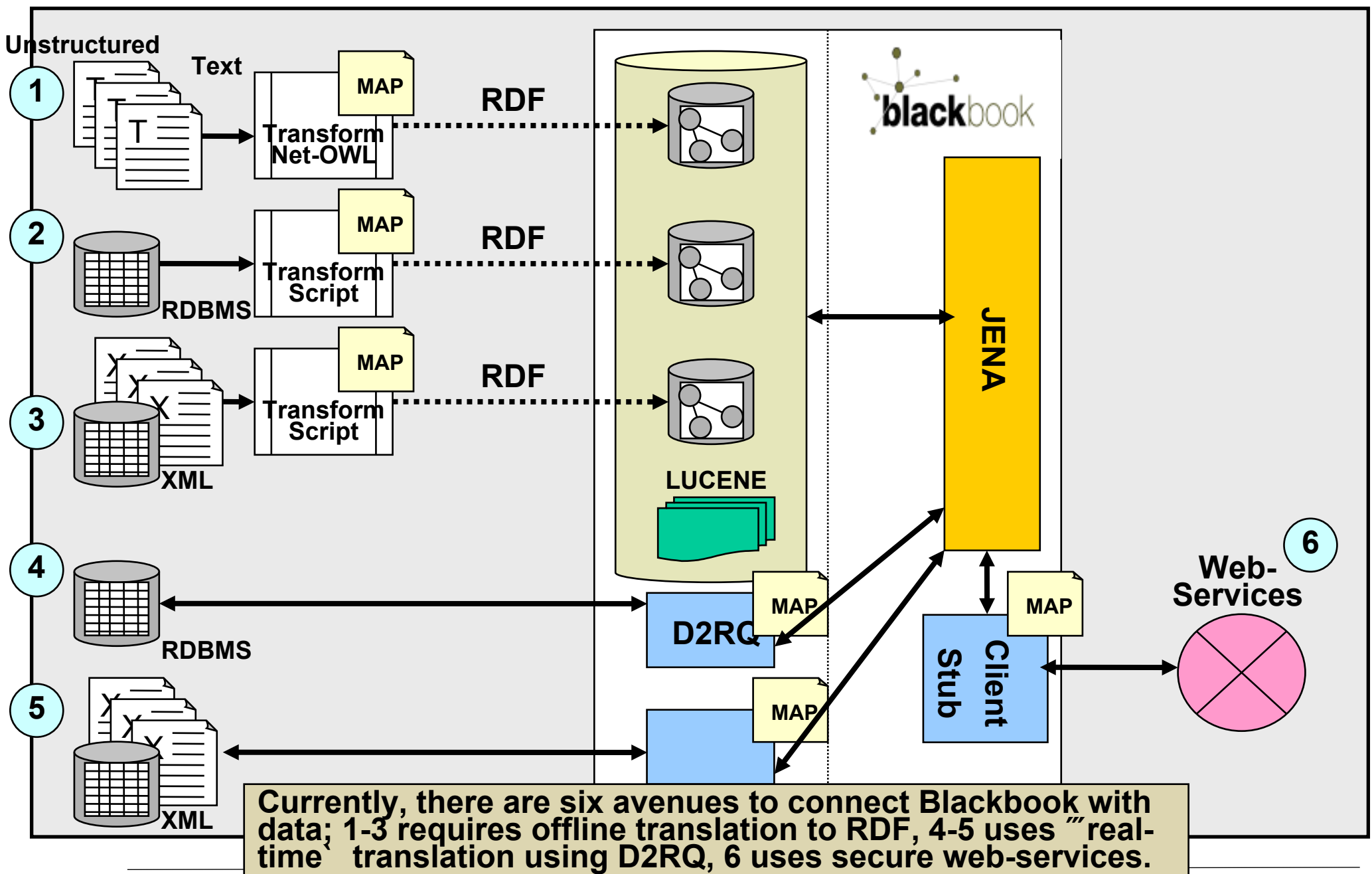
Translation



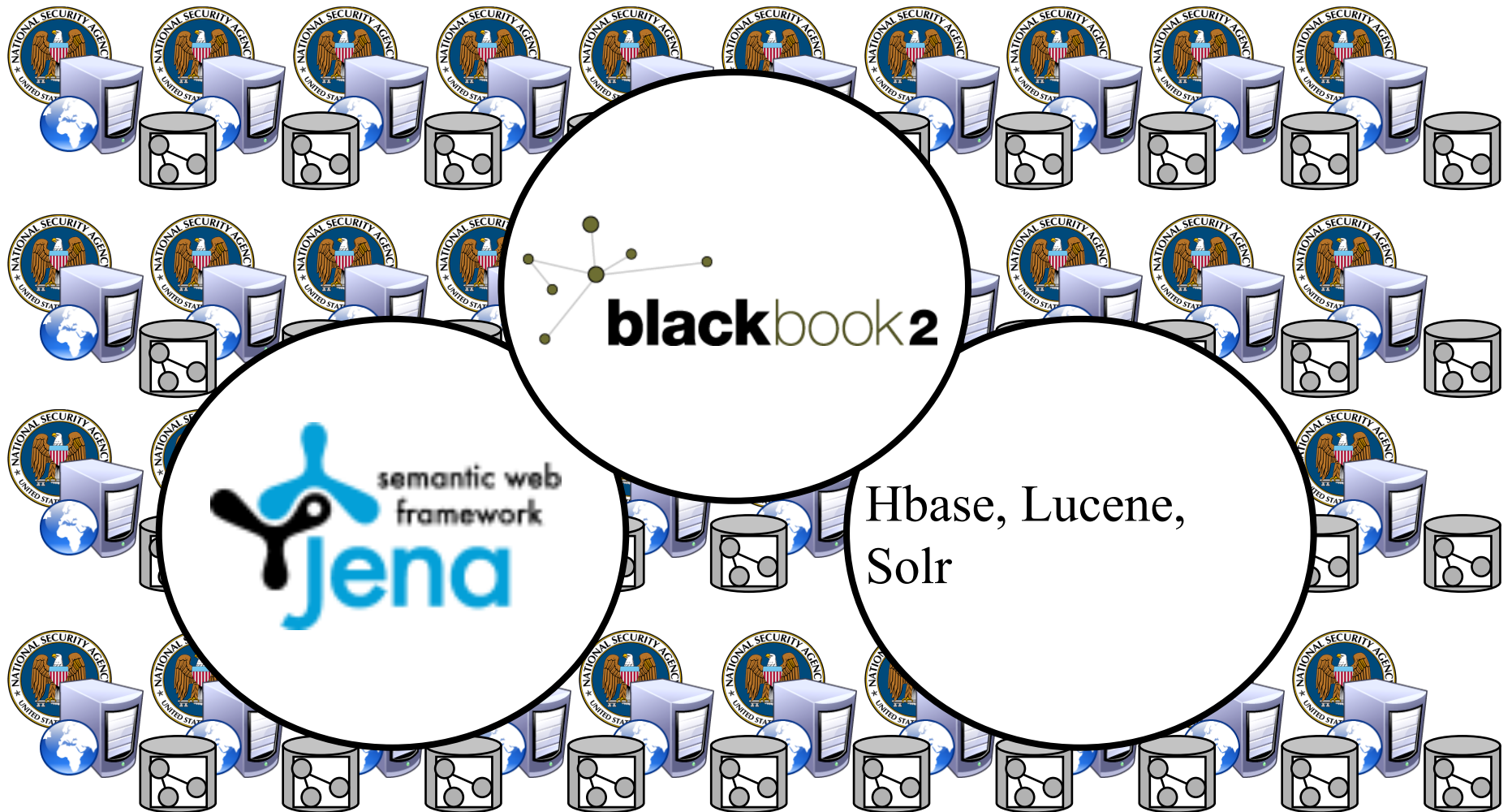
RDF

Unlike most applications, Blackbook performs queries on data in RDF form, not relational form.

# Datasource Connectivity



# Blackbook and Alternate Stores





# Scalability using Hadoop

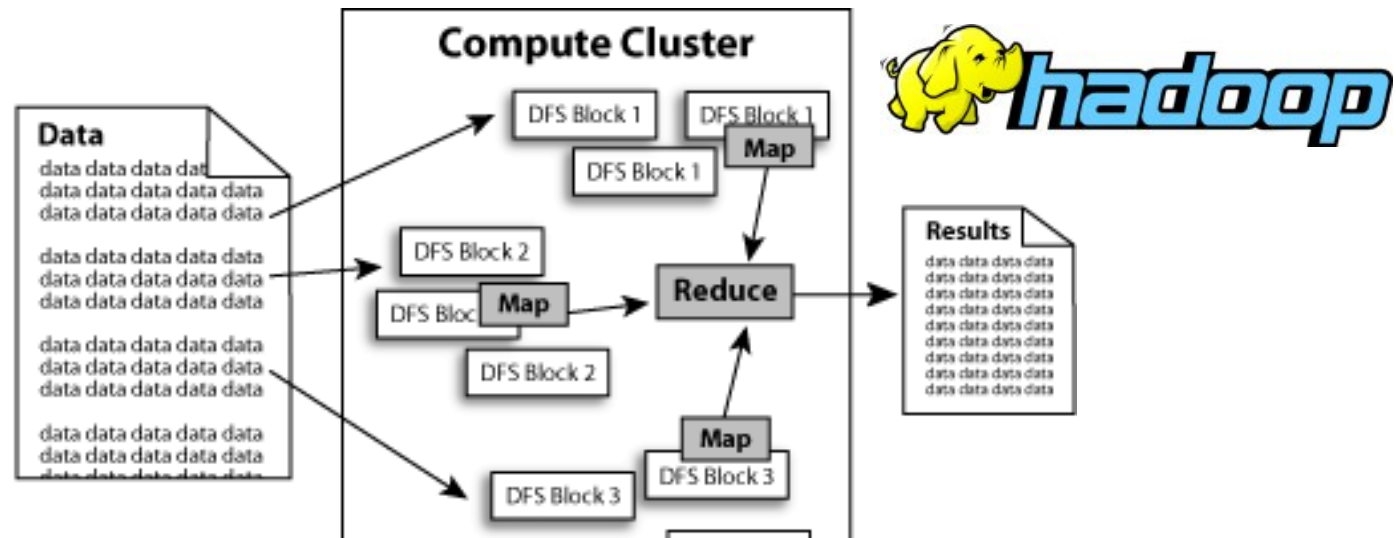
Hadoop implements MapReduce, using the Hadoop Distributed File System (*HDFS*). MapReduce divides applications into many small blocks of work. HDFS creates multiple replicas of data blocks for reliability, placing them on compute nodes around the cluster. MapReduce can then process the data where it is located.

**Scalable:** Hadoop can reliably store and process petabytes.

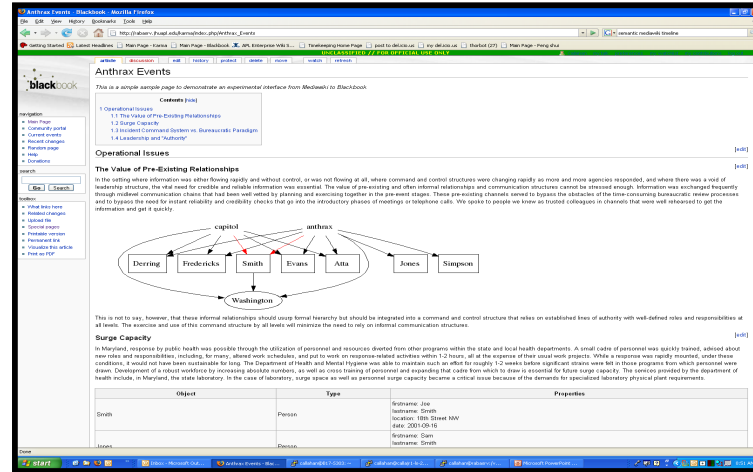
**Economical:** It distributes the data and processing across clusters of commonly available computers. These clusters can number into the thousands of nodes.

**Efficient:** By distributing the data, Hadoop can process it in parallel on the nodes where the data is located. This makes it extremely rapid.

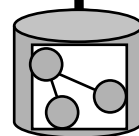
**Reliable:** Hadoop automatically maintains multiple copies of data and automatically redeploys computing tasks based on failures.



# Blackbook and Wikis

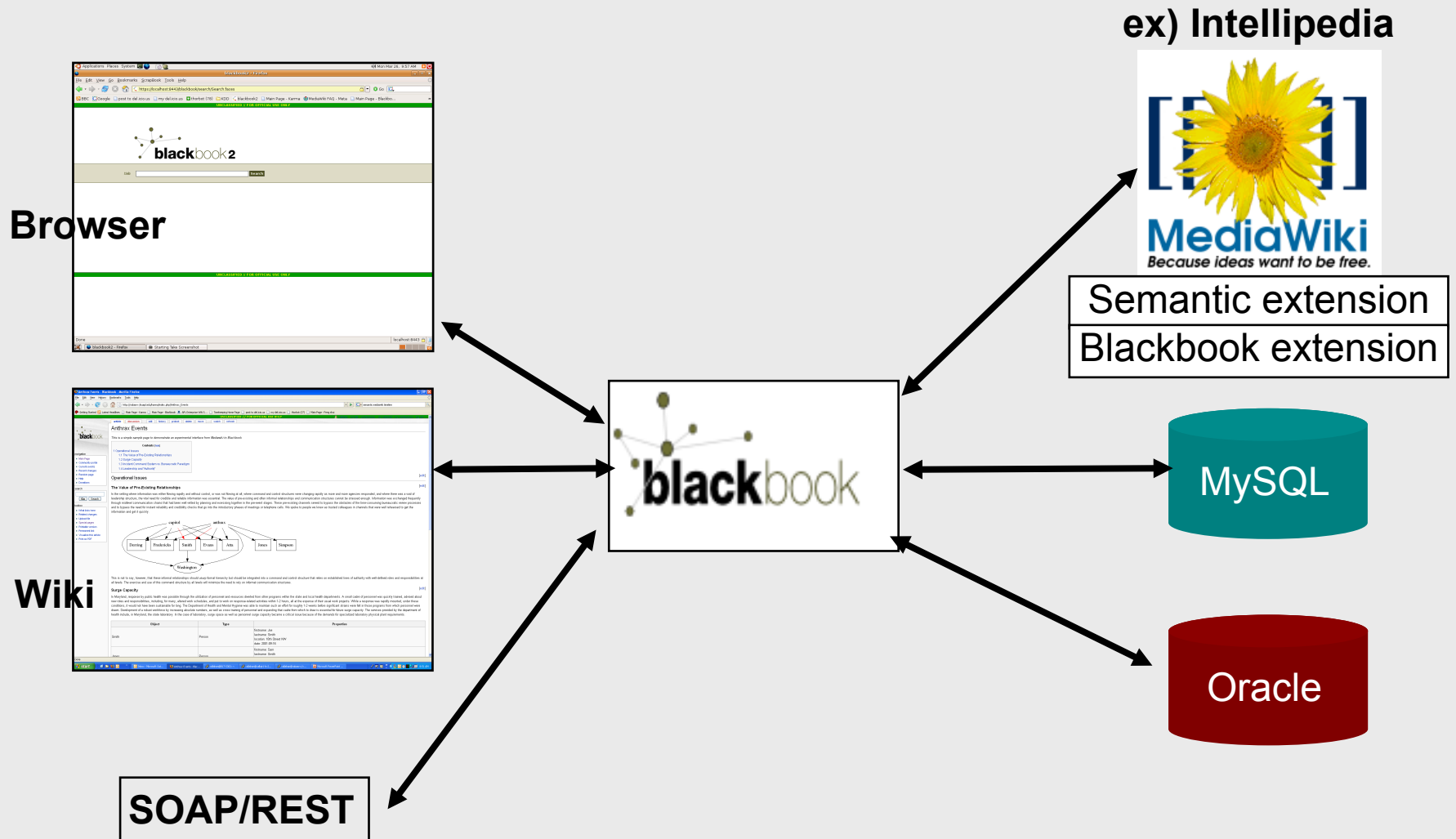


Wiki%~~100~~00



Wikis

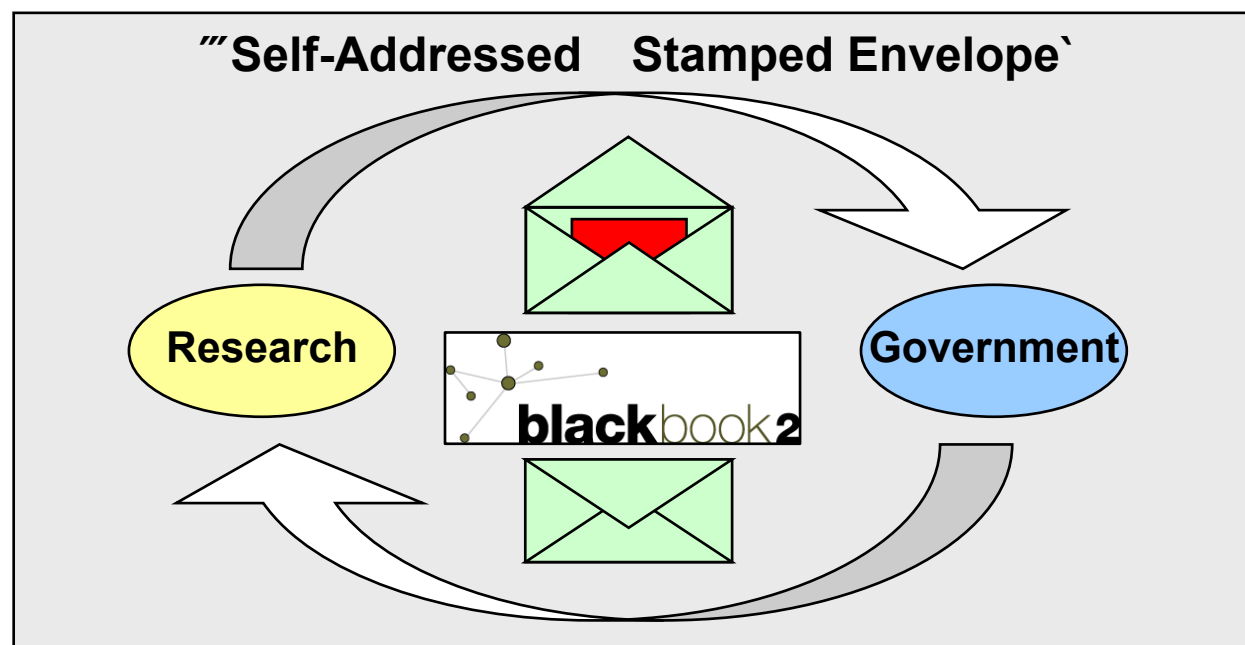
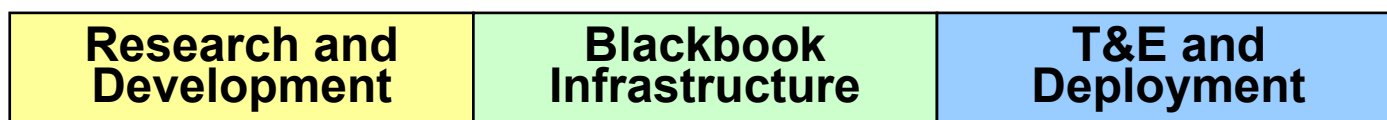
# Blackbook and Wikis



Like browsers, "Wiki's can be a front-end to Blackbook. Wiki's can also be a datasource. Wiki extensions can be utilized to enable Semantic and Blackbook features.

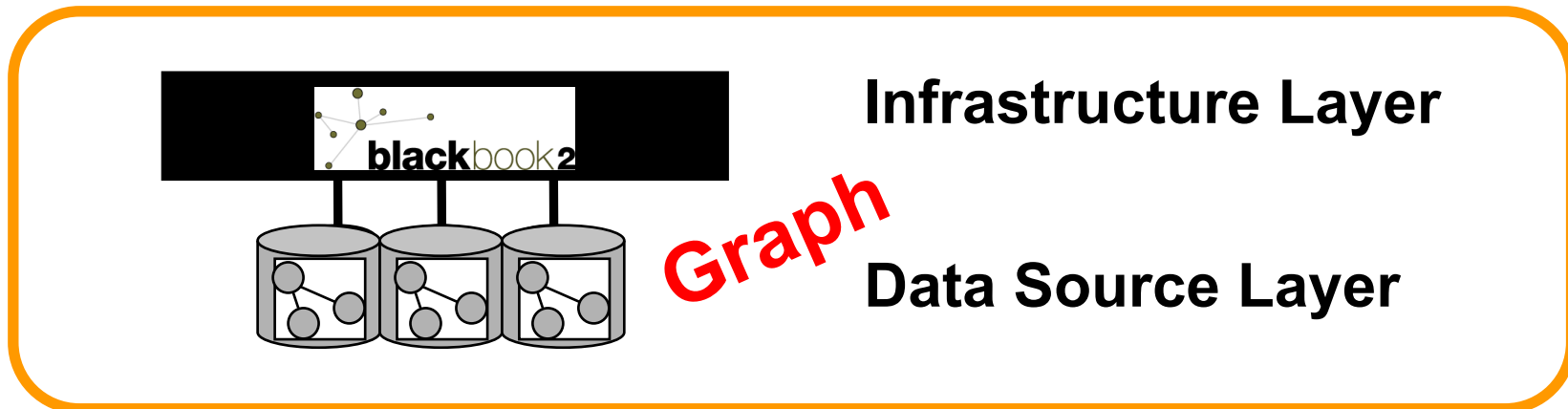
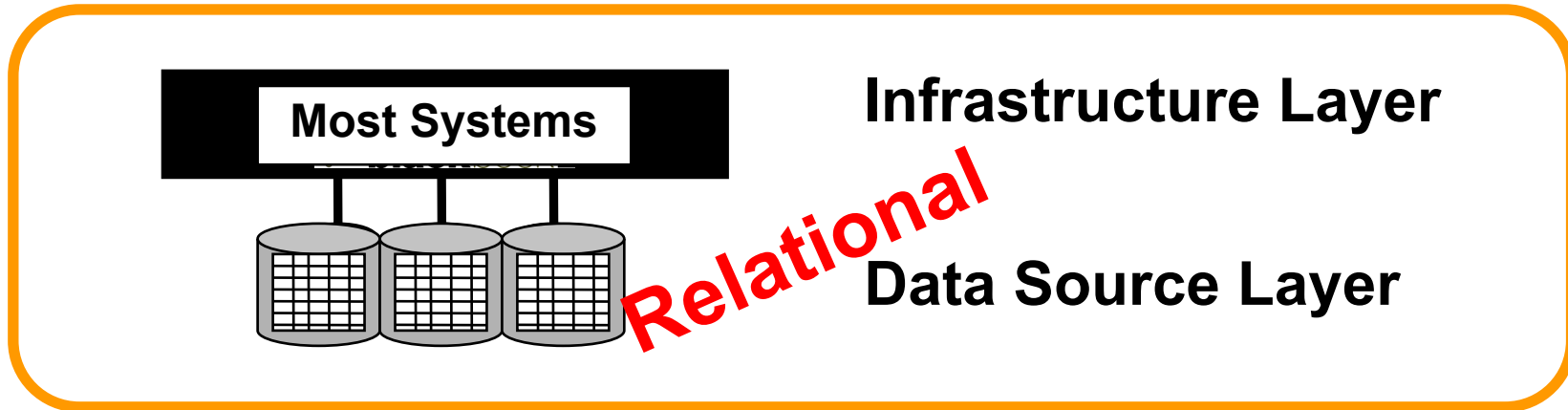
# Revolutionize Technology Transfer

Improve Intelligence Analysis by Coordinated Exposition of Multiple Data Sources Across Intelligence Community Agencies



A research product (red), such as a new and improved algorithm or visualization, can easily be transferred from research to government using the Blackbook “envelope” .

# Relational vs. Graph-based Systems



Blackbook2 is a JEE server-based RDF processor that provides an asynchronous interface to back-end datasources.