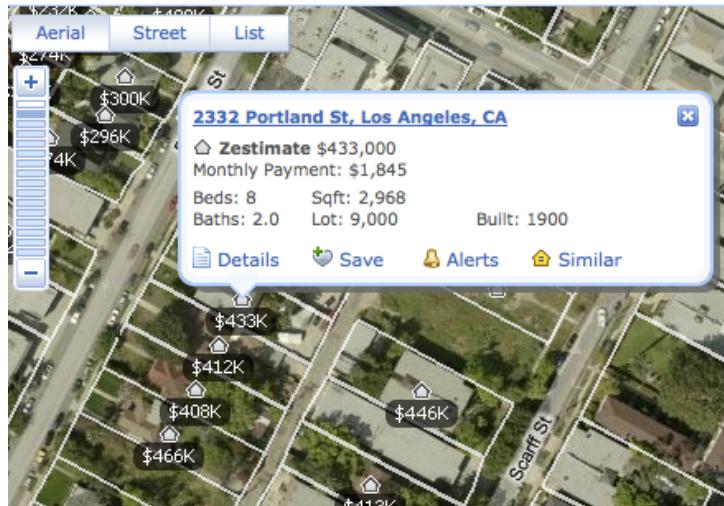


# Building Geospatial Mashups to Visualize Information for Crisis Management

**Shubham Gupta and Craig A. Knoblock**  
University of Southern California

# WHAT IS A GEOSPATIAL MASHUP?

**Integrated View** of data combined from diverse data sources  
on a **Map**



- Real Estate Listing
- Property Tax

Name	Lifts
Mountain High	3
Snow Summit	5
Mt Baldy	3
Snow Valley	5
Shirley Meadows Ski Area	0
Bear Mountain/Big Bear	0



- Snow Report
- Weather (Weather Underground)
- Snow Resorts

# MOTIVATION: CRISIS MANAGEMENT

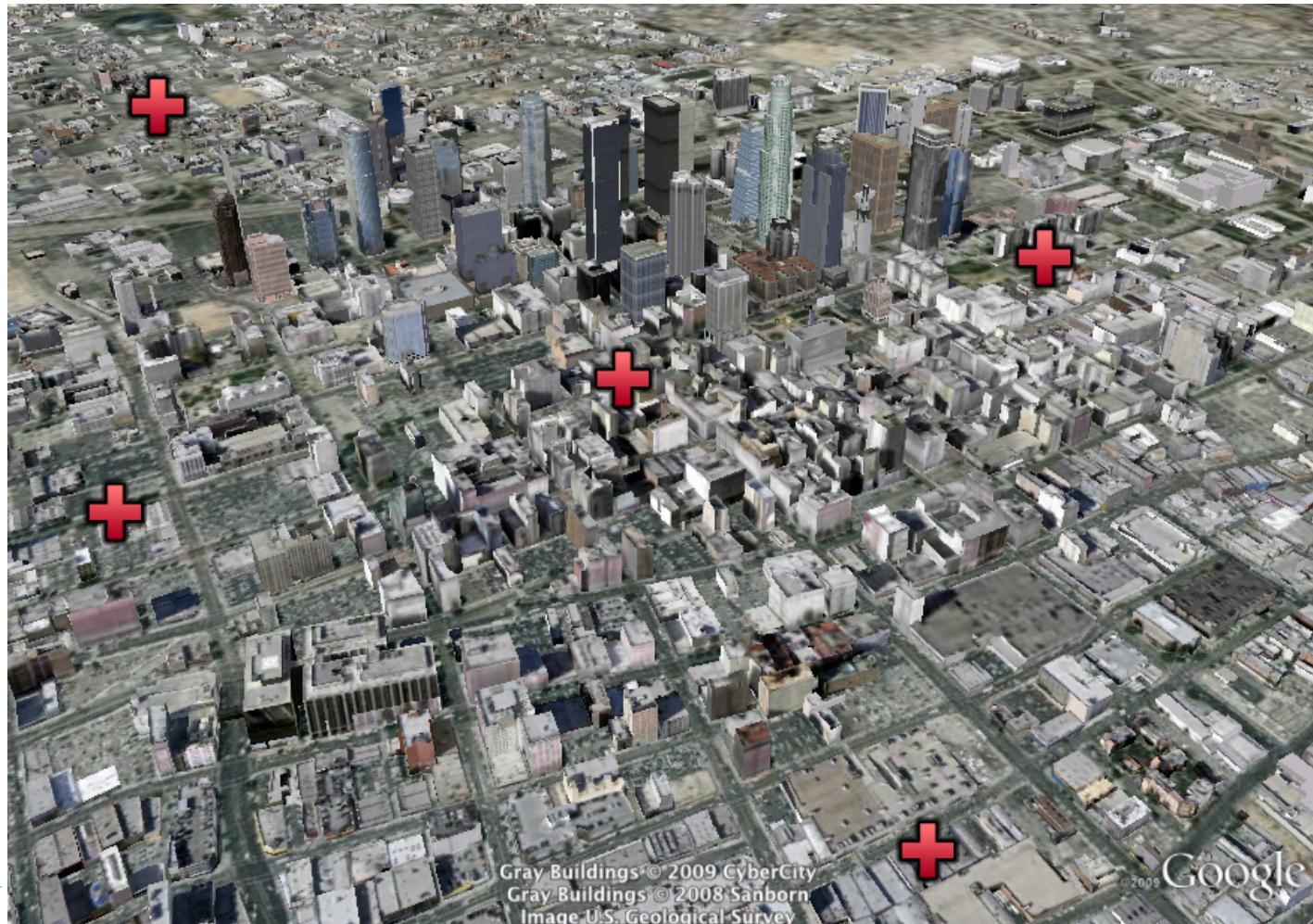
---

Characteristics of decision-making in such scenarios:

- ▶ Employs **multiple data sources** with heterogeneous data access and data representation types
- ▶ Takes place in **time-sensitive** environment: Every decision has to be taken promptly
- ▶ Various **data visualization** techniques are used such as graphs, timeline charts, etc. to understand the data better

# MOTIVATION: CRISIS MANAGEMENT

- ▶ An imaginary earthquake relief scenario in Los Angeles.



# MOTIVATION: CRISIS MANAGEMENT

- Analyst employs multiple data sources for decision-making

The image shows a 'TIME CONSUMING APPROACH' on the left and a 'GEOSPATIAL MASHUPS CAN SAVE TIME!' solution on the right.

**TIME CONSUMING APPROACH:** This section shows a screenshot of a Google News interface and a separate database table. The news interface includes a map of Los Angeles with several red pins indicating locations. A large callout bubble highlights three issues:

- X Managing multiple software packages
- X Need to switch back and forth b/w sources
- X No good data visualization may be provided

A green arrow points from the 'X Managing multiple software packages' text to the database table. Another green arrow points from the 'X Need to switch back and forth b/w sources' text to the news interface. A third green arrow points from the 'X No good data visualization may be provided' text to the map.

**GEOSPATIAL MASHUPS CAN SAVE TIME!:** This section shows a geospatial mashup interface. It features a map of Southern California with red pins, a table of 'Latest Injuries Statistics', and a chart titled 'CURRENT STATISTICS' showing 'INJURIES' and 'CASUALTIES' over time. A large callout bubble highlights three benefits:

- ✓ Only Web browser required
- ✓ Integrated view of all sources
- ✓ Customize visualization as per needs

A green arrow points from the '✓ Only Web browser required' text to the chart. Another green arrow points from the '✓ Integrated view of all sources' text to the map. A third green arrow points from the '✓ Customize visualization as per needs' text to the table.

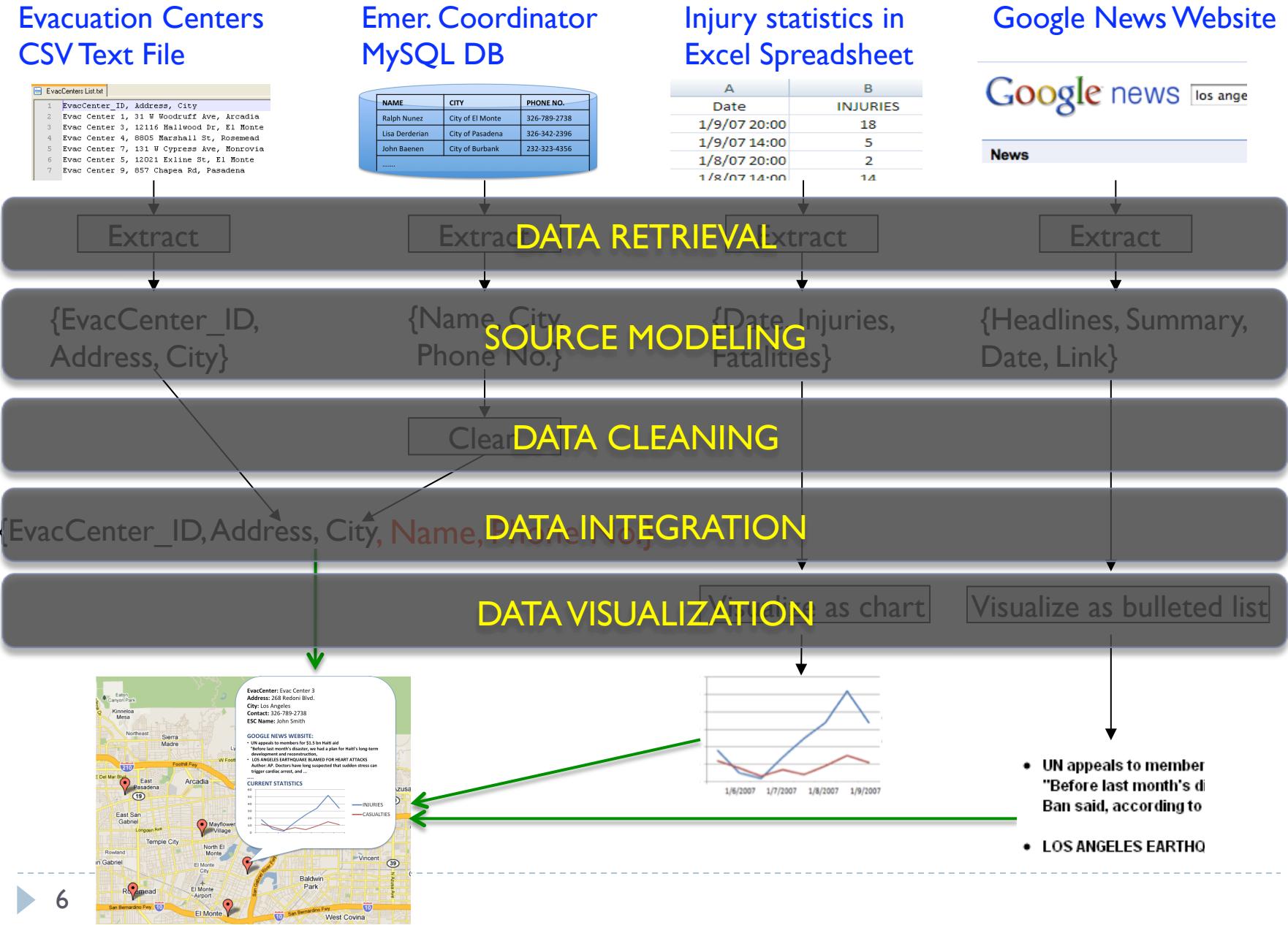
**Database Table:** A table titled 'PHONE NO.' containing four rows of data.

PHONE NO.
326-789-2738
326-342-2396
232-323-4356

**CSV Text File:** A table titled 'City' listing addresses:

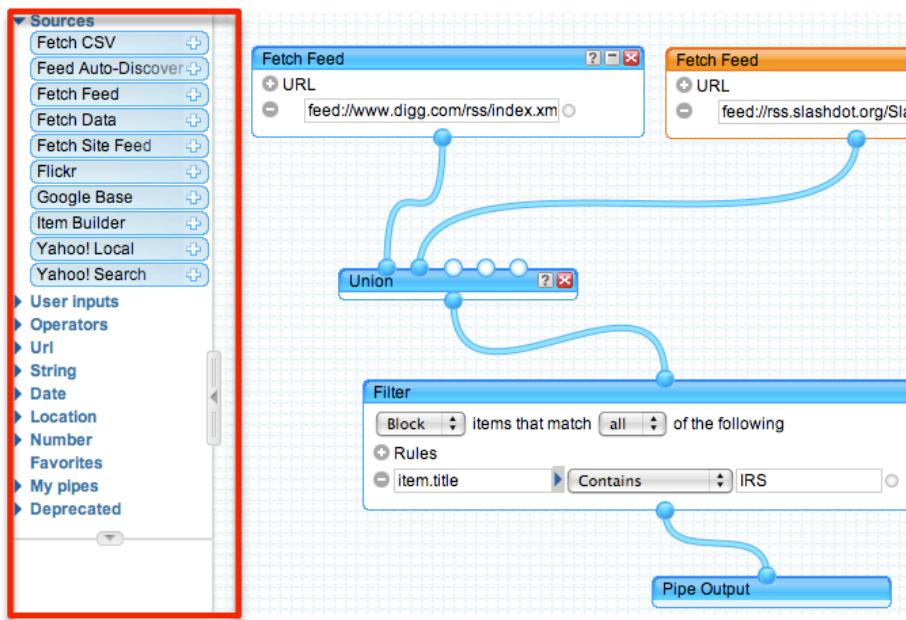
City
druff Ave, Arcadia
llwood Dr, El Monte
shall St, Rosemead
ress Ave, Monrovia
line St, El Monte
ea Rd, Pasadena

# MASHUP BUILDING ISSUES



# EXISTING APPROACHES

**Existing solutions require some prior programming knowledge**



- I. Mostly based on Widget Paradigm**
  - Widgets (i.e., 43 for Pipes, 300+ for MS) represents an operation on the data
  - Locating and learning to customize widget can be **time consuming**
- 2. None of the current tools support customized data visualization!**

Can we come up with a framework that addresses all of the issues while still making the Mashup building process easy?

# OUR APPROACH

---

- ▶ In our earlier work, we introduced **Programming-by-Demonstration approach** to Mashup building.
  - ▶ Does not require any programming or widget knowledge.
  - ▶ Enables user to create a mashup by simply demonstrating the system on one or more concrete example.
- ▶ We present a **Programming by Demonstration approach** to support **customized Data Visualization**.

# KARMA USER INTERFACE

The screenshot displays the Karma User Interface. At the top, there's a menu bar with 'File', 'Edit', 'View', 'Tools', and 'Help'. Below the menu is a toolbar with icons for 'Import', 'Clean', 'Integrate', and 'Publish'. The main area features a 'Table' view with two tabs: 'Source1' and 'Source2'. The 'Source1' tab is active, showing a grid of data with columns: PR-TimeDuration, PR-Number, Data Type, and Column Name. The data includes rows for dates from July 2009 to May 2009 and corresponding StockValue values. A callout bubble points to this table with the text 'Data Table – Spreadsheet Type Interface'. In the bottom right corner of the main window, there's a large empty text area. A callout bubble points to this area with the text 'Various Information Integration Operations'. At the bottom of the interface, there's a row of buttons labeled 'Wrapper', 'Database', 'Excel', 'Text File', and 'KML File'. A callout bubble points to the 'Excel' button with the text 'Data Source Types Currently Supported by Karma'. The bottom left corner of the slide has a small blue arrow icon followed by the number '9'.

Data Table – Spreadsheet Type Interface

Various Information Integration Operations

Data Source Types Currently Supported by Karma

9

# DATA RETRIEVAL

- ▶ Karma facilitates retrieval of data from structured data-sources, such as Excel spreadsheets, MySQL databases and CSV files.
- ▶ Karma also facilitates the extraction of data from semi-structured data sources such as web pages.

	EvacCenter_ID, Address, City
1	Evac Center 1, 31 W Woodruff Ave, Arcadia
2	Evac Center 3, 12116 Hallwood Dr, El Monte
4	Evac Center 4, 8805 Marshall St, Rosemead
5	Evac Center 7, 131 W Cypress Ave, Monrovia
6	Evac Center 5, 12021 Exline St, El Monte
7	Evac Center 9, 857 Chapea Rd, Pasadena

CSV Text File

A	B	C
Date	INJURIES	FATALITIES
1/9/07 20:00	18	12
1/9/07 14:00	5	8
1/8/07 20:00	2	3
1/8/07 14:00	14	7
1/7/07 20:00	25	4
1/7/07 14:00	34	9
1/6/07 20:00	52	15
1/6/07 14:00	34	11

10      Excel Spreadsheet

NAME	CITY	PHONE NO.
Ralph Nunez	City of El Monte	326-789-2738
Lisa Derderian	City of Pasadena	326-342-2396
John Baenen	City of Burbank	232-323-4356
.....		

MySQL Database

Google news

News

[Top Stories](#) [More sections ▾](#)

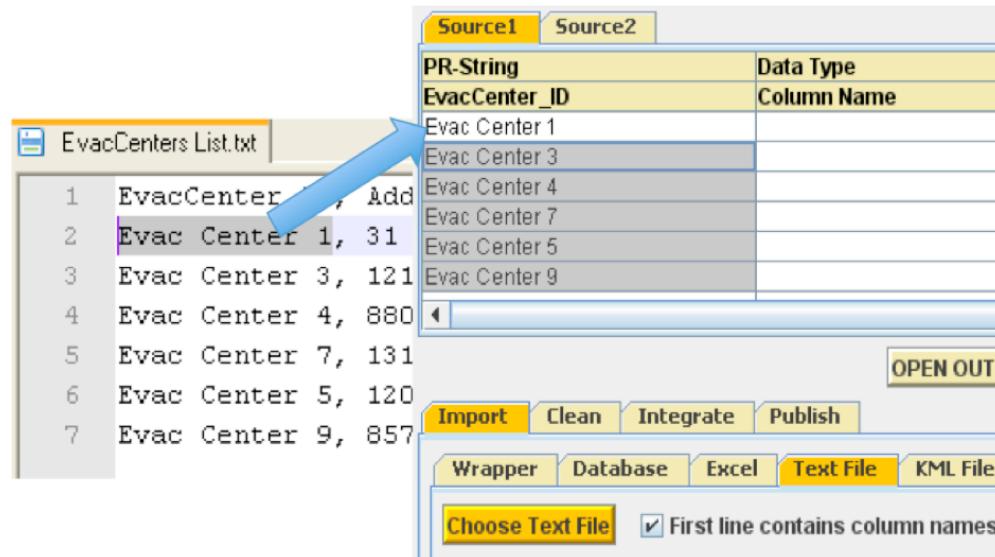
[All news](#) [Images](#)

 **UN appeals to members**  
Press TV - 2 hours ago  
"Before last month's disaster, we reconstruction," Ban said, accord

HTML Web Page

# DATA RETRIEVAL

- ▶ The retrieval of data from structured data-sources, such as Excel sheets and CSV files is done through a drag and drop mechanism.
- ▶ The user is only required to select a sample data-element and drop it into Karma's data table.



# DATA RETRIEVAL

- Extraction of data from HTML web-pages is done using software wrappers.

Table

Source1	Source2	PR-String	PR-String	PR-String	PR-String	PR-String	Data 1 Column
Headline	Summary	Url	Datetime	Source			
UN appeals to member...	"Before last month's dis...	http://news.google.com/...	9 hours ago	Monsters and Critics.com			
LOS ANGELES EARTH...	Author: AP. Doctors hav...	http://news.google.com/...	1 hour ago	Los Angeles Times			
Not guilty plea...Digging...	LOS ANGELES (AP) — ...	http://news.google.com/...	17 hours ago	9&10 News			
Style Star: Rihanna	Rihanna performs on st...	http://news.google.com/...	2 hours ago	Access Hollywood			
6.0 Earthquake in Califo...	According the the US G...	http://news.google.com/...	Feb 4, 2010	Gather.com			
Haiti is a reminder of ho...	In the aftermath of the d...	http://news.google.com/...	Feb 8, 2010	Los Angeles Times			
Earthquakes Weekly Up...	... former Seattle Sound...	http://news.google.com/...	17 hours ago	OurSports Central (pres...			
US & World News, Tue...	Meanwhile, about 500 r...	http://news.google.com/...	1 hour ago	WTWW			
Shake, Shake, Shake	But the US Geological S...	http://news.google.com/...	15 hours ago	Newspaper.com			
At a Haiti school's reop...	(Carolyn Cole / Los Ang...	http://news.google.com/...	Feb 1, 2010	L...			

Import Clean Integrate Publish

Wrapper Database Excel Text File KML File

Load Fetch Agents Repository

Fetch Agents

Wrapper Name	Input Parameter Names	Output Parameter Names	Input values	Execute
Alacra2	CompanyType, Compan...	CompanyName,Addres...		Execute
flightStats2	airport, airlines			Execute
GoogleNews	q	Headline,Summary,Url,...	Los Angeles Earthquake	Execute
GoogleSearch	searchInfo			Execute
LATimes	SearchQuery			Execute
SICCodeDescription	psic	Category,IndustryGroup,...		Execute

Enter values for input parameters in yellow column to execute wrapper!

# SOURCE MODELING

- ▶ Karma automatically generates the semantic types of each attribute to learn the underlying model of the data source.
- ▶ Supervised machine learning techniques are used to generate a set of patterns for each semantic type from training data.

Source1	Source2	Source3
PR-String	PR-String	
StreetAddress	City	
2353 Portland St.	Los Angeles	
543 Orchard Ave.	Burbank	
417 Glenmoor Cir.	Milpitas	
6472 Hawthorne Blvd	Pasadena	
325 Abbot Ave.	Santa Clara	

Initial Semantic Type



Source1	Source2	Source3
PR-Address	PR-String	
StreetAddress	City	
2353 Portland St.	Los Angeles	
543 Orchard Ave.	Burbank	
417 Glenmoor Cir.	Milpitas	
6472 Hawthorne Blvd	Pasadena	
325 Abbot Ave.	Santa Clara	

Manually label the data with the correct semantic type for training

Source1	Source2	Source3	Source4
PR-Address	PR-String		
Address	EvacCenter_ID		
31 W Woodruff Ave	Evac Center 1		
12116 Hallwood Dr	Evac Center 3		
8805 Marshall St	Evac Center 4		
131 W Cypress Ave	Evac Center 7		
12021 Exline St	Evac Center 5		
857 Chapea Rd	Evac Center 9		

When the new data is imported of same type, Karma automatically labels it correctly

# DATA CLEANING

- ▶ Karma performs the data cleaning by learning and applying the transformation rules that are learnt from the user specified example.
- ▶ Example:

Source1	Source2
PR-String	PR-String
City	Coordinator_Name
City of Agoura Hills	Carol Tubelis
City of Alhambra	Bat. Chief Ray Mosack
City of Arcadia	Chief Dave Odell
City of Artesia	Madalena Galindo
City of Duarte	Brian Villalobos
City of El Monte	Ralph Nunez

Initial data source



Source1	Source2	Cleaning Table
PR-String	Data Type	
City	User Defined Values	
City of Agoura Hills	Agoura Hills	
City of Alhambra		
City of Arcadia		
City of Artesia		
City of Duarte		
City of El Monte		

User provides example



Source1	Source2	Source3
PR-String	PR-String	
City	Coordinator_Name	
Agoura Hills	Carol Tubelis	
Alhambra	Bat. Chief Ray Mosack	
Arcadia	Chief Dave Odell	
Artesia	Madalena Galindo	
Duarte	Brian Villalobos	
El Monte	Ralph Nunez	

Data source after cleaning



Source1	Source2	Cleaning Table
PR-String	Data Type	
City	User Defined Values	
City of Agoura Hills	Agoura Hills	
City of Alhambra	Alhambra	
City of Arcadia	Arcadia	
City of Artesia	Artesia	
City of Duarte	Duarte	
City of El Monte	El Monte	

Karma applies the learned transformation rules over remaining data

# DATA INTEGRATION

- ▶ Karma detects and ranks its potential integration with other sources based on the common attribute names and matching semantic types.
- ▶ Karma suggests potential joins between the current data sources in the form of column completions.

Source1	Source2		
PR-String	PR-Address	PR-String	Data Type
EvacCenter_ID	Address	City	Column Name
Evac Center 1	31 W Woodruff Ave	Arcadia	
Evac Center 3	12116 Hallwood Dr	El Monte	
Evac Center 4	8805 Marshall St	Rosemead	
Evac Center 7	131 W Cypress Ave	Monrovia	
Evac Center 5	12021 Exline St	El Monte	
Evac Center 9	857 Chapea Rd	Pasadena	

Column completion

# DATA INTEGRATION

Karma suggests the possible column completions in a drop down list

MySQL Database loaded as a another source in Karma

Karma executes the join query once the user selects an option and displays the result

R-String	PR-Address	PR-String	Data Type
vacCenter_ID	Address	City	Select Column Name
vac Center 1	31 W Woodruff Ave	Arcadia	Select Column Name
vac Center 3	12116 Hallwood Dr	El Monte	Name
vac Center 4	8805 Marshall St	Rosemead	Phone No.
vac Center 7	131 W Cypress Ave	Monrovia	
vac Center 5	12021 Exline St	El Monte	
vac Center 9	857 Chapea Rd	Pasadena	

NAME	CITY	PHONE NO.
Ralph Nunez	City of El Monte	326-789-2738
Lisa Derderian	City of Pasadena	326-342-2396
John Baenen	City of Burbank	232-323-4356
.....		

PR-String	PR-Address	PR-String	PR-String
EvacCenter_ID	Address	City	Name
Evac Center 1	31 W Woodruff Ave	Arcadia	Chief Dave Odell
Evac Center 3	12116 Hallwood Dr	El Monte	Ralph Nunez
Evac Center 4	8805 Marshall St	Rosemead	Donna Wagner
Evac Center 7	131 W Cypress Ave	Monrovia	Dave Dennis
Evac Center 5	12021 Exline St	El Monte	Ralph Nunez
Evac Center 9	857 Chapea Rd	Pasadena	Lisa Derderian

# DATA VISUALIZATION

---

- ▶ Definition: Representation of data in a visual form to amplify cognition.
- ▶ Example: Timeline charts, Graphs, Tables, etc.
- ▶ Application:
  - ▶ Summarization of large data
  - ▶ Providing qualitative overview
  - ▶ Detecting patterns, anomalies and relationships in data
  - ▶ Identify regions of interest in large and complex datasets

**CRITICAL TOOL FOR EFFICIENT DATA ANALYSIS!**



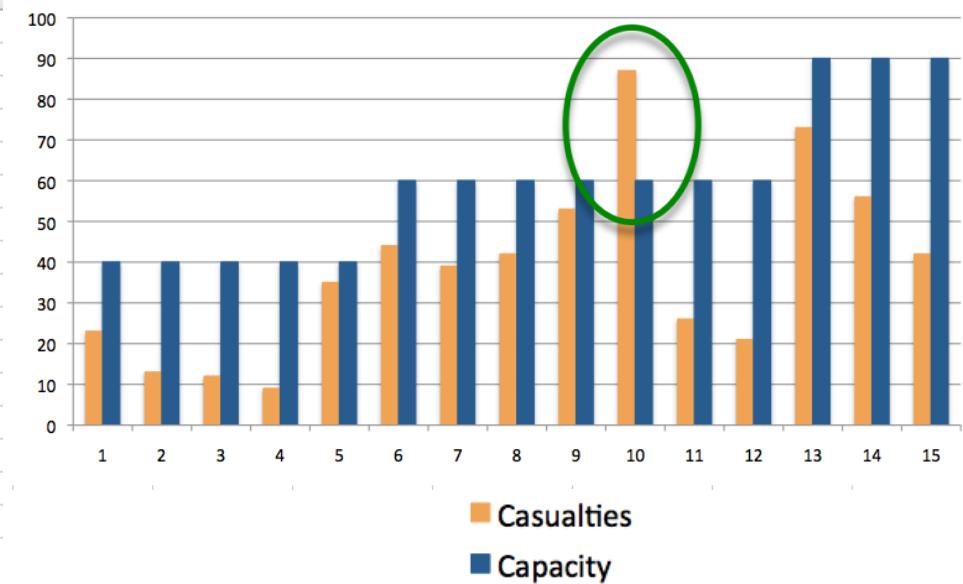
**CRITICAL IN CRISIS AND DISASTER SCENARIOS**

# DATA VISUALIZATION

## ▶ Example:

Which evacuation center is handling more casualties than its capacity?

	A	B	C
1	Evacuation Center ID	Casualties	Capacity
2	1	23	40
3	2	13	40
4	3	12	40
5	4	9	40
6	5	35	40
7	6	44	60
8	7	39	60
9	8	42	60
10	9	53	60
11	10	87	60
12	11	26	60
13	12	21	60
14	13	73	90
15	14	56	90
16	15	42	90



BETTER VISUALIZATION!

# VISUALIZATION BY DEMONSTRATION

- The user demonstrates the desired data visualization by specifying some examples using a **drag and drop** mechanism. E.g. Visualize News data as bulleted lists in final mashup:

The screenshot shows a data visualization tool interface. On the left, there is a 'Preview Generated' window displaying a table of news data. The table has columns: PR-String, PR-Date, PR-String, PR-URL, and Data Type. The Data Type column contains 'Link'. A large speech bubble labeled 'Output Preview Window' points to this preview area. Below the preview window, there is a 'Publish' section with buttons for 'Import', 'Clean', 'Integrate', and 'Publish'. Under 'Publish', there are buttons for 'HTML', 'KML', 'XML', 'CSV Text File', and 'Database'. Below these buttons are 'Publish HTML' and 'Save HTML' buttons. On the right, there is a green-bordered window titled 'LIST FORMAT' with the sub-instruction '(Drag data to this section for List format)'. Below it is a 'PARAGRAPH FORMAT' section with the sub-instruction '(Drag data to this section for Paragraph format)'. At the bottom is a 'TABLE FORMAT' section with the sub-instruction '(Drag data to this section for Table format)'. The data in the preview window includes:

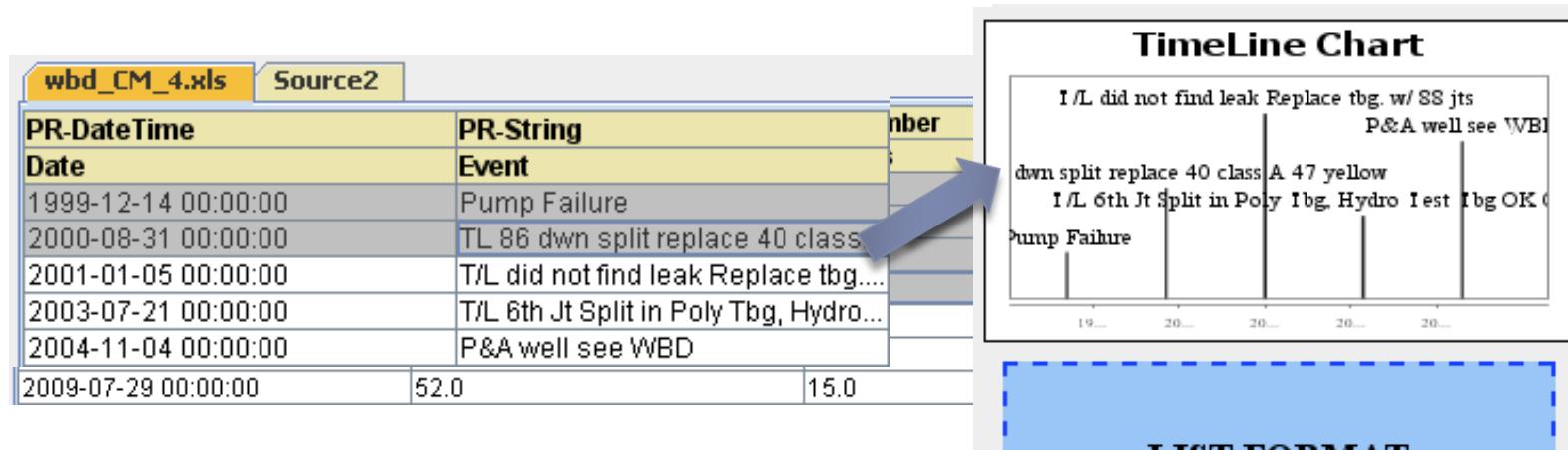
PR-String	PR-Date	PR-String	PR-URL	Data Type
Headline	Date	Summary	Link	
KILLER QUAKE: AFTERSH... LOS ANGELES EARTHQUA...	1/9/2007 1/9/2007	The anniversary Tuesday of... Author: AP. Doctors have lo...	<a href="#">http://www.reut...</a>	
Man Indicted in Earthquake...	1/9/2007	This is considered the larg...	<a href="#">http://www.reuters.com/arti...</a>	
4.9 Earthquake Rattles Muc...	1/9/2007	A series of aftershocks fro...	<a href="#">http://www.staugustine.co...</a>	
MINOR EARTHQUAKE SHA...	1/9/2007	A moderate earthquake jolt...	<a href="#">http://www.reuters.com/arti...</a>	

# VISUALIZATION BY DEMONSTRATION

Karma currently supports four types of visualization formats:

## I. Chart Format:

- Useful for visualizing:
  - Numerical Statistics
  - Time Based Events
- Karma uses Google Charts API to dynamically generate charts in final mashup.
- Karma attempts to find the best suitable chart for data using its semantic types.



# VISUALIZATION BY DEMONSTRATION

Karma currently supports four types of visualization formats:

## 2. Paragraph Format:

- Useful for visualizing descriptive text data such as Wikipedia definitions.

The screenshot shows the Karma interface with two tabs: 'Source1' and 'Source2'. Under 'Source1', there is a table with columns 'PR-String' and 'PR-String'. The first row contains 'Paragraph1' and 'Paragraph1'. The second row contains 'Los Angeles (pronounced /lɒs ˈændʒələs/ los-AN-jə-ləs; Spanish: [los ˈanxeles], Spanish for "the angels") is the largest city in the state of California and the second largest in the United States.[1] Often abbreviated as L.A. and nicknamed The City of Angels, Los Angeles has an estimated population of 3.8 million[2] and spans over'.

**Paragraph1:** Los Angeles (pronounced /lɒs ˈændʒələs/ los-AN-jə-ləs; Spanish: [los ˈanxeles], Spanish for "the angels") is the largest city in the state of California and the second largest in the United States.[1] Often abbreviated as L.A. and nicknamed The City of Angels, Los Angeles has an estimated population of 3.8 million[2] and spans over

## 3. List Format:

- Useful for visualizing information in a bulleted list such as list of summarized news articles .

The screenshot shows the Karma interface with two tabs: 'Source1' and 'Source2'. Under 'Source1', there is a table with columns 'PR-String' and 'PR-String'. The rows contain 'Headline' and 'Date'. The first row contains 'KILLER QUAKE: AFTERSHOCKS OF CHAOS, FEAR WORLD AWAY BUT ...' and '1/9/2007'. The second row contains 'LOS ANGELES EARTHQUAKE' and '1/9/2007'. The third row contains 'Man Indicted in Earthquake' and '1/9/2007'. The fourth row contains '4.9 Earthquake Rattles Muc...' and '1/9/2007'. The fifth row contains 'MINOR EARTHQUAKE SHA...' and '1/9/2007'.

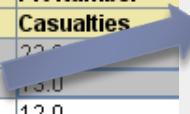
- KILLER QUAKE: AFTERSHOCKS OF CHAOS, FEAR WORLD AWAY BUT ...  
The anniversary Tuesday of the terrible Los Angeles earthquake was supposed to be a day of celebration, a time to look back proudly on the

# VISUALIZATION BY DEMONSTRATION

Karma currently supports four types of visualization formats:

## 4. Table Format:

- Useful for visualizing information that is best presented in a row-and-column format such as numerical values etc.



Source1	Source2
PR-Number	PR-Number
Evacuation Center ID	Casualties
1.0	22.0
2.0	13.0
3.0	12.0
4.0	9.0
5.0	35.0
6.0	44.0

Evacuation Center ID	Casualties	Capacity
1.0	23.0	40.0
2.0	13.0	40.0
3.0	12.0	40.0
4.0	9.0	40.0

# RESULTS CAN BE PUBLISHED IN MULTIPLE FORMATS

- ▶ Karma lets you export your final mashup in variety of formats:
  - HTML Page
  - Database table
  - KML Layer
  - XML File
  - CSV Text File

The screenshot shows the Karma interface. At the top, there are four tabs labeled 'Source1', 'Source2', 'Source3', and 'Source4'. Below these tabs is a data grid with the following columns and data:

PR-String	PR-Address	PR-String	PR-String	PR-String
EvacCenter_ID	Address	City	Coordinator_Name	Phone
Evac Center 1	31 W Woodruff Ave	Arcadia	Chief Dave Odell	(626) 574-5102
Evac Center 3	12116 Hallwood Dr	El Monte	Ralph Nunez	(626) 580-2065
Evac Center 4	8805 Marshall St	Rosemead	Donna Wagner	(626) 569-2102
Evac Center 7	131 W Cypress Ave	Monrovia	Dave Dennis	(626) 256-8104
Evac Center 5	12021 Exline St	El Monte	Ralph Nunez	(626) 580-2065
Evac Center 9	857 Chapea Rd	Pasadena	Lisa Derderian	(626) 744-7276

Below the data grid is a toolbar with buttons for 'Import', 'Clean', 'Integrate', and 'Publish'. The 'Publish' button is highlighted. A secondary toolbar below it contains buttons for 'HTML', 'KML', 'XML', 'CSV Text File', and 'Database'. The 'XML' button is highlighted. A callout bubble points to the 'Save XML' button, which is located at the bottom of the secondary toolbar. The 'Save XML' button has a small icon of a file with a plus sign.

Different mashup publishing options

# AUTOMATICALLY FINDS GEOSPATIAL REFERENCES

- ▶ Final mashup output in HTML web page format:

The screenshot shows a user interface for data integration and visualization. On the left, there's a table titled "Data Table" with columns: PR-String, PR-Address, PR-String, PR-String, PR-String, Data Type, and Column. A green oval highlights the "Address" column. Below the table are buttons for Import, Clean, Integrate, and Publish, with "Publish" highlighted. Underneath are buttons for HTML, KML, XML, CSV Text File, and Database, with "HTML" highlighted. Buttons for Publish HTML and Save HTML are also present.

On the right, there's a news snippet about an anniversary earthquake, a line chart titled "Line Chart" comparing Casualties and Injuries over time, and a summary box containing bulleted text.

- Geospatial references automatically detected using semantic types.
- Google geocoding service used to get coordinates.
- Markers are placed on those coordinates in the final mashup.

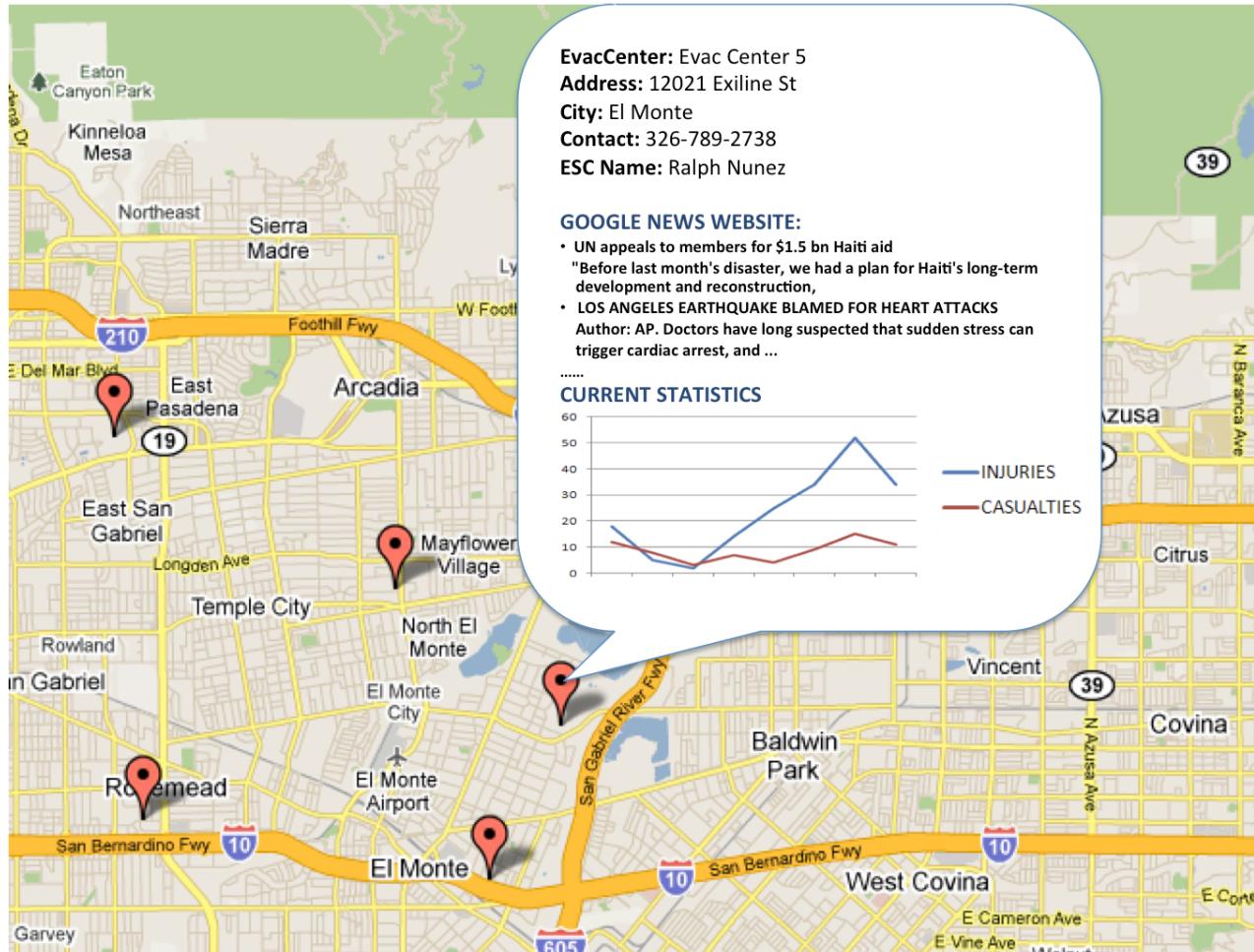
## PARAGRAPH FORMAT

(Drag data to this section for Paragraph format)

▶ 24

# CONSTRUCTS A MAP WITH USER-DEFINED LAYOUT

- ▶ Final mashup output in HTML web page format:



## RELATED WORK

---

- ▶ Existing and Past mashup building tools:



- ▶ Lot of research has been done on data visualization, esp. within data mining and knowledge discovery communities.
  - ▶ Improved Sammon algorithm for visualizing sparse data [Martin et al. 2004].
  - ▶ Interactive High dimensional data visualization [Buja et al. 1996]
- ▶ None of the current (or past) mashup building tools support customized data visualization in the final mashup output.

# CONCLUSION

---

- ✓ Geospatial Mashups can be instrumental for efficient and judicious decision-making in crisis management.
- ✓ Visualization By Demonstration approach lets the user customize the mashup output without any programming.
- ✓ Users can now construct a mashup in one seamless integrated process including data visualization.

## FUTURE WORK

---

- ▶ Integrate more visualization formats such as scatter plots, 2D/3D isosurfaces, etc.
- ▶ Provide the ability to save the plans for extracting and integrating the data so that they can be applied to the new data.
- ▶ Provide the ability to support dynamic view of data (to reflect the changes made to data in the mashup).
- ▶ Support the integration of geospatial data in mashups (i.e., vector layers, raster layers).

# THANK YOU!

---

**Any Questions? Comments?**