



# Interactive Information Integration for Network Extraction, Discovery, and Analysis

Craig Knoblock  
Information Sciences Institute  
University of Southern California

Based on joint work with Pedro Szekely, Shubham Gupta,  
and Rattapoom Tuchinda



# Network Extraction, Discovery, and Analysis

- Much of the previous work has assumed that this should be an automatic process
  - The problem is that there is a huge amount of data it continues to grow
- For some types of tasks, we should consider how we can allow analysts to drive the process
  - Exploit the knowledge and skills of an analyst
  - Make the analysts more effective
  - Support the extraction, integration and analysis of highly diverse sources of data



# Motivating Example: Analyzing Movement Data

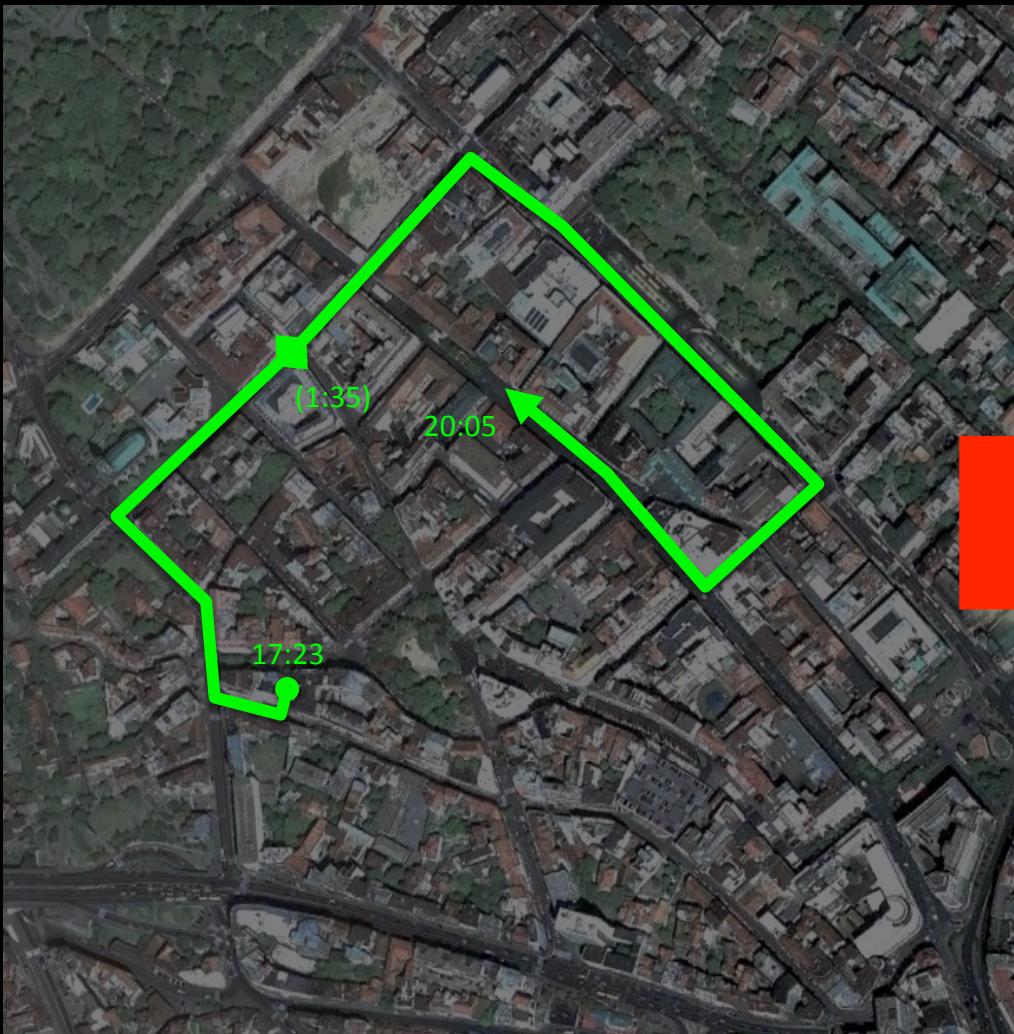
Without context  
the data is  
meaningless

Track	Latitude	Longitude	Time
tr-001	20.4573536672	44.8177358741	17:23:00
tr-001	20.4570188882	44.8177212897	17:24:12
tr-001	20.4566247835	44.8174301050	17:24:47
tr-001	20.4561817212	44.8171662676	17:28:08
tr-001	20.4555110995	44.8168148316	17:29:30
tr-001	20.4554824243	44.8166732121	17:23:00
tr-001	20.4565241020	44.8163966851	19:58:45
tr-001	20.4577144223	44.8159429667	19:58:53
tr-001	20.4586475431	44.8152713395	19:59:01
tr-001	20.4594042182	44.8156705582	19:59:09
tr-001	20.4586262967	44.8163906038	20:04:17
tr-001	20.4581171801	44.8167898621	21:04:34
tr-001	20.4577474239	44.8170976552	22:04:51
tr-001	20.4582442353	44.8174771842	23:05:08

# Visualizing MTI Data on Imagery

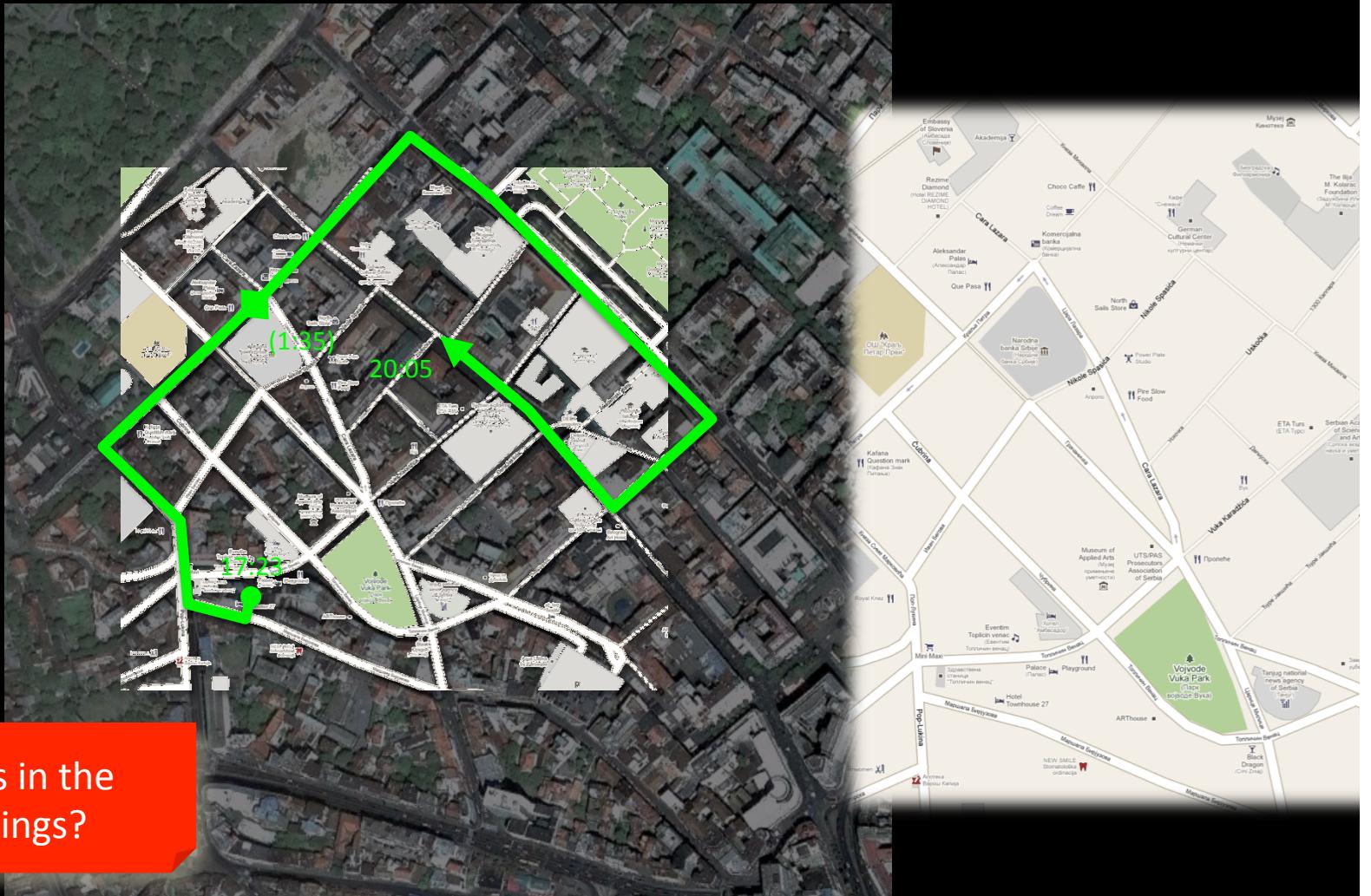
Imagery

	Time
41	17:23:00
97	17:24:12
50	17:24:47
76	17:28:08
16	17:29:30
21	17:23:00
51	19:58:45
67	19:58:53
95	19:59:01
82	19:59:09
38	20:04:17
21	21:04:34
52	22:04:51
42	23:05:08



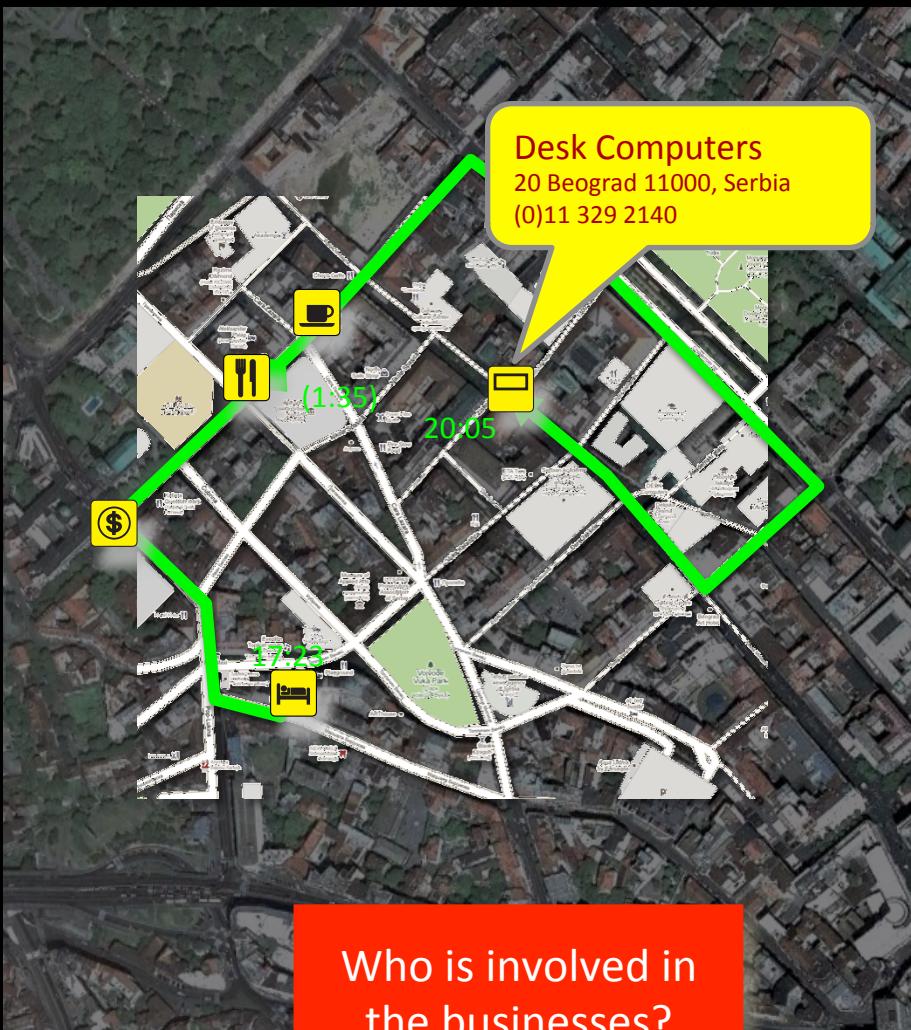
# Road Network Identifies Streets

- Imagery
- Road Network



# Phone Book Data To Identify Buildings

- Imagery
- Road Network
- Phone Books



**YellowPages.rs** Ako nije ovde, možda i ne postoji.

Advanced search:  By address  By brand  By phone

Beograd vuka karadzica 2 Location Address

[List your business](#)

Refine search:  
[By category](#) [By industry:](#)

- Banks
- Air conditioners and air conditioning

[By location:](#)

**Chinese restaurant Peking**  
Location: Beograd, Stari Grad, Vuka Karadžić  
Recommendations: [Read recommendations](#)  
Description: U srcu Beograda, u pešačkoj zoni svojoj tajnovitosti nalazi se najstariji restoran u Beogradu koji preko dve decenije svojim kvalitetom daje poseban ugodaj gostima. Na...  
Keywords: Restaurants Catering supplies Chinese food

**988info.rs** MESTO ZA VAŠ OGLAS

Fizička lica

Izaberite kriterijume za pretragu

Mrežna grupa: Beograd 011 Mesto: Stojnik (Beograd)

Telefon: \_\_\_\_\_ Ime: \_\_\_\_\_ Prezime: \_\_\_\_\_

Molimo Vas koristite srpske karaktere (Please use Serbian characters) - [View](#)

**VAŽNO:** Za probleme prilikom pokretanja pretrage i promene mrežne grupe, omogućite pristup na ovu stranicu.

Broj strane [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#) [18](#)

ANTONIJEVIĆ, MILOMIR	011 825
MILJE UROŠEVIĆA 24, STOJNIK	
ANTONIJEVIĆ, RADA	011 825
MILJE UROŠEVIĆA 32, STOJNIK	
ANTONIJEVIĆ, RADMILA	011 825
MILJE UROŠEVIĆA 27, STOJNIK	
ANTONIJEVIĆ, RADOLJUB	011 825
MILJE UROŠEVIĆA 31, STOJNIK	
ANTONIJEVIĆ, SLOBODAN	011 825
STOJNIK BB, STOJNIK	
ARANĐELOVIĆ, MILENA	011 825

ANALITIČKI INSTITUT VEDA, DOO, STOJNIK



# Domain Registration Links Businesses to Cyber-World



- Imagery
- Road Network
- Phone Books
- Cyber Data

Desk Computers  
20 Beograd 11000, Serbia  
(0)11 329 2140  
Web site: [www.desk.rs](http://www.desk.rs)  
ID Number: 6326099, Tax ID: 101520010  
Admin: Miljan Jovanović



HOME RESEARCH MONITOR

Domain name: desk.rs.  
Domain status: Active  
Registration date: 10.03.2008 12:15:41  
Modification date: 08.02.2011 14:37:50  
Expiration date: 10.03.2012 12:15:41  
Registrar: Absolut OK d.o.o.

Owner: Desk doo  
Address: Bulevar despota Stefana 110, Beograd  
ID Number: 6326099  
Tax ID: 101520010

DNS: ns1.absolutok.com. - 80.93.224.1  
DNS: ns2.absolutok.com. - 80.93.224.2

Administrative contact: Miljan Jovanović, D  
Address: Bulevar despota Stefana 110, Beograd

Technical contact: Tehnička podrška, Absolut  
Address: Palmira Toljatija 5/III, Novi Beograd



# Social Networks Provide Information about Businesses and People

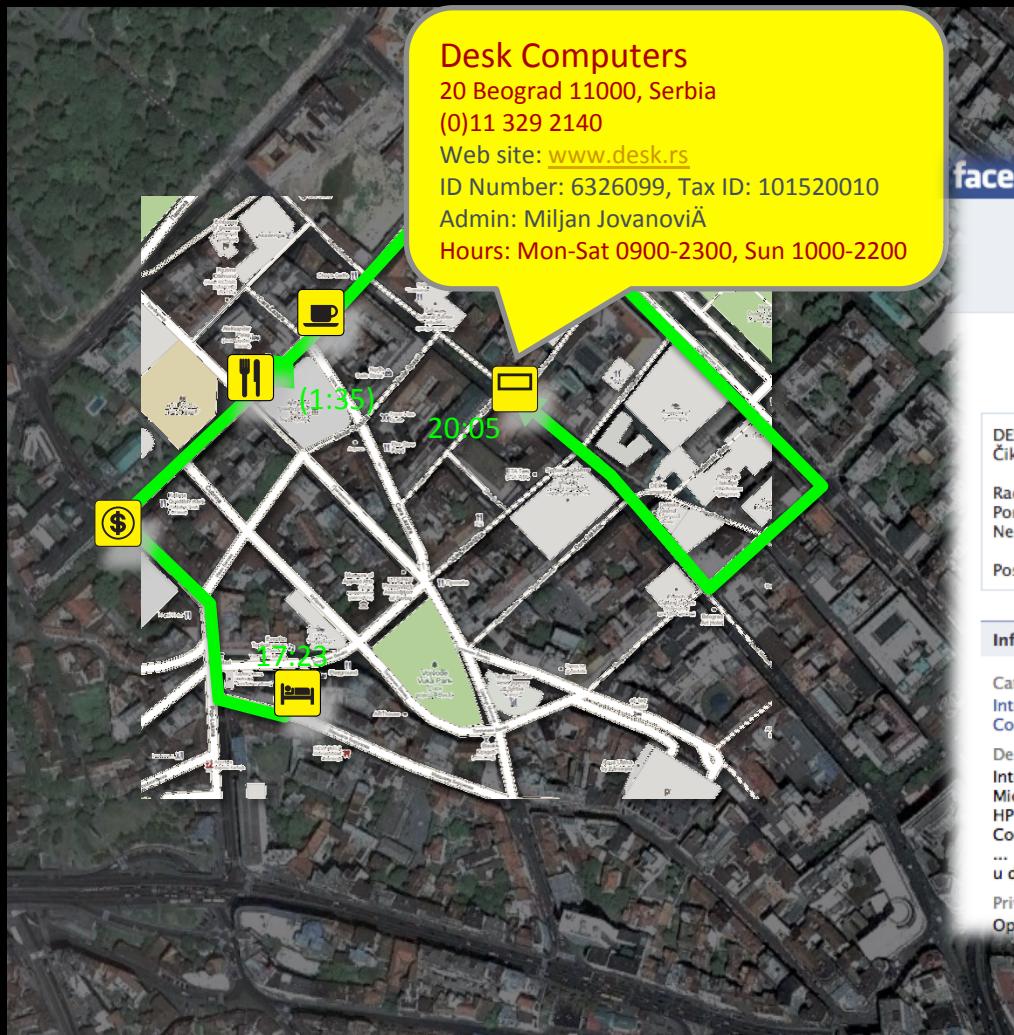
- Imagery
- Road Network
- Phone Books
- Cyber Data
- Social Network

AINS

ija

.o.  
ija

.o.o.  
bija



facebook Search

**DESK computers GRUPA**

**Basic Info**

Name: DESK Computer  
Category: Internet & Technology  
Description: Intel Premier Partner, Microsoft Gold Certified Partner, HP Gold Preferred partner, Cooler Master Golden Partner, ... u dobrom ste društvu!

Radno vreme:  
Pon-Sub: 9:00 – 23:00h  
Nedelja: 10:00 – 22:00h  
Posetite nas!!!

**Contact Info**

Email: marketing@desk.rs  
Website: <http://www.desk.rs>  
Location: Bulevar Mihajla Pupina 18-20, Čika Ljubina 18-20

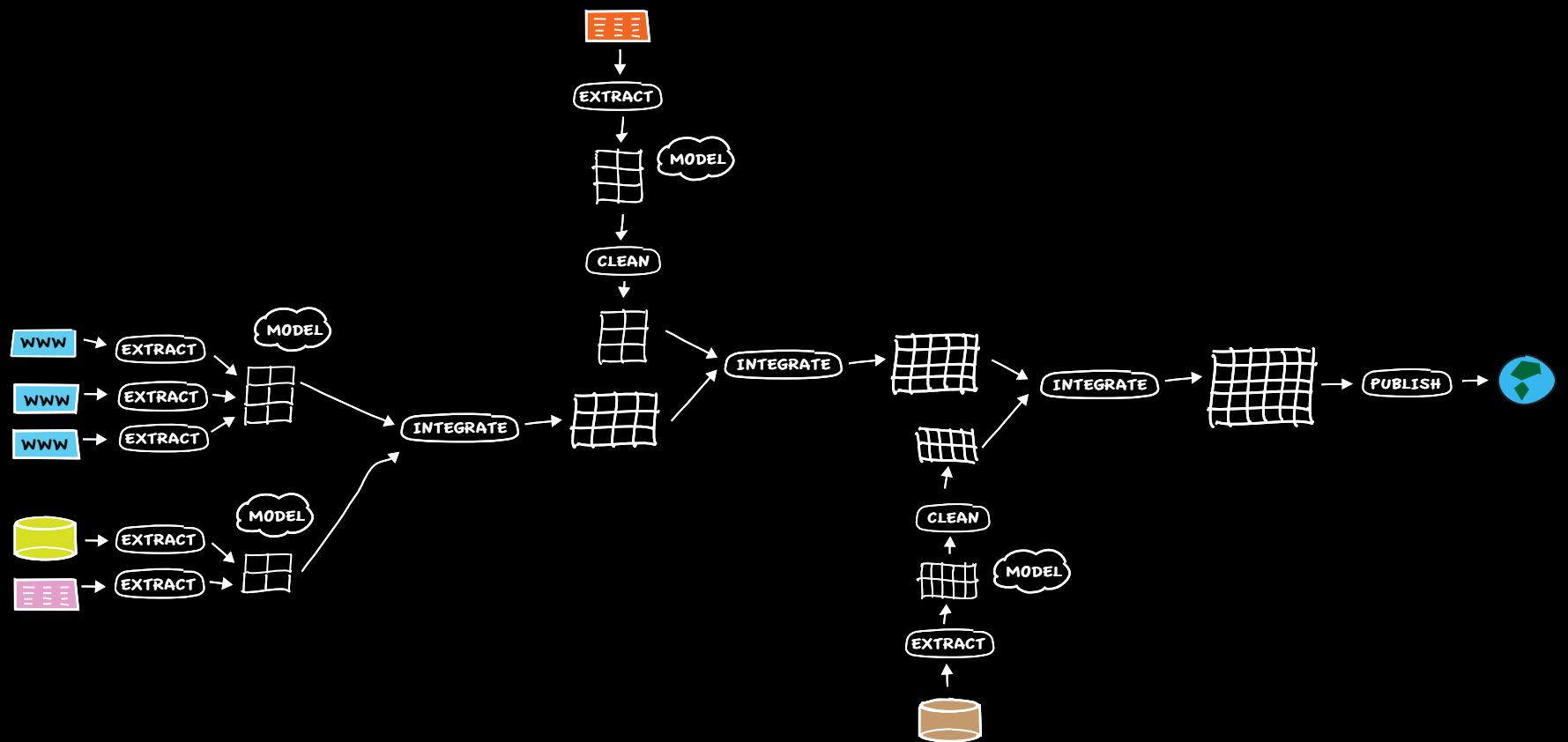
**Recent News**

News: Promenjeno je ime na DESK SHOP PLATO

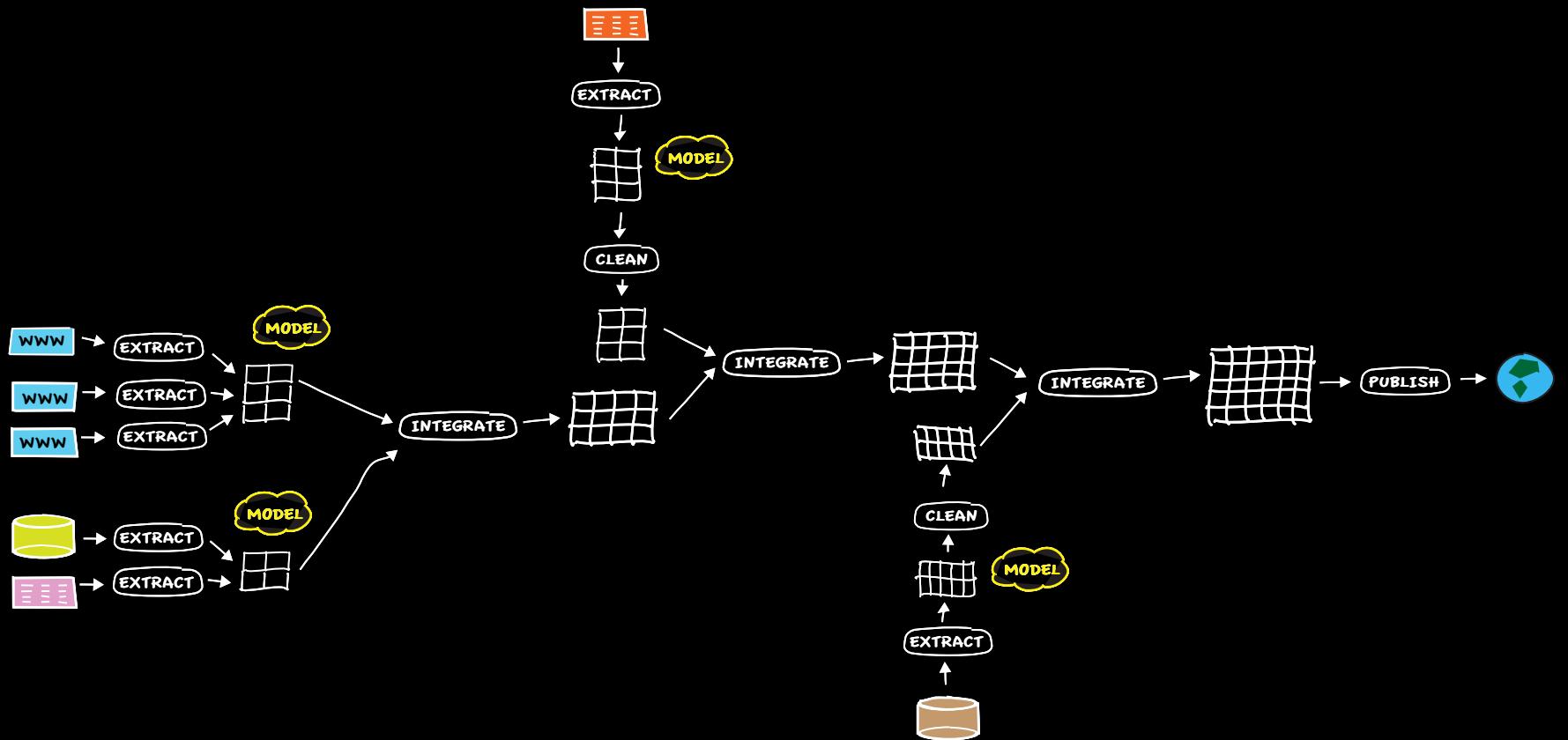
Privacy Type: Open: All content is public.

NOVA prava za DESK SHOP PLATO

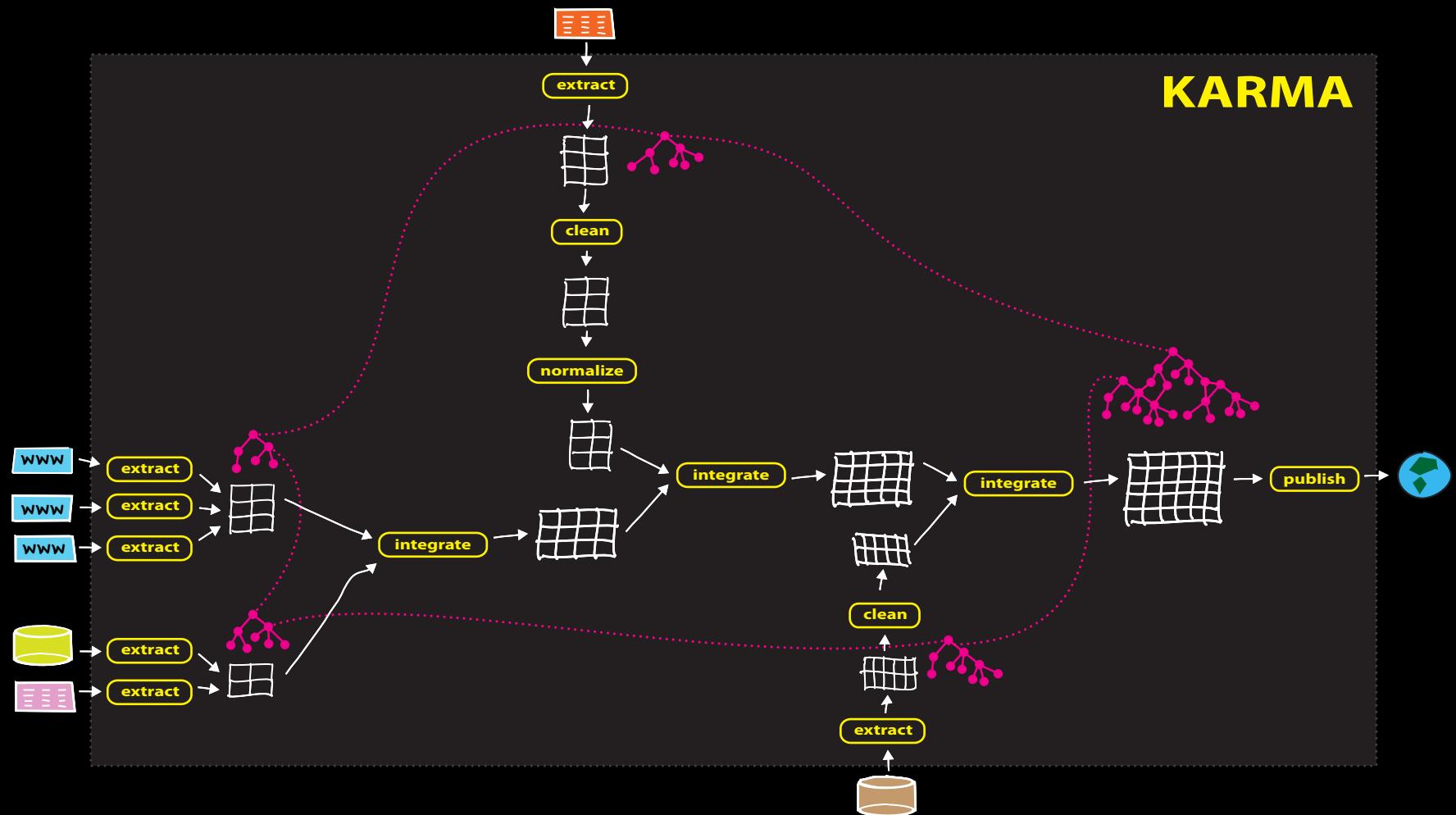
# Information Integration Process



# Users Model Data in Their Head

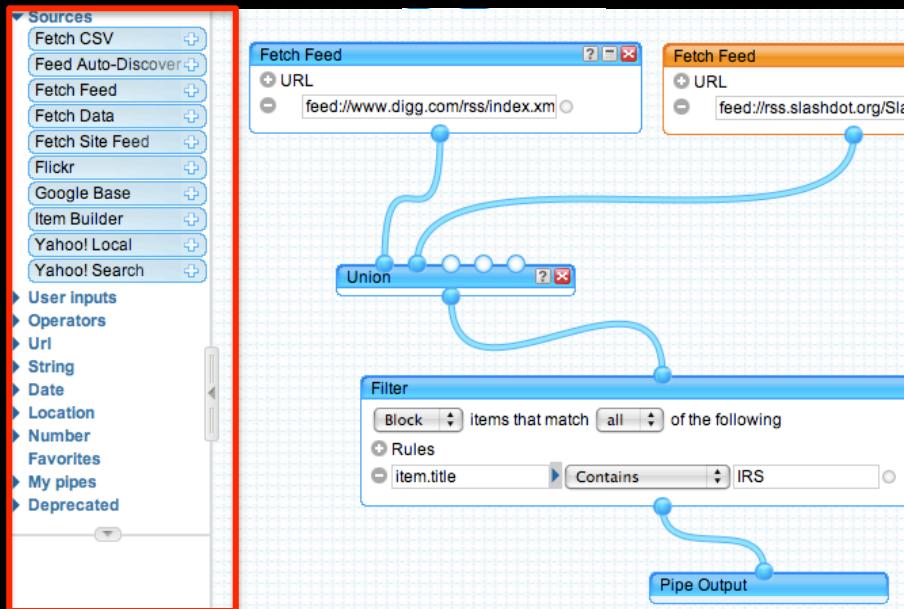


# Goal: “By Example” Support for All Data Integration Tasks



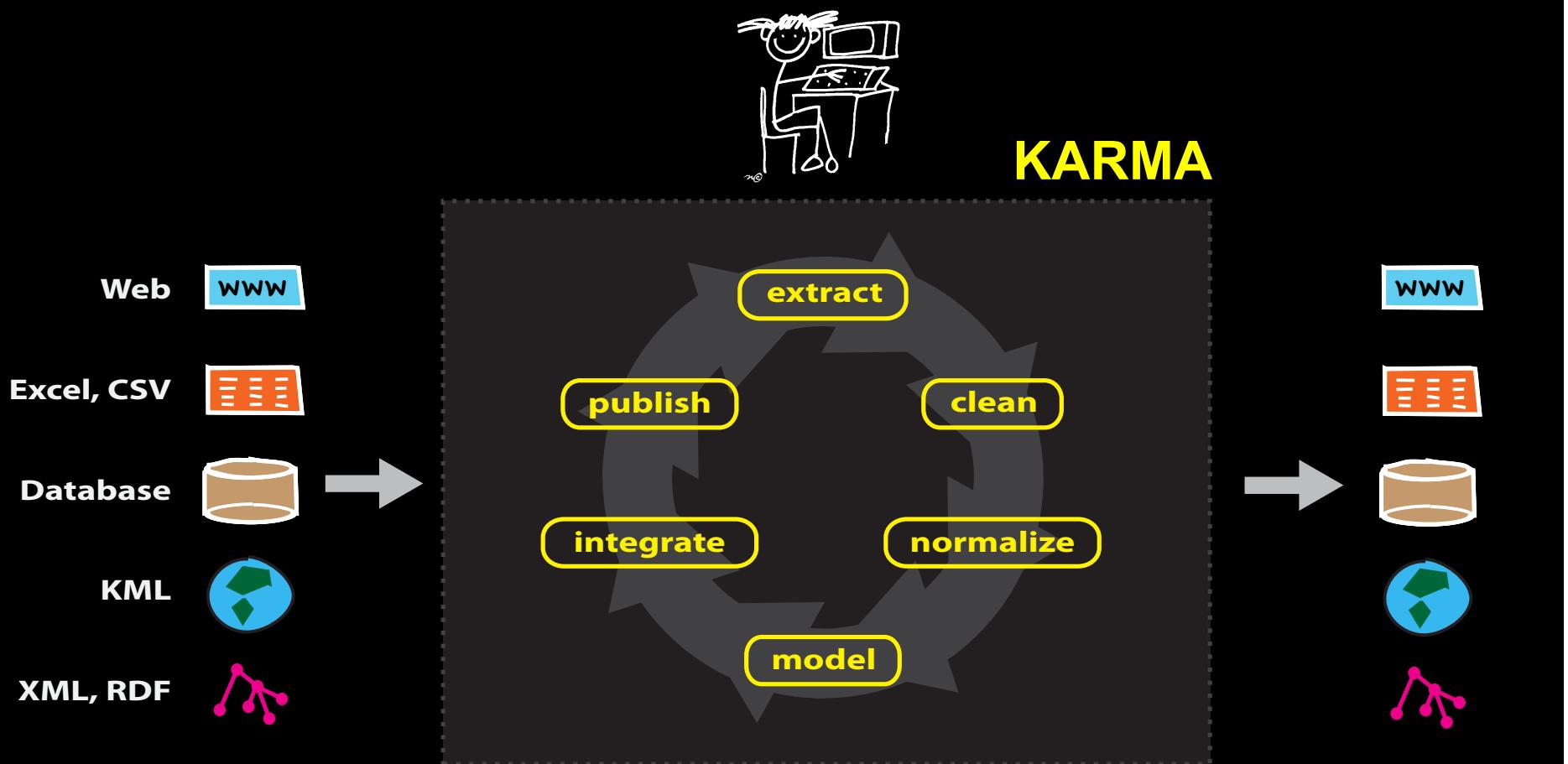
# Existing Approaches

Existing solutions require programming knowledge

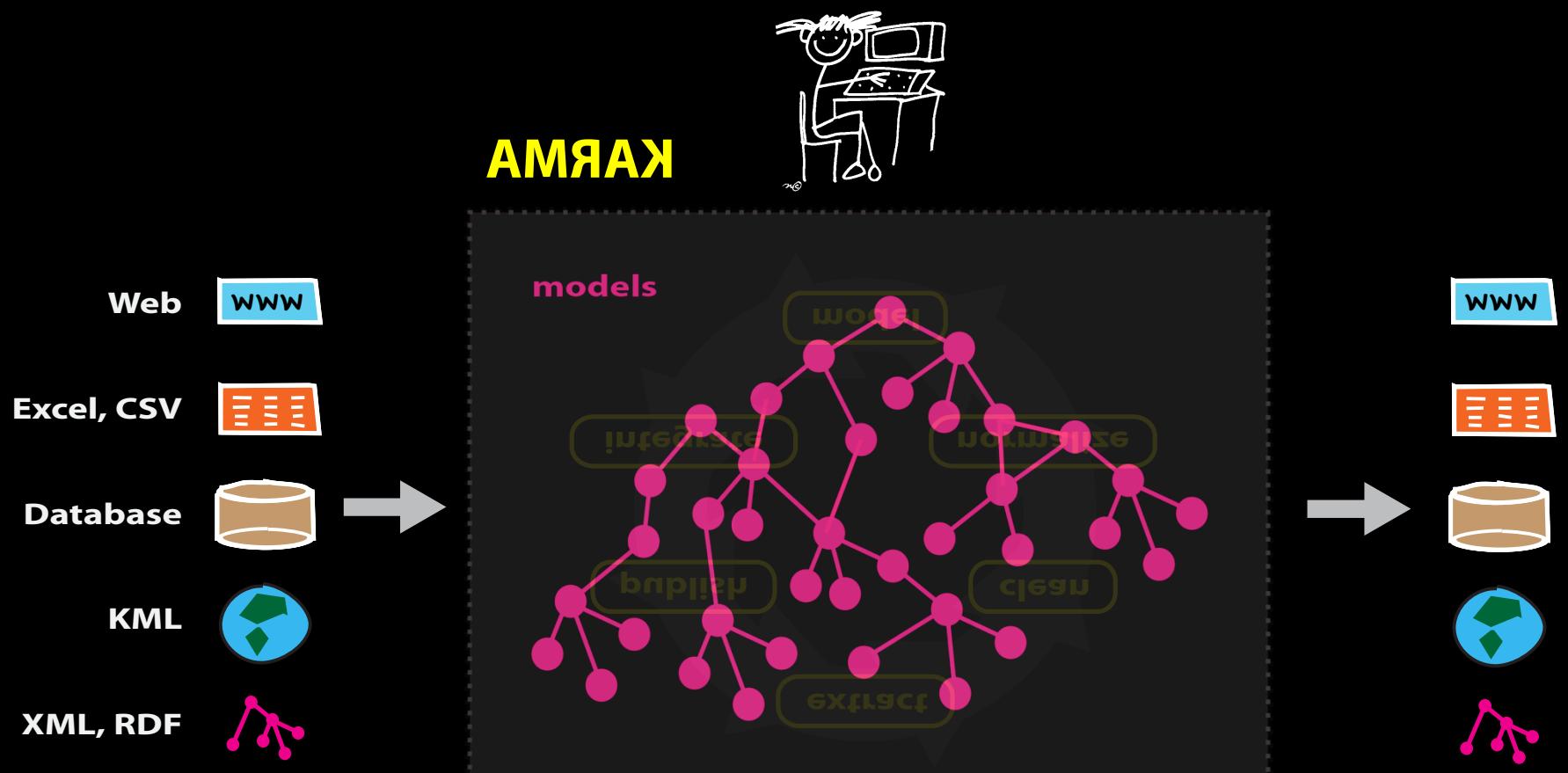


- Widget Paradigm
  - 43 for Pipes
  - 300+ for MS
- Focus on data manipulation
  - Difficult to find appropriate widgets

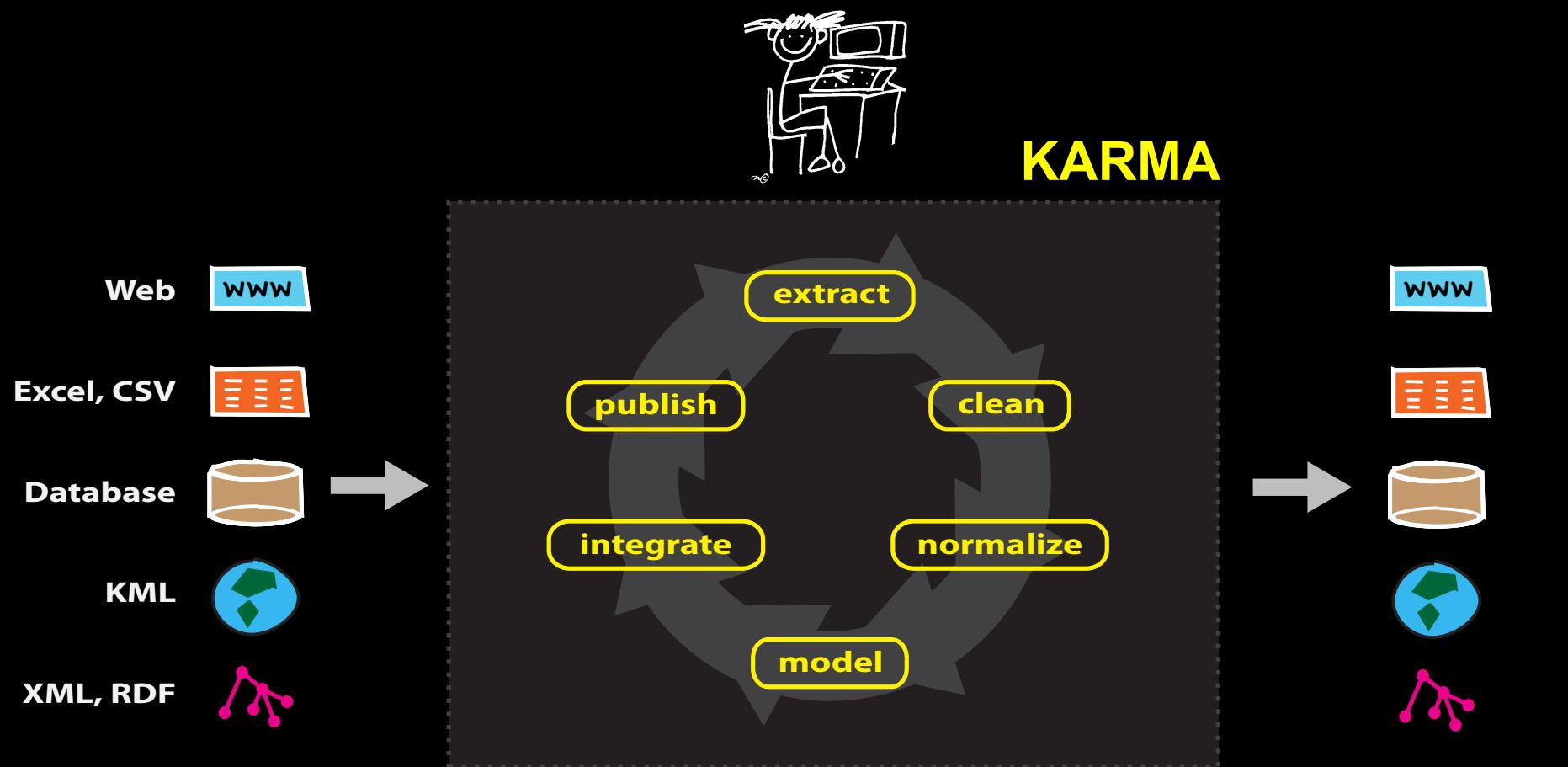
# Goal: “By Example” Support for All Data Integration Tasks



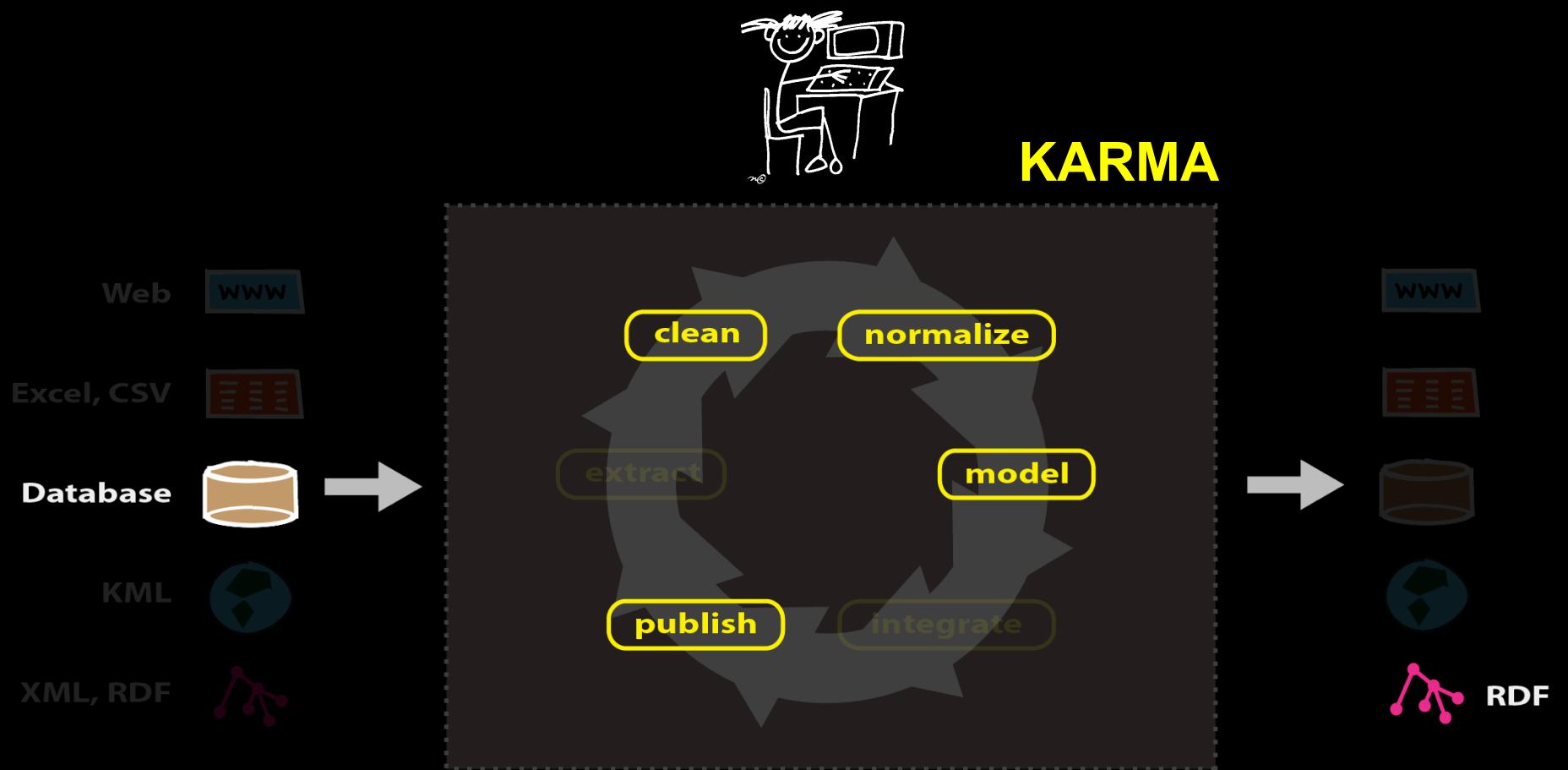
# Models, the Key Ingredient



# Goal: “By Example” Support for All Data Integration Tasks



# Focus of our KDD Effort

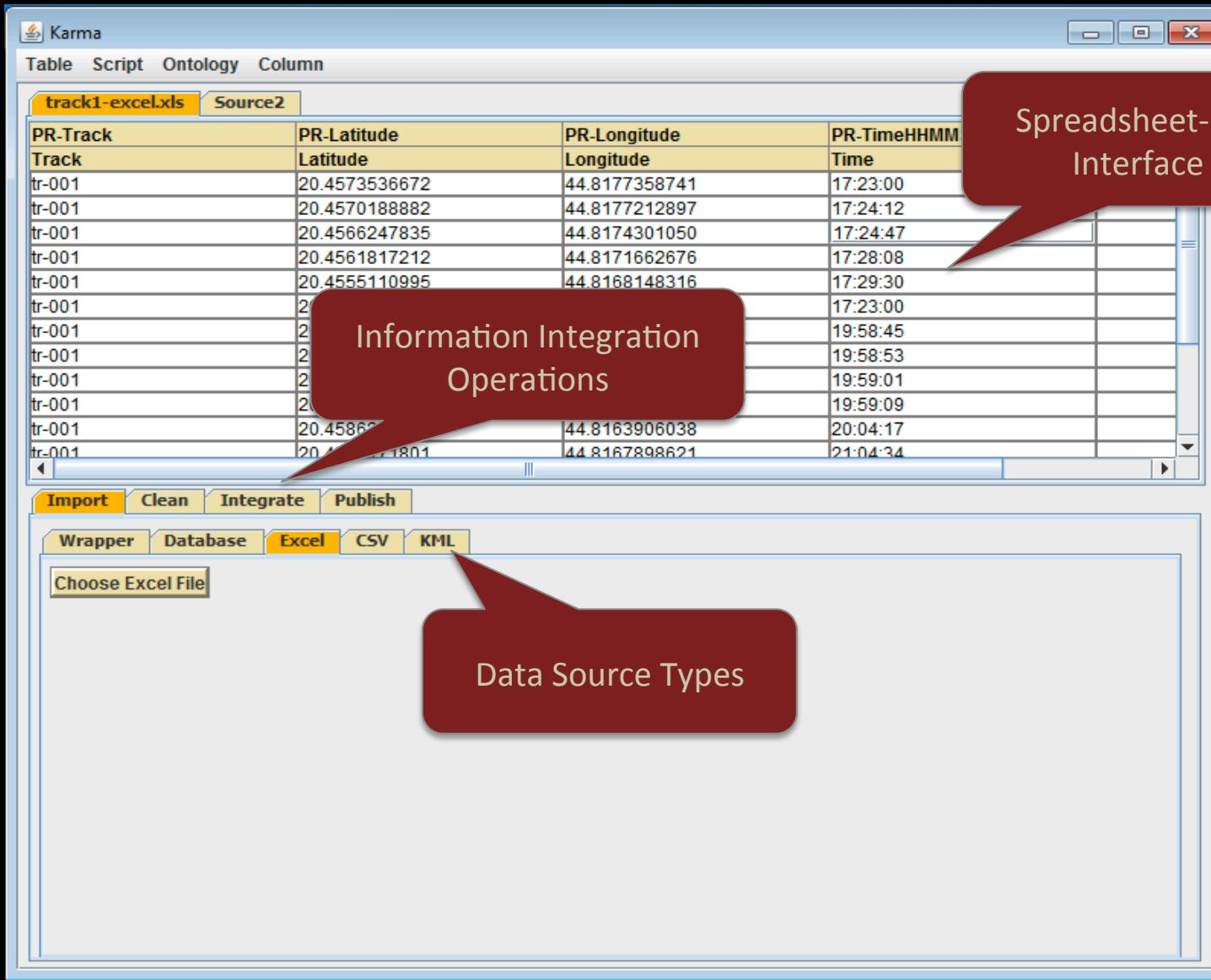




# Information Integration by Example

- We are developing an integration-by-example approach to extracting, modeling, cleaning, and integrating sources
  - Does not require any programming
  - Enables user to create an integrated view of the sources by demonstrating the task on one or more examples
  - Allows a user to interactively apply their knowledge of a task and the sources
  - Results can be immediately displayed or exported in various formats

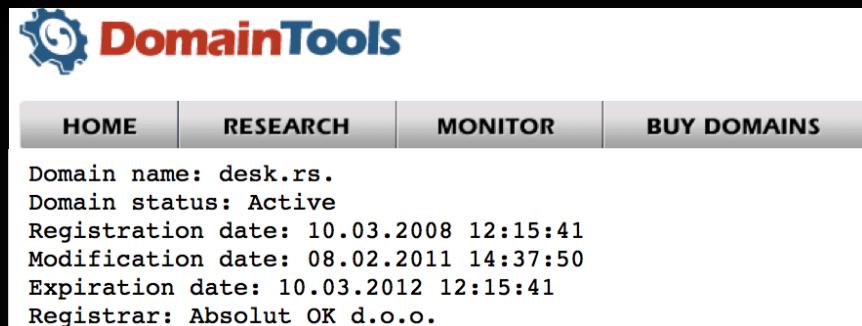
# Karma System



# Data Import

	A	B	C	D
1	Track	Latitude	Longitude	Time
2	tr-001	20.4573536672	44.8177358741	17:23:00
3	tr-001	20.4570188882	44.8177212897	17:24:12
4	tr-001	20.4566247835	44.8174301050	17:24:47
5	tr-001	20.4561817212	44.8171662676	17:28:08
6	tr-001	20.4555110995	44.8168148316	17:29:30
7	tr-001	20.4554824243	44.8166732121	17:23:00

Spreadsheet &amp; CSV



The screenshot shows the DomainTools website interface. At the top, there is a navigation bar with links for HOME, RESEARCH, MONITOR, and BUY DOMAINS. Below the navigation bar, the domain name "desk.rs." is displayed along with its status as "Active". Detailed registration information is provided, including the registration date (10.03.2008 12:15:41), modification date (08.02.2011 14:37:50), expiration date (10.03.2012 12:15:41), and the registrar (Absolut OK d.o.o.).

HTML Web Page



NAME	ADRESS	PHONE NO.
ADAMOVIĆ, VLAD	CARA LAZARA 15/2	011 2645986
ALEKSIĆ, DOBRILA	NIKOLE SPASIĆA 2	011 3263422
ZIMIĆ, FEĐA	CARA LAZARA 13/21	011 2372934
.....		

Database

```
<lineString>
  <extrude>1</extrude>
  <tessellate>1</tessellate>
  <altitudeMode>absolute</altitudeMode>
  <coordinates> -112.2550785337791,36.07954952145647,2357
    -112.2549277039738,36.08117083492122,2357
    -112.2552505069063,36.08260761307279,2357
  </coordinates>
</LineString>
</Placemark>
```

KML

# Import By Example

Examples provided using drag & drop



	A	B	C	PR-Track	PR-Latitude	PR-Longitude
1	Track	Latitude	Longitude	Track	Latitude	Longitude
2	tr-001	20.4573536672	44.8177358741	tr-001	20.4573536672	44.8177358741
3	tr-001	20.4570188882	44.8177212897	tr-001	20.4570188882	44.8177212897
4	tr-001	20.4566247835	44.8174301050	tr-001	20.4566247835	44.8174301050
5	tr-001	20.4561817212	44.8171662676	tr-001	20.4561817212	44.8171662676
6	tr-001	20.4555110995	44.8168148316	tr-001	20.4555110995	44.8168148316
7	tr-001	20.4554824243	44.8166732121	tr-001	20.4554824243	44.8166732121
8	tr-001	20.4565241020	44.8163966851	tr-001	20.4565241020	44.8163966851
9	tr-001	20.4577144223	44.8159429667	tr-001	20.4577144223	44.8159429667

# Import from Web Sources

The screenshot shows the Karma application interface. At the top, there's a menu bar with 'Table', 'Script', 'Ontology', and 'Column'. Below the menu is a tab bar with 'YellowPages' (selected), 'DomainRegistry', and 'Source3'. The main area contains a data grid with columns: PR-URL, PR-BusinessName, PR-Address, PR-PhoneNumber, and PR-PersonName. The grid lists various business entities with their addresses and contact details. A red callout bubble points to the 'DomainRegistry' tab with the text: "Wrappers used like spreadsheet formulas". Below the grid, there are buttons for 'Import', 'Clean', 'Integrate', and 'Publish'. At the bottom, there's a 'Wrapper' tab selected, followed by tabs for 'Database', 'Excel', 'CSV', and 'KML'. A table titled 'Fetch Agents' lists several wrappers with their input and output parameter names and execution options.

Wrapper Name	Input Parameter Names	Output Parameter Names	Input values	Execute
Alacra	CompanyName	Name, CompanyType, Q...	Manually Input Value	Execute
DomainRegistry	URL	WebAddress, Business...	WebAddress	Execute
GoogleNews	Keyword	HeadLines, Summary, U...	Manually Input Value	Execute
FlightStats	Airline	Airport, Code, Time, Len...	WebAddress	Execute
SwissPfamB	SwissProtEntryName	Source, Domain, Start, End	BusinessName	Execute

# Source Modeling: Semantic Typing

- Assign semantic type of data in columns
- Infer types using supervised machine learning

PR-String
Address
CARA LAZARA 15/2, BEOGRAD
CARA LAZARA 13/21, BEOGRAD
NIKOŁE SPASIĆA 2, BEOGRAD
USKOĆKA 7/008, BEOGRAD
NIKOŁE SPASIĆA 4/9, BEOGRAD
NIKOŁE SPASIĆA 5, BEOGRAD
CARA LAZARA 7, BEOGRAD

Initial Semantic Type



PR-Address'
Address
CARA LAZARA 15/2, BEOGRAD
CARA LAZARA 13/21, BEOGRAD
NIKOŁE SPASIĆA 2, BEOGRAD
USKOĆKA 7/008, BEOGRAD
NIKOŁE SPASIĆA 4/9, BEOGRAD
NIKOŁE SPASIĆA 5, BEOGRAD
CARA LAZARA 7, BEOGRAD

Manual labeling

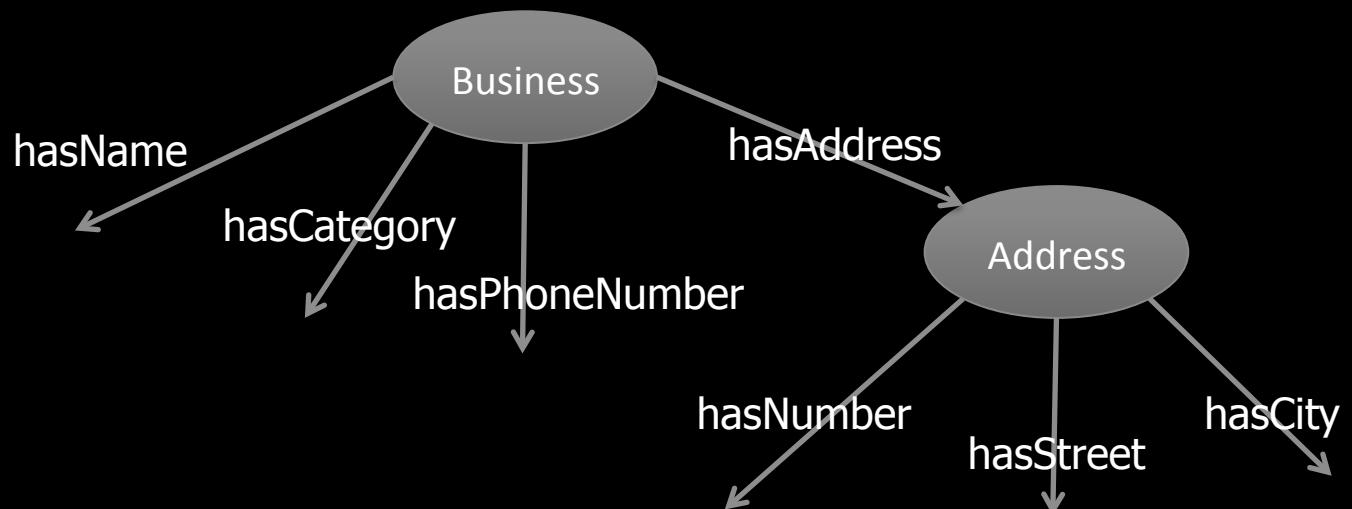
PR-Address
address
AILOVI... Francuska 23,Beograd
Kraljice Natalije 39,Beograd
GRAD Zemunski put 3,Beograd
Bulevar umetnosti 16,Beograd
Kraljice Marije 1/9,Beograd
RGOVI... Gospodar Jovanova 11,Beograd

New data set:  
semantic type correctly assigned

# Source Modeling

yellowpage.csv		Source2			
PR-hasName	PR-hasCategory	PR-hasPhoneNumber	PR-hasNumber	PR-hasStreet	PR-hasCity
Name	Category	Phone	Number	Street	City
Drzavna Lutrija Srbije	igre na srecu	011 202 9292	6	Uskocka	Beograd
Fun casino	kazina	011 627 605	4	Uskocka	Beograd
Grand t.t.	turisticke agencije	011 328 4955	7	Uskocka	Beograd
Menjacnica srbijalot	menjacnice	011 263 1846	4	Uskocka	Beograd
Optika	optika	011 263 3625	7	Uskocka	Beograd
Udruzenje samostalnih u...	fotografi	011 263 2562	5	Uskocka	Beograd
Anicic Slavko	advokati	063 372 034	3	Cara Lazara	Beograd

BELGRADE YELLOW PAGES DATA AFTER SEMANTIC TYPING



ONTOLOGY FOR DESCRIBING A SOURCE

# Source Modeling

yellowpage.csv		Source2			
PR-hasName	PR-hasCategory	PR-hasPhoneNumber	PR-hasNumber	PR-hasStreet	PR-hasCity
Name	Category	Phone	Number	Street	City
Drzavna Lutrija Srbije	igre na srecu	011 202 9292	6	Uskocka	Beograd
Fun casino	kazina	011 627 605	4	Uskocka	Beograd
Grand t.t.	turisticke agencije	011 328 4955	7	Uskocka	Beograd
Menjacnica srbijalot	menjacnice	011 263 1846	4	Uskocka	Beograd
Optika	optika	011 263 3625	7	Uskocka	Beograd
Udruzenje samostalnih u...	fotografi	011 263 2562	5	Uskocka	Beograd
Anicic Slavko	advokati	063 372 034	3	Cara Lazara	Beograd

BELGRADE YELLOW PAGES DATA AFTER SEMANTIC TYPING

yellowpage.csv		Source2			
Business					
(hasName) Thing	(hasCategory) Thing	(hasPhoneNumber) Thing	(hasAddress) Address		
PR-hasName	PR-hasCategory	PR-hasPhoneNumber	PR-hasNumber	PR-hasStreet	PR-hasCity
Name	Category	Phone	Number	Street	City
Drzavna Lutrija Srbije	igre na srecu	011 202 9292	6	Uskocka	Beograd
Fun casino	kazina	011 627 605	4	Uskocka	Beograd
Grand t.t.	turisticke agencije	011 328 4955	7	Uskocka	Beograd
Menjacnica srbijalot	menjacnice	011 263 1846	4	Uskocka	Beograd

BELGRADE YELLOW PAGES DATA AFTER SOURCE MODELING

# Data Cleaning

Cleaning rules learned from the user examples

PR-String	PR-Address
Name	Address
ADAMOVIĆ, VLADIMIR	CARA LAZARA 15/2, BEOGRAD
ALEKSIĆ, DOBRILA	CARA LAZARA 13/21, BEOGRAD
ZIMIĆ, FEĐA	NIKOŁE SPASIĆA 2, BEOGRAD
RISTIĆ, LJILJANA	USKOČKA 7/008, BEOGRAD
ĆIRIĆ, OLIVERA	NIKOŁE SPASIĆA 4/9, BEOGRAD

Initial data source



PR-Address	Data Type
Address	User Defined Values
CARA LAZARA 15/2, BEOGRAD	CARA LAZARA'
CARA LAZARA 13/21, BEOGRAD	
NIKOŁE SPASIĆA 2, BEOGRAD	
USKOČKA 7/008, BEOGRAD	
NIKOŁE SPASIĆA 4/9, BEOGRAD	

User provides example



PR-String	PR-Address
Name	Address
ADAMOVIĆ, VLADIMIR	CARA LAZARA
ALEKSIĆ, DOBRILA	CARA LAZARA
ZIMIĆ, FEĐA	NIKOŁE SPASIĆA
RISTIĆ, LJILJANA	USKOČKA
ĆIRIĆ, OLIVERA	NIKOŁE SPASIĆA

Data source after cleaning



PR-Address	Data Type
Address	User Defined Values
CARA LAZARA 15/2, BEOGRAD	CARA LAZARA'
CARA LAZARA 13/21, BEOGRAD	CARA LAZARA
NIKOŁE SPASIĆA 2, BEOGRAD	NIKOŁE SPASIĆA
USKOČKA 7/008, BEOGRAD	USKOČKA
NIKOŁE SPASIĆA 4/9, BEOGRAD	NIKOŁE SPASIĆA

Learned rule applied to remaining data

# Data Integration

- Column augmentation
  - Adding columns via joins across sources
- Row augmentation
  - Adding rows via union of sources with the same type of information

PR-String	PR-String	PR-String	PR-String
Coordinates	PlacemarkType	Name	Column Name
20.453399419784546,...	Line String	cara lazara	
20.45507311820984,4...	Line String	nikole spasica	
20.45578122138977,4...	Line String	uskocka'	

Column augmentation

Row augmentation

# Data Integration: Join Across Sources

	PR-String	PR-String	PR-String	Data Type
tes	PlacemarkType	street	Description	Select Column Name
9419784546,...	Line String	cara lazara		Select Column Name
311820984,4...	Line String	nikole spasica		Phone
122138977,4...	Line String	uskocka		Name

Karma suggests possible column augmentations



Name	Street	Phone
ADAMOVIĆ, VLAD	CARA LAZARA	011 2645986
ALEKSIĆ, DOBRILA	NIKOLE SPASIĆA	011 3263422
ZIMIĆ, FEĐA	CARA LAZARA	011 2372934
.....		

Karma collects joinable columns



PR-String	PR-String	PR-String	PR-String	PR-String
Coordinates	PlacemarkType	Street	Description	Name
20.4533994197845...	Line String	cara lazara		ADAMOVIĆ, VLADIMIR
20.4533994197845...	Line String	cara lazara		ALEKSIĆ, DOBRILA
20.4533994197845...	Line String	cara lazara		ANDREJIĆ, LJILJANA
20.4533994197845...	Line String	cara lazara		ANTIĆ, KATICA
20.4533994197845...	Line String	cara lazara		APOSTOLOVIĆ, DUŠAN
20.4533994197845...	Line String	cara lazara		APOSTOLOVIĆ, MILAN
20.4533994197845...	Line String	cara lazara		APOSTOLOVIĆ, MILAN
20.4533994197845...	Line String	cara lazara		PARIĆ, MIURA

Karma executes join queries when users select the column to add

# Data Integration: Union of Sources

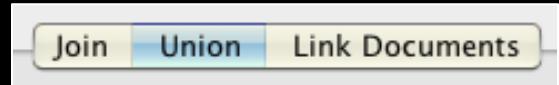
Union of sources based on common attribute names

PR-Address	PR-String	PR-String	PR-String	PR-Address
Name	category	Phone	weblink	Street
Fun casino	kazina	011 627 605		Uskočka
Grand t.t.	turističke agencije	011 328 4955		Uskočka
Menjačnica srbijalot	menjačnice	011 263 1846		Uskočka
Optika	optika	011 263 3625		Uskočka
Udruženje samostalnih umetničkih radnika	fotografi	011 263 2562		Uskočka
Anićić Slavko	advokati	063 372 034		Cara Lazara
Apropo	knjižare	011 262 5839		Cara Lazara
Atlantik	igre na sreću	011 262 3053		Cara Lazara
Chip card	bankarski sistemi	011 262 8961		Cara Lazara
Drakulić Branko	obućari	011 625 157		Cara Lazara

Yellow Pages

PR-String	PR-String	PR-Address
Phone	Name	Street
011 2624121	ADAMOVIĆ, VLADIMIR	CARA LAZARA
011 2624782	ALEKSIĆ, DOBRILA	CARA LAZARA
011 2621134	ZIMIĆ, FEĐA	NIKOLE SPASIĆA
011 2624019	RISTIĆ, LJILJANA	USKOČKA
011 2631334	ČIRIĆ, OLIVERA	NIKOLE SPASIĆA
011 2631334	ČVIIĆ, OLIVERA	NIKOLE SPASIĆA
011 2627416	ANDREJIĆ, LJILJANA	CARA LAZARA
011 2620376	ANTIĆ, KATICA	CARA LAZARA

White Pages



PR-String	PR-String	PR-Address
Phone	Name	Street
011 2624018	SERDAR, VLADIMIR	USKOČKA
011 2620570	TADIĆ, SMILJKA	USKOČKA
011 2623130	TOMIĆ, MILOŠ	USKOČKA
011 2623251	ČURČIN, BOGDAN	USKOČKA
011 627 605	Fun casino	Uskočka
011 328 4955	Grand t.t.	Uskočka
011 263 1846	Menjačnica srbijalot	Uskočka
011 263 3625	Optika	Uskočka
011 263 2562	Udruženje samostalnih umetničkih radnika	Uskočka
063 372 034	Anićić Slavko	Cara Lazara
011 262 5839	Apropo	Cara Lazara

People from  
White Pages

Businesses from  
Yellow Pages

# Data Publishing

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

## Different mashup publishing options

# Data Display

- Data displayed on a map or image based on geospatial attributes
- Display continuously updated

Karma

Table Script Ontology Column

YellowPages DomainRegistry Source3

PR-BusinessName	PR-Address	PR-PhoneNumber	PR-URL
BusinessName	Address	Phone	WebAddress
Desk Computers	110 Bulevard despota Stefan...	(0)11 329 2140	<a href="http://www.desk.rs">http://www.desk.rs</a>
UTS/PAS Prosecutors Ass...	19 Alekse Nenadovica, Beo...	(0)11 303 1909	<a href="http://www.uts-pas.org">http://www.uts-pas.org</a>
Galerija 12	12 Vuka Karadžića Belgrade	(0)11 262 6527	<a href="http://www.galerija12.com">http://www.galerija12.com</a>
Apropo	10 Nikola Spasica Belgra...	(0)11 262 5839	<a href="http://www.apropo.rs">http://www.apropo.rs</a>
Akademija	10 Cara Lazara Belgrade ...	(0)64 647 8870	<a href="http://www.akademija.rs">http://www.akademija.rs</a>
Palace Hotel	23 Topliin Wreath, Belgrad...	(0)11 218 5585	<a href="http://www.palace.rs">http://www.palace.rs</a>
Tanjug news agency of Se...	2 Cara Wreath, Belgrade, ...	(0)11 328 7437	<a href="http://www.tanjug.rs">http://www.tanjug.rs</a>
NEW SMILE Stomatoloka ...	45/2 Marshal Birjuzova, Be...	(0)11 305 8118	<a href="http://www.smile.rs">http://www.smile.rs</a>
Ethno Fusion Fest	19 Marshal Birjuzova, Belg...	(0)11 381 1124	<a href="http://www.ethnofusionfest.com">http://www.ethnofusionfest.com</a>
Majestic	28 Obiličev venac, Belgrade...	(0)11 328 5777	<a href="http://www.majestic.rs">http://www.majestic.rs</a>



# Data Display

- Data displayed on a map or image based on geospatial attributes
- Display continuously updated

Karma

Table Script Ontology Column

YellowPages DomainRegistry Source3

PR-BusinessName	PR-Address	PR-PhoneNumber	PR-URL
BusinessName	Address	Phone	WebAddress
Desk Computers	110 BulevardespostaStefan...	(0)11 329 2140	<a href="http://www.desk.rs">http://www.desk.rs</a>
UTS/PAS Prosecutors Ass...	19 AlekseNenadovica,Beo...	(0)11 303 1909	<a href="http://www.uts-pas.org">http://www.uts-pas.org</a>
Galerija 12	12 Vuka Karadja Belgrade	(0)11 262 6527	<a href="http://www.galerija12.com">http://www.galerija12.com</a>
Apropo	10 Nikola Spasica Belgra...	(0)11 262 5839	<a href="http://www.apropo.rs">http://www.apropo.rs</a>
Akademija	10 Cara Lazara Belgrade ...	(0)64 647 8870	<a href="http://www.akademija.rs">http://www.akademija.rs</a>
Palace Hotel	23 Topliin Wreath, Belgrad...	(0)11 218 5585	<a href="http://www.palace.rs">http://www.palace.rs</a>
Tanjug news agency of Se...	2 Cara Wreath, Belgrade, ...	(0)11 328 7437	<a href="http://www.tanjug.rs">http://www.tanjug.rs</a>
NEW SMILE Stomatoloka ...	45/2 Marshal Birjuzova, Be...	(0)11 305 8118	<a href="http://www.smile.rs">http://www.smile.rs</a>
Ethno Fusion Fest	19 Marshal Birjuzova, Belg...	(0)11 381 1124	<a href="http://www.ethnofusionfest.com">http://www.ethnofusionfest.com</a>
Majestic	28 Obiliev venac, Belgrade...	(0)11 328 5777	<a href="http://www.majestic.rs">http://www.majestic.rs</a>



# Current Work

- Incorporate explicit semantics
  - Build and use a domain ontology for source modeling, data cleaning, and data integration
  - Using this domain ontology, represent, reason, and publish data using RDF
- Develop improved machine learning methods for source modeling
  - Build and use graphical models to improve the automatic modeling of new sources
  - Develop an interactive approach to source alignment using the domain ontology
- Data Normalization
  - Build methods for learning transformations by example
- Data Cleaning
  - Learn to identify data to be cleaned and learn methods for cleaning

# Discussion

- There is a role for discovery and analysis over massive networks to identify activities that might go unnoticed
- But...we also need to think about tools that allow analysts to drive the process and exploits their knowledge and intuitions
- This requires tools that
  - Provide access to a wide variety of sources
  - Makes it possible to easily combine the data in novel ways
  - Allows an analyst to visualize and understand data

