



# Knowledge Graph Enrichment Scenarios with PoolParty Semantic Integrator

The background of the slide is a light blue-grey color. It features a top-down view of a person's hands interacting with a tablet. The tablet displays a colorful Venn diagram with four overlapping circles in green, yellow, red, and blue, each containing a number. To the right of the tablet, there is a complex network diagram consisting of white nodes connected by thin lines, with a globe-like structure on the far right. In the bottom right corner, there is a white circular icon containing a gear and a document symbol.

**Tomas Knap, PhD**  
Software Engineer and Researcher  
Semantic Web Company

# Introducing Semantic Web Company

**Founded** in **2004**

**Based** in Vienna

**Privately held**

**50+** FTE

Software Engineers & Consultants for **NLP**, **Semantics** and **Machine learning**

Developer & Vendor of **PoolParty Semantic Suite**



Participating in projects with €2.5 million funding for **R&D**



**~35%**  
revenue growth/year



SWC named to KMWorld's **'100 Companies That Matter in Knowledge Management'** in 2016, until 2019



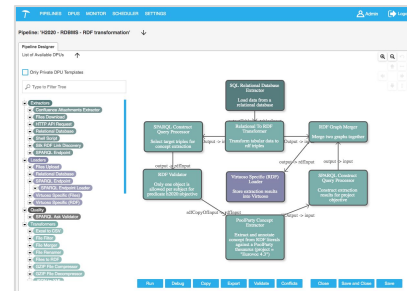
**ISO 27001:2013** certified



4,500 followers

# Agenda

- ▶ PoolParty Semantic Suite
- ▶ Knowledge Graphs (KG)
- ▶ KG Enrichment Scenarios
  - ▷ Unstructured Data
    - Discover concepts in the documents
    - Link concepts to external knowledge graphs
    - ...
  - ▷ Structured Data - CSV, JSON, relational tables
    - Align schemas of entities
    - Link attributes to the entities in the KG
      - Discover duplicates in the KG
    - Fuse different representations of entities
    - ...
- ▶ Summary



Bain Capital is a venture capital company based in Boston, MA. Since inception it has invested in hundreds of companies including AMC Entertainment, Brookstone, and Burger King. The company was co-founded by Mitt Romney.

Vendor City	Vendor State	Vendor Zip Code
Fairfax	VA	22030
Washington	DC	20016
Mount Airy	MD	21771
Washington	DC	20022
Rockville	MD	20851
Landover Hills	MD	20784
Potomac	MD	20854

# PoolParty Suite

# Fact sheet: PoolParty Semantic Suite

Most complete and secure  
**Semantic Middleware** on  
the Global Market



## Semantic AI:

Fusion Graphs, NLP,  
and Machine Learning

**ISO** 27001:2013 certified  
**W3C** standards compliant

First release in **2009**

Current version **7.1**

**On-premises** or  
**cloud-based**



Over **200** installations  
world-wide



Named as Sample  
Vendor in  
**Gartner's Hype  
Cycle for AI 2019**



Named as  
Representative  
Vendor in **Gartner's  
Market Guide for  
Hosted AI Services  
2018**

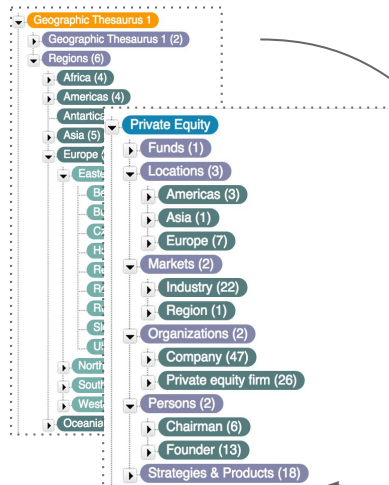


**KMWorld** listed  
PoolParty as  
**Trend-Setting  
Product** 2015, 2016,  
2017, 2018, and 2019

# How does it work?

The Knowledge Engineer's perspective

## Taxonomy & Ontology Server



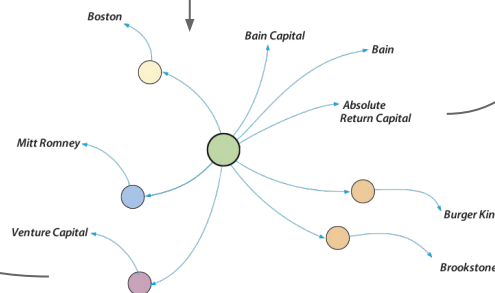
Identify new candidate concepts to be included in a controlled vocabulary

## Entity Extractor & Semantic Classifier

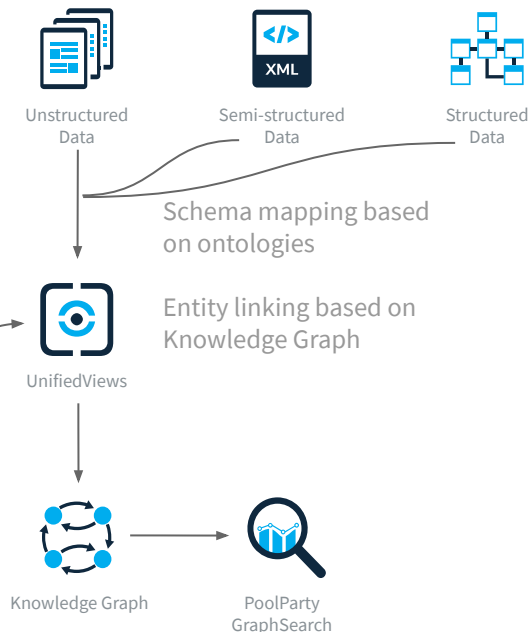
Controlled vocabularies as a basis for highly precise knowledge extraction and text classification

**Factsheet**

Bain Capital is a venture capital company based in Boston, MA. Since inception it has invested in hundreds of companies including AMC Entertainment, Brookstone, and Burger King. The company was co-founded by Mitt Romney.

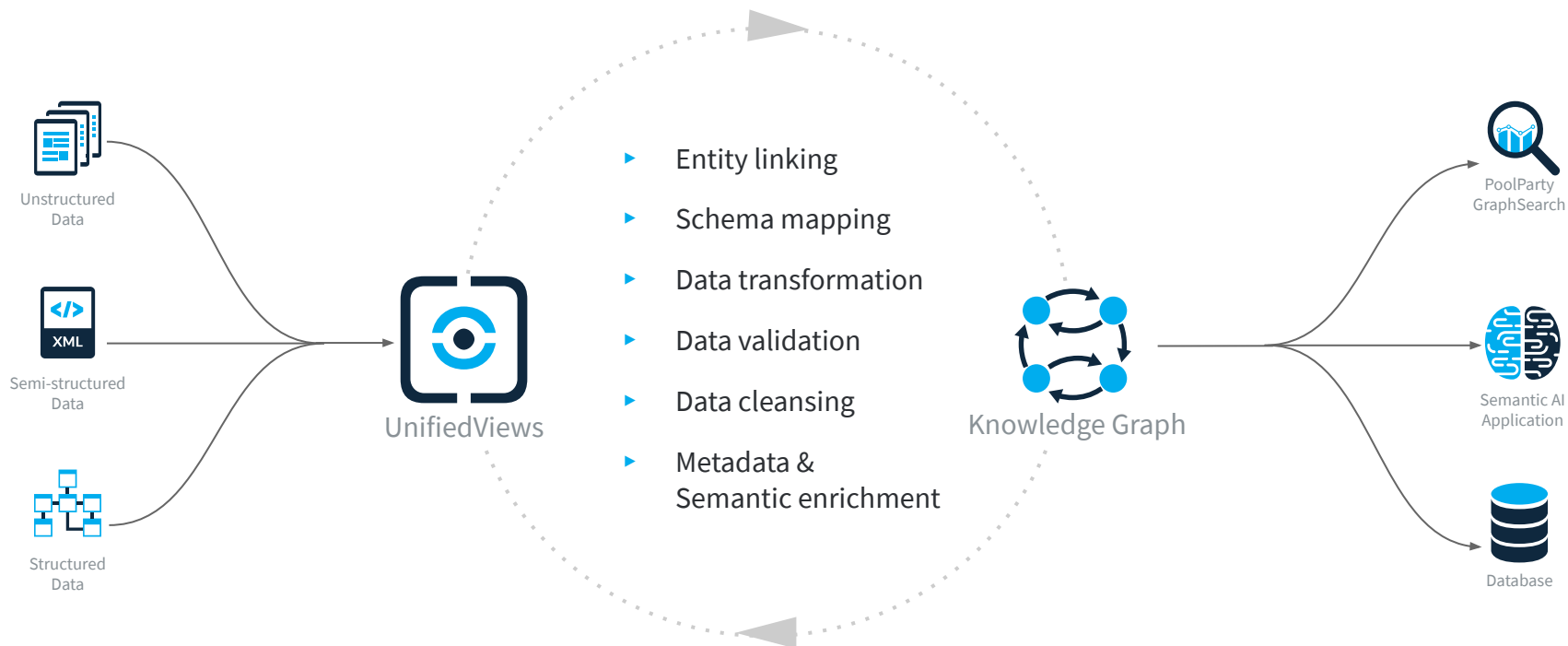


## Data Integration & Data Linking



# How does it work?

The Data Engineer's perspective



# PoolParty UnifiedViews for Data Engineers

## Knowledge Graph Automation and Linked Data Orchestration

PIPELINES DPUS MONITOR SCHEDULER SETTINGS

QUICK START

MANUAL

SUPPORT

FAQ

VIDEOS

WEBSITE

Semantic Web Company GmbH  
Vienna, Austria  
support@poolparty.biz

Actions	Name	Last run time
	CitySPIN Energy Pipe	0.00.00
	H2020 - RDBMS - RDF transformation	0.00.03
	JSON to XML	
	MS Azure Activity	
	RDF Data Loading Example	
	Test	0.00.01
	Test R2Rml	0.00.00
	test xsit	
	test xsit 2	

PIPELINES DPUS MONITOR SCHEDULER SETTINGS Admin Logout

Pipeline: 'H2020 - RDBMS - RDF transformation'

Pipeline Designer

List of Available DPUs

Only Private DPU Templates

Type to Filter Tree

- Extractors
  - Confluence Attachments Extractor
  - Files Download
  - HTTP API Request
  - Relational Database
  - Shell Script
  - Silk RDF Link Discovery
  - SPARQL Endpoint
- Loaders
  - Files Upload
  - Relational Database
  - SPARQL Endpoint
  - SPARQL Endpoint Loader
  - Virtuoso Specific (Files)
  - Virtuoso Specific (RDF)
- Quality
  - SPARQL Ask Validator
- Transformers
  - Excel to CSV
  - File Filter
  - File Merger
  - File Renamer
  - Files to RDF
  - GZIP File Compressor
  - GZIP File Decompressor

Run Debug Copy Export Validate Conflicts Close Save and Close Save

First Previous Page: 1 / 1 Next Last



# Knowledge Graphs

# Knowledge Graphs (KG)

- ▶ “Network of all kinds of things which are relevant to the organization”
  - ▶ created by subject-matter experts with the help of tools and machine learning algorithms
- ▶ KG provides common interface to query the data



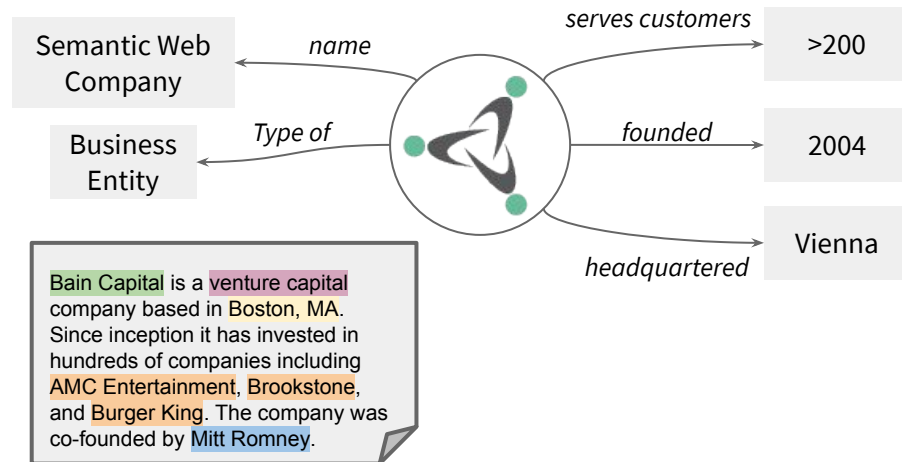
<https://graphviews.poolparty.biz/>

<https://semantic-web.com/2014/07/15/from-taxonomies-over-ontologies-to-knowledge-graphs/>

# Knowledge Graphs

- ▶ **Abstract concepts** from thesauri/taxonomies, together with **related** concepts
- ▶ Real world **entities**, with its **attributes and relations, classes of entities**
  - ▶ Semantic Web Company, Business Entity, Projects
- ▶ **Documents**
  - ▶ links to concepts/entities discovered

The image shows two screenshots. On the left is the 'STW Thesaurus for Economics' interface, displaying a hierarchical tree structure of categories such as 'A General descriptors (1)', 'B Business economics (11)', 'G Geographic names (8)', 'N Related subject areas (11)', 'P Commodities (24)', 'V Economics (19)', and 'W Economic sectors (35)'. The 'W.02 Forestry (19)' category is expanded, showing sub-categories like 'Agroforestry (0)', 'Deforestation (0)', 'Forest (1)', 'Forest damage (0)', 'Forest economics (0)', 'Forest policy (1)', 'Forestry (2)', and 'Forestry legislation (0)'. On the right is the 'Agroforestry' SKOS (Simple Knowledge Organization System) interface. It shows the 'Details' tab with sections for 'Broader Concepts' (Cultivation system, Forestry, Tropical agriculture, W.01.03 Farming Systems, W.02 Forestry), 'Narrower Concepts' (empty), and 'Related Concepts' (Tropical forest). The 'Preferred Label' is 'Agroforestry', and there are 'Alternative Labels', 'Hidden Labels', 'Notation', and 'Scope Notes' sections.

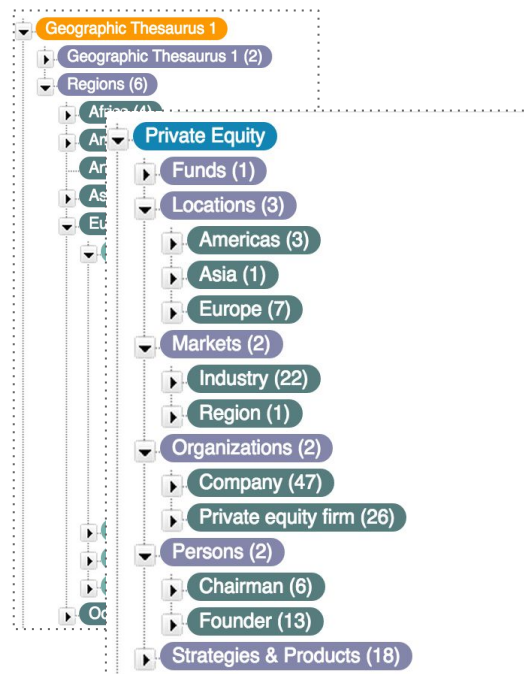


# KB Enrichment Scenarios

Unstructured Data

- ▶ Discover concepts in the documents based on the KG, concept co-occurrences
- ▶ Suggest new concepts
- ▶ **Link to concepts in external knowledge bases (DBpedia)**
  - Fetch attributes/relations
- ▶ Recognize (named) entities
- ▶ Classify documents

## Knowledge Graph



## Entity Extraction & Text Mining

Bain Capital is a venture capital company based in Boston, MA. Since inception it has invested in hundreds of companies including AMC Entertainment, Brookstone, and Burger King. The company was co-founded by Mitt Romney.

# Linking Concepts to DBpedia

- A guide to classic cocktails
  - Beverages (5)
  - Cocktails (6)
    - Apéritif and digestif (3)
    - Duo and trio cocktails (3)
    - IBA Official Cocktail (3)
    - Contemporary Classics (30)
      - Bellini (0)
      - Black Russian (0)
      - Bloody Mary (0)
      - Caipirinha (0)
      - Champagne cocktail (0)
      - Cosmopolitan (0)
      - Cuba Libre (0)
      - French 75 (0)
      - French Connection (0)
      - Godfather (0)
      - Godmother (0)
      - Golden dream (0)
      - Grasshopper (0)
      - Harvey Wallbanger (0)
      - Hemingway Special (0)
      - Horse's Neck (0)
      - Irish coffee (0)
      - Kir Royale (1)
      - Long Island Iced Tea (0)
      - Margarita (0)
      - Mimosa (0)
      - Mint julep (0)
      - Mojito (0)

## Cosmopolitan Approved

<http://vocabulary.semantic-web.at/cocktails/4f85cadd-4fd4-4ffc-a78d-cddb0763fa4b>

Assign Approve Reject Add to Collection Add to Blacklist Add to ExactMatch

Cocktail

### Lookup

#### Linked Data Settings

EnDBpedia

Cosmopolitan

Search

Select Scheme:

SKOS

Label	URI	Mapping Type
Cosmopolitan (cocktail)	<a href="http://dbpedia.org/resource/Cosmopolitan_(cocktail)">http://dbpedia.org/resource/Cosmopolitan_(cocktail)</a>	
Cosmopolitan	<a href="http://dbpedia.org/resource/Cosmopolitan">http://dbpedia.org/resource/Cosmopolitan</a>	
Bolboschoenus maritimus	<a href="http://dbpedia.org/resource/Bolboschoenus_maritimus">http://dbpedia.org/resource/Bolboschoenus_maritimus</a>	

Save

Save and Close

Cancel

- A guide to classic cocktails
- Beverages (5)
- Cocktails (6)
  - Apéritif and digestif (3)
  - Duo and trio cocktails (3)
  - IBA Official Cocktail (3)
    - Contemporary Classics (30)
      - Bellini (0)
      - Black Russian (0)
      - Bloody Mary (0)
      - Caipirinha (0)
      - Champagne cocktail (0)
      - Cosmopolitan (0)
      - Cuba Libre (0)
      - French 75 (0)
      - French Connection (0)
      - Godfather (0)
      - Godmother (0)
      - Golden dream (0)
      - Grasshopper (0)
      - Harvey Wallbanger (0)
      - Hemingway Special (0)
      - Horse's Neck (0)
      - Irish coffee (0)
      - Kir Royale (1)
      - Long Island Iced Tea (0)
      - Margarita (0)
      - Mimosa (0)
      - Mint julep (0)
      - Mojito (0)
      - Moscow mule (0)
      - Piña colada (0)
      - Rose (0)
      - Sea Breeze (0)

Cosmopolitan Approved

<http://vocabulary.semantic-web.at/cocktails/4f85cadd-4fd4-4ffc-a78d-cddb07631a4b>

Cocktail

Details Notes Documents **Linked Data** Triples Visualization Quality Management History

EnDBPedia EnDBPediaCategories GeoNames



Lookup

Linked Data Settings

EnDBPedia

Select Scheme:

Lookup

Linked Data Settings

EnDBPedia

Select Scheme:

Cosmopolitan Club	<a href="http://dbpedia.org/resource/Cosmopolitan_Club">http://dbpedia.org/resource/Cosmopolitan_Club</a>	<input type="text"/>
The Cosmopolitan	<a href="http://dbpedia.org/resource/The_Cosmopolitan">http://dbpedia.org/resource/The_Cosmopolitan</a>	<input type="text"/>
Cosmopolitan Bowl	<a href="http://dbpedia.org/resource/Cosmopolitan_Bowl">http://dbpedia.org/resource/Cosmopolitan_Bowl</a>	<input type="text"/>
Cosmopolitan Church	<a href="http://dbpedia.org/resource/Cosmopolitan_Church">http://dbpedia.org/resource/Cosmopolitan_Church</a>	<input type="text"/>
Cosmopolitan Dock	<a href="http://dbpedia.org/resource/Cosmopolitan_Dock">http://dbpedia.org/resource/Cosmopolitan_Dock</a>	<input type="text"/>
Cosmopolitan Girl	<a href="http://dbpedia.org/resource/Cosmopolitan_Girl">http://dbpedia.org/resource/Cosmopolitan_Girl</a>	<input type="text"/>

# About: Cosmopolitan (cocktail)

An Entity of Type : Food, from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](http://dbpedia.org)

A cosmopolitan, or informally a cosmo, is a cocktail made with vodka, triple sec, cranberry juice, and freshly squeezed or sweetened lime juice.

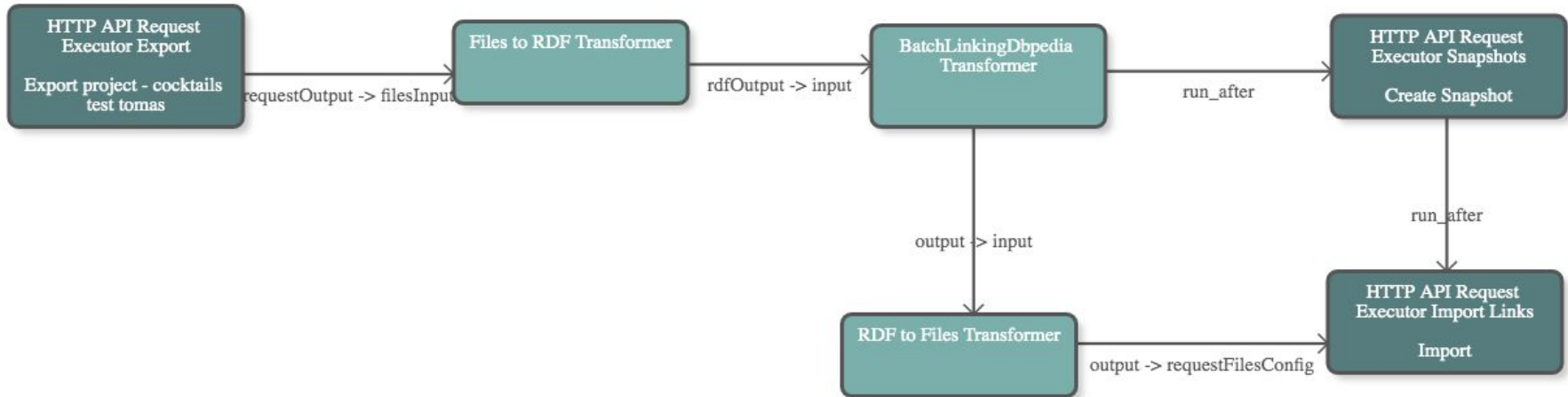
Property	Value
<a href="#">dbo:abstract</a>	<ul style="list-style-type: none"><li>A cosmopolitan, or informally a cosmo, is a cocktail made with vodka, triple sec, cranberry juice, and freshly squeezed or sweetened lime juice. <small>(en)</small></li></ul>
<a href="#">dbo:thumbnail</a>	<ul style="list-style-type: none"><li><a href="https://commons.wikimedia.org/wiki/Special:FilePath/Cosmopolitan_(5076906532).jpg?width=300">wiki-commons:Special:FilePath/Cosmopolitan_(5076906532).jpg?width=300</a></li></ul>
<a href="#">dbo:wikiPageID</a>	<ul style="list-style-type: none"><li>409113 <small>(xsd:integer)</small></li></ul>
<a href="#">dbo:wikiPageRevisionID</a>	<ul style="list-style-type: none"><li>732587202 <small>(xsd:integer)</small></li></ul>
<a href="#">dbp:caption</a>	<ul style="list-style-type: none"><li>A cosmopolitan garnished with a lemon twist</li></ul>
<a href="#">dbp:drinkware</a>	<ul style="list-style-type: none"><li>cocktail</li></ul>
<a href="#">dbp:garnish</a>	<ul style="list-style-type: none"><li>Lemon slice, lime wedge</li></ul>
<a href="#">dbp:iba</a>	<ul style="list-style-type: none"><li>yes</li></ul>
<a href="#">dbp:ingredients</a>	<ul style="list-style-type: none"><li>* 4 cl Vodka Citron * 1.5 cl Cointreau * 1.5 cl Fresh lime juice * 3 cl Cranberry juice</li></ul>





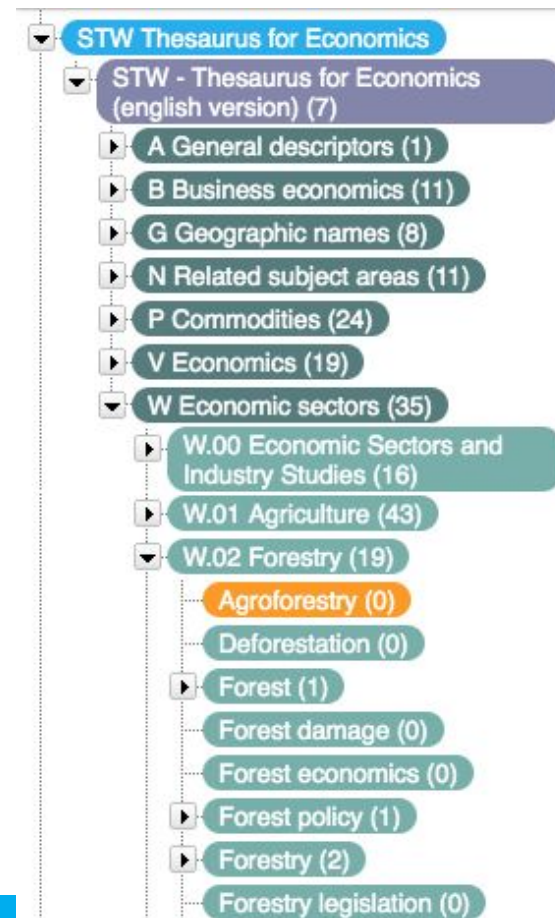
# Linking Concepts to DBpedia - Batchlinking

- ▶ UnifiedViews pipeline for batchlinking the thesaurus to DBpedia
  - ▶ Solr to provide efficient search for potential candidates
  - ▶ Spreading activation to take into account context of the concept



# Linking Concepts - Batchlinking Evaluation

- ▶ STW Thesaurus for Economics
  - ▷ <http://zbw.eu/stw/version/latest/about>
  - ▷ ~3500 links discovered
  - ▷ Compared against gold standard provided by STW (~2600 links) and manually extended (~900 links)
- ▶ Results:
  - ▷ Precision **96%**
    - The suggested link was correct
  - ▷ Recall almost 40%

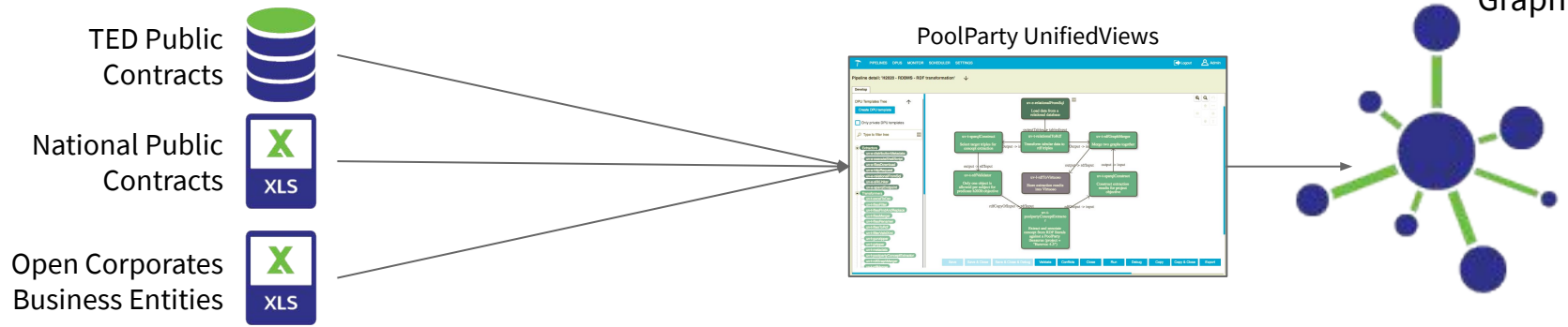


# KB Enrichment Scenarios

Structured Data

# KG Enrichment Scenarios for Structured Data

- ▶ Data collection/pre-processing
- ▶ **Schema mapping, Entity linking, Data fusion**
- ▶ Quality assessment
- ▶ Loading data to the KG



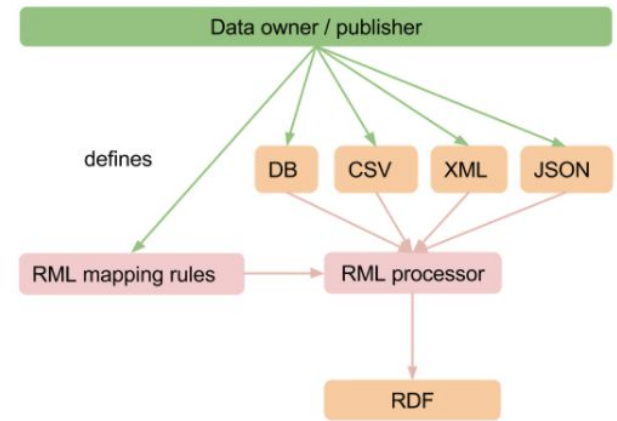
# Schema Mapping - Idea

- ▶ Map schemas of the input data sources to the target schemas maintained by PoolParty
  - ▶ Support for arbitrary sources (CSV, JSON, XML, relational data sources)
  - ▶ Suggestion of mapping rules
  - ▶ Support for data transformations

Contract Number	Vendor	Vendor City	Vendor State	Vendor Zip Code
1069374	Tom Woll and Associates	Fairfax	VA	22030
1053701	Setty & Associates International PLLC	Washington	DC	20016
1065909	Keller Brothers Inc	Mount Airy	MD	21771
9508000263AA	Ariana's Transportation Services, LLC	Washington	DC	20022
1069194	Community Ministries of Rockville, Inc.	Rockville	MD	20851
1020522	Family Services, Inc.	Landover Hills	MD	20784
1080123	Pallavi Belur	Potomac	MD	20854

# Schema Mapping - UnifiedViews

- ▶ Schema mapping DPUs
  - ▷ R2RML (W3C Rec)
  - ▷ RML - generalization of R2RML
    - <http://rml.io>

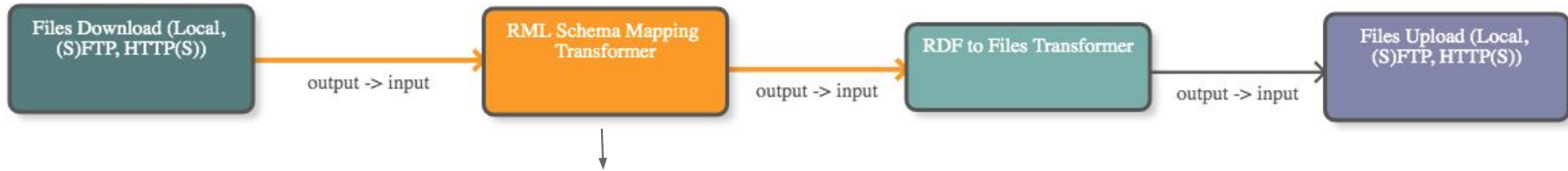


## Step 2: Define mappings for subject columns

Subject CSV column	Object CSV column	Property
Contract Number	Vendor	* Supplier
Vendor	Vendor Address Line 2	* Address
Vendor Address Line 2	Vendor Address Line 1	* Street
Vendor Address Line 2	Vendor Zip Code	* Zip code
Vendor Address Line 2	Vendor City	* City
Vendor Address Line 2	Vendor State	* State

```
@prefix rml: <http://semweb.mmlab.be/ns/rml#>.
@prefix ql: <http://semweb.mmlab.be/ns/ql#>.
@prefix rr: <http://www.w3.org/ns/r2rml#>.
@base <http://unifiedviews.eu/rml/csv/base>.
<#Contract+Number>
rml:logicalSource [
  rml:source "file.csv" ;
  rml:referenceFormulation ql:CSV
];
rr:subjectMap [
  rr:template "http://demo.semantic-web.com/data/contract/{Contract Number}";
  rr:class <http://purl.org/procurement/public-contracts#Contract>
];
rr:predicateObjectMap [
  rr:predicate <http://purl.org/procurement/public-contracts#supplier>;
  rr:objectMap [
    rr:parentTriplesMap <#Vendor>
  ]
].
```

# Schema Mapping - UnifiedViews Pipeline



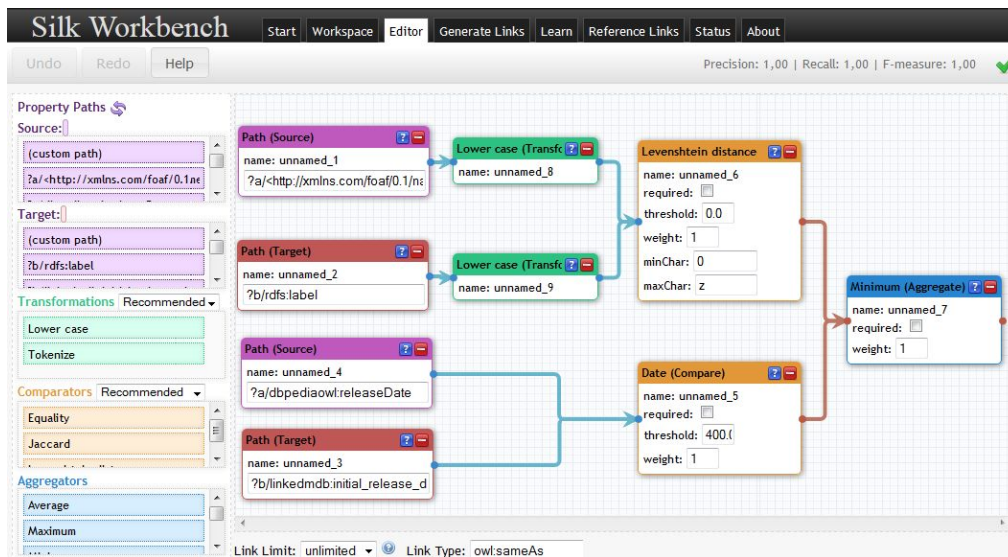
## Step 2: Define mappings for subject columns

Subject CSV column	Object CSV column	Property
Contract Number	Vendor	Supplier
Vendor	Vendor Address Line 2	Address
Vendor Address Line 2	Vendor Address Line 1	Street
Vendor Address Line 2	Vendor Zip Code	Zip code
Vendor Address Line 2	Vendor City	City
Vendor Address Line 2	Vendor State	State

```
@prefix rml: <http://semweb.mmlab.be/ns/rml#>.
@prefix ql: <http://semweb.mmlab.be/ns/ql#>.
@prefix rr: <http://www.w3.org/ns/r2rml#>.
@base <http://unifiedviews.eu/rml/csv/base>.
<#Contract+Number>
rml:logicalSource [
  rml:source "file.csv" ;
  rml:referenceFormulation ql:CSV
];
rr:subjectMap [
  rr:template "http://demo.semantic-web.com/data/contract/{Contract Number}";
  rr:class <http://purl.org/procurement/public-contracts#Contract>
];
rr:predicateObjectMap [
  rr:predicate <http://purl.org/procurement/public-contracts#supplier>;
  rr:objectMap [
    rr:parentTriplesMap <#Vendor>
  ]
].
```

# Entity Linking - Idea

- ▶ Link attributes to the entities in the existing KG
- ▶ Discover duplicates between the acquired data and the KG



```
<Interlinks>
<Interlink id="cities">
  <LinkType>owl:sameAs</LinkType>

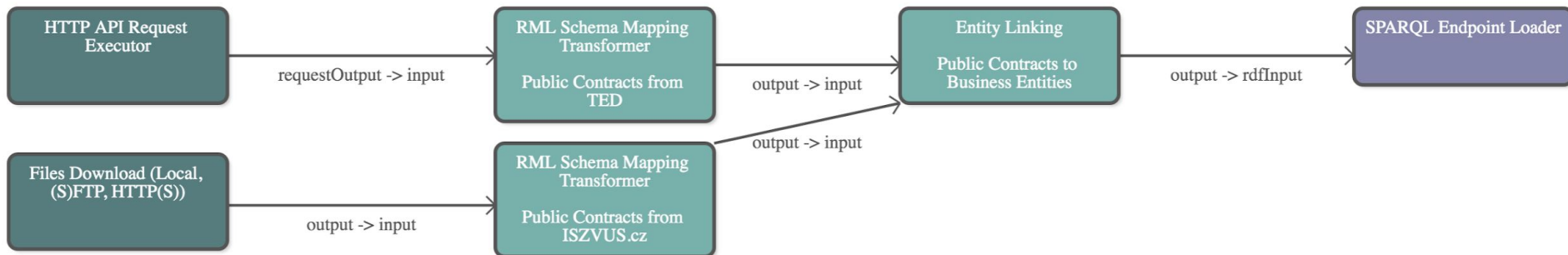
  <SourceDataset dataSource="dbpedia" var="a">
    <RestrictTo>
      ?a rdf:type dbpedia:City
    </RestrictTo>
  </SourceDataset>

  <TargetDataset dataSource="geonames" var="b">
    <RestrictTo>
      ?b rdf:type gn:P
    </RestrictTo>
  </TargetDataset>

  <LinkageRule>
    <Aggregate type="average">
      <Compare metric="levenshteinDistance" threshold="1">
        <Input path="?a/rdfs:label" />
        <Input path="?b/gn:name" />
      </Compare>
      <Compare metric="num" threshold="1000">
        <Input path="?a/dbpedia:populationTotal" />
        <Input path="?b/gn:population" />
      </Compare>
    </Aggregate>
  </LinkageRule>
</Interlinks>
```



- ▶ UnifiedViews DPUs
  - ▶ DPUs executing SPARQL (Update) queries
  - ▶ DPUs using Silk, <http://silkframework.org/>



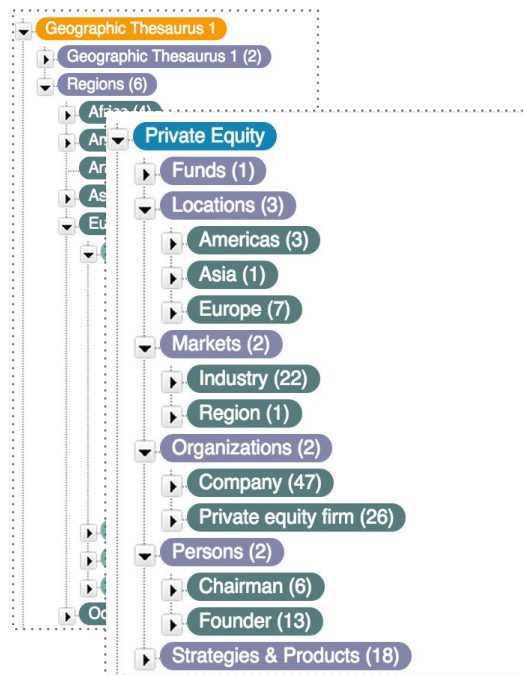
- ▶ Fuse different representations of the same resources
  - ▷ Define what happens when there are conflicting attributes/relations
  - ▷ Apply a conflict resolution policy

Company Name	Founded	Number of employees	Address
Food Council, Inc.	12/20/2004	30	...
Montgomery county Food Council, Inc.	12/21/2004	35	...



## Knowledge Graph

- ▶ Unstructured data with:
  - Recognized (named) entities
  - Recognized attributes and relations of (named) entities
- ▶ Follow the recipe for structured data
  - Schema mapping
  - Entity Linking
  - Data Fusion



## Entity Extraction & Text Mining

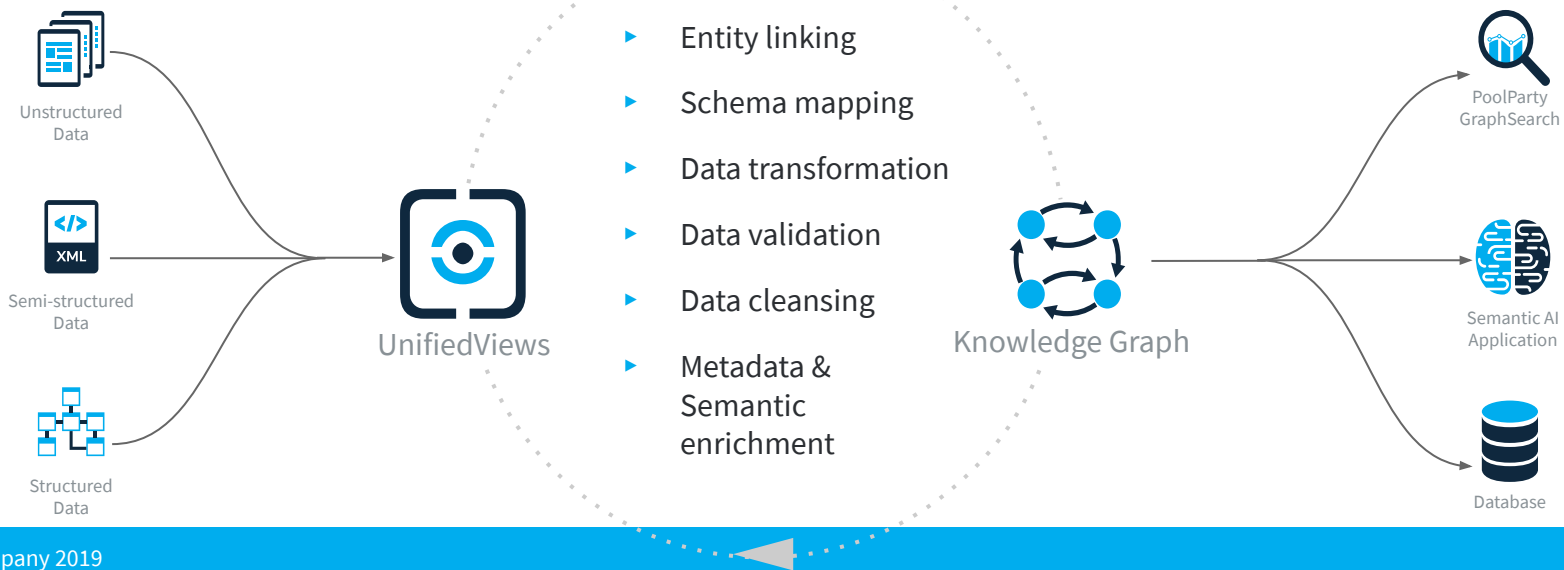
Bain Capital is a venture capital company based in Boston, MA. Since inception it has invested in hundreds of companies including AMC Entertainment, Brookstone, and Burger King. The company was co-founded by Mitt Romney.

# Summary

# Knowledge Graphs

- ▶ “PoolParty brings the power of Google into the enterprise environment through Knowledge Graphs.”

Knowledge Architect,  
Top 3 Oil & Gas Company





# Thank You!



Tomas Knap, PhD

Software Engineer and Researcher,  
Semantic Web Company

[tomas.knap@semantic-web.com](mailto:tomas.knap@semantic-web.com)

