Semantifying Content Management Systems

Axel-Cyrille Ngonga Ngomo[1,2], Norman Heino[1], Klaus Lyko[1],
René Speck[1], Martin Kaltenböck[3]

[1]University of Leipzig, [2]University of Mainz,[3]Semantic Web Company

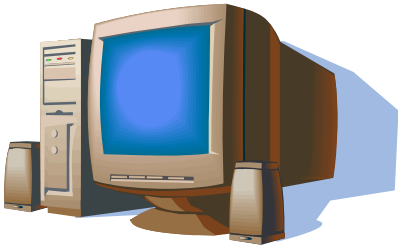
ISWC2011, Bonn, Germany
October 27th, 2011

Take-Home Message

- **Two-fold tasks for CMS**
  1. Make content easily processable for humans
  2. Provide content in machine-readable format

- Present **SCMS** framework
  - **Generic solution** for adding semantic capabilities to CMS
  - **High flexibility** ensured by communication via RDF
  - **High accuracy** ensured by the FOX framework

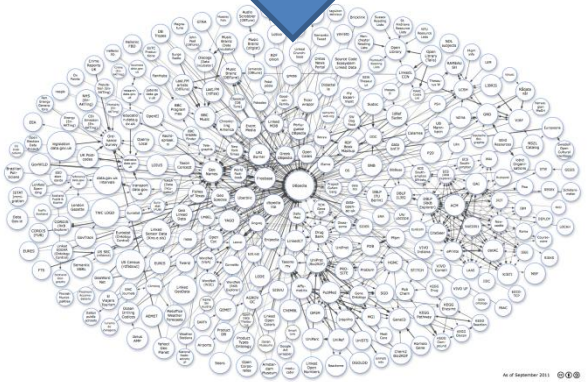
Introduction



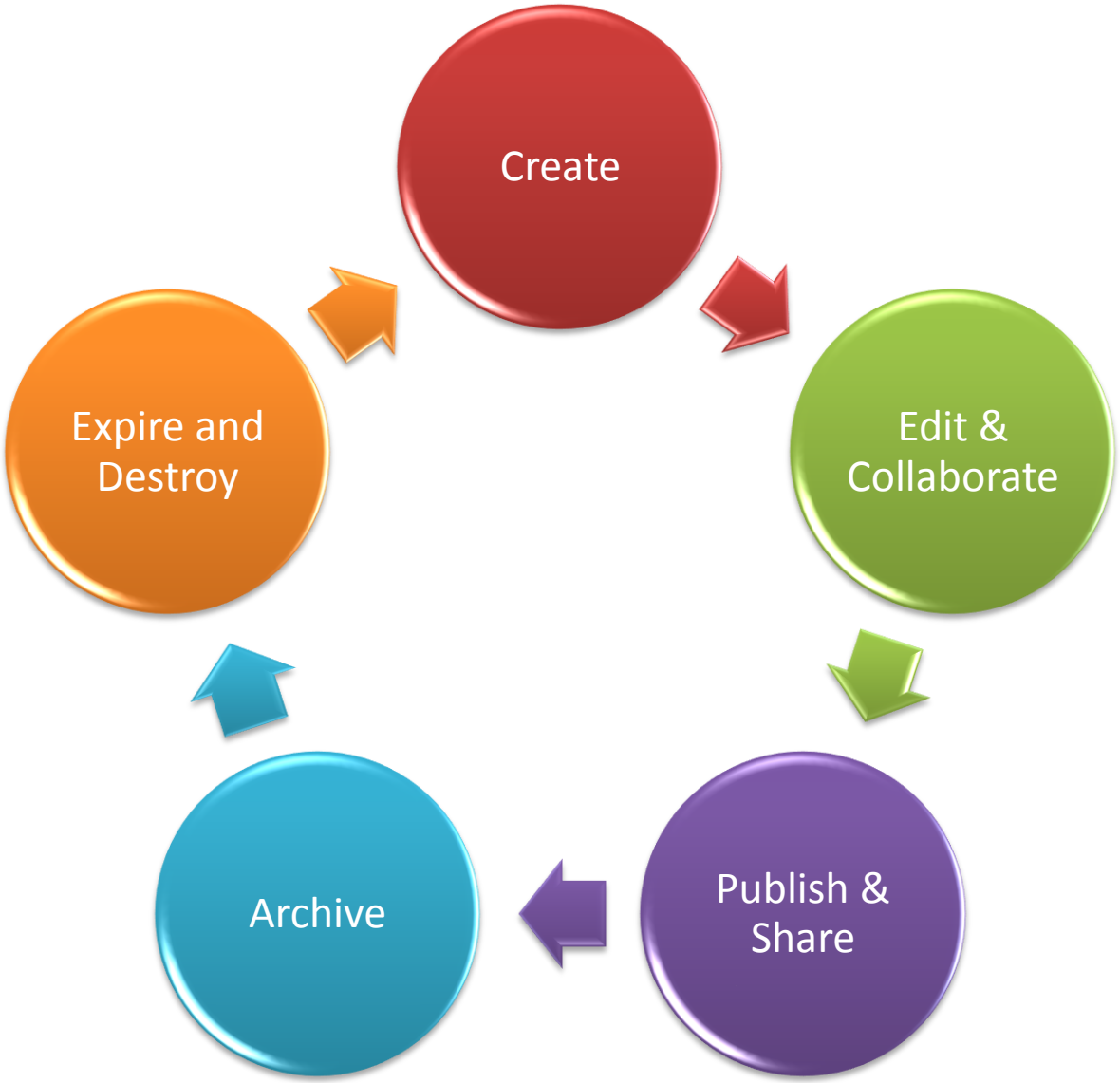
CMS

Human Web

SCMS



Introduction

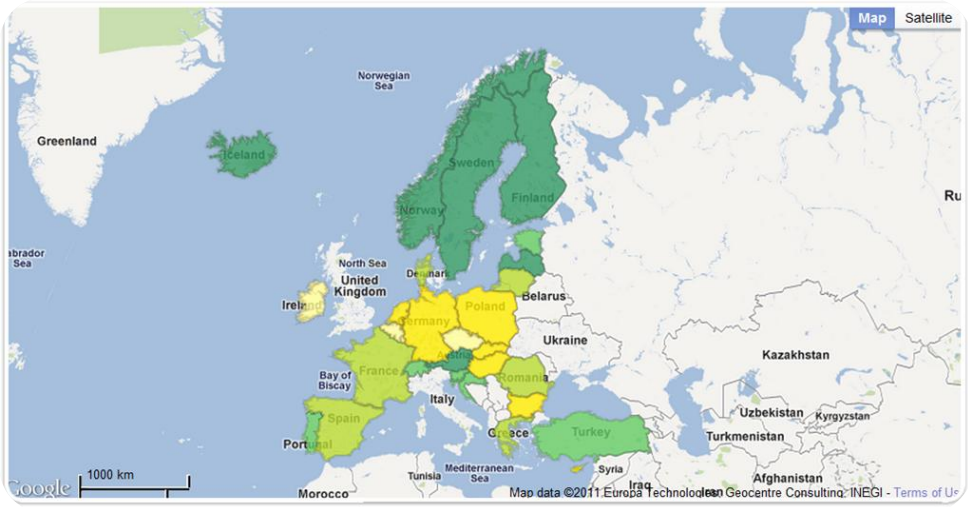


Requirements

- **Flexibility**
  - Several hundreds CMS
  - Written in different languages
  - Easy integration
- **Accuracy**
  - Low false positive rate
  - Low false negative rate
  - Knowledge curation

Use Cases

- Renewable Energy
  - http://www.reeegle.info
  - Description of actors
  - Description of countries

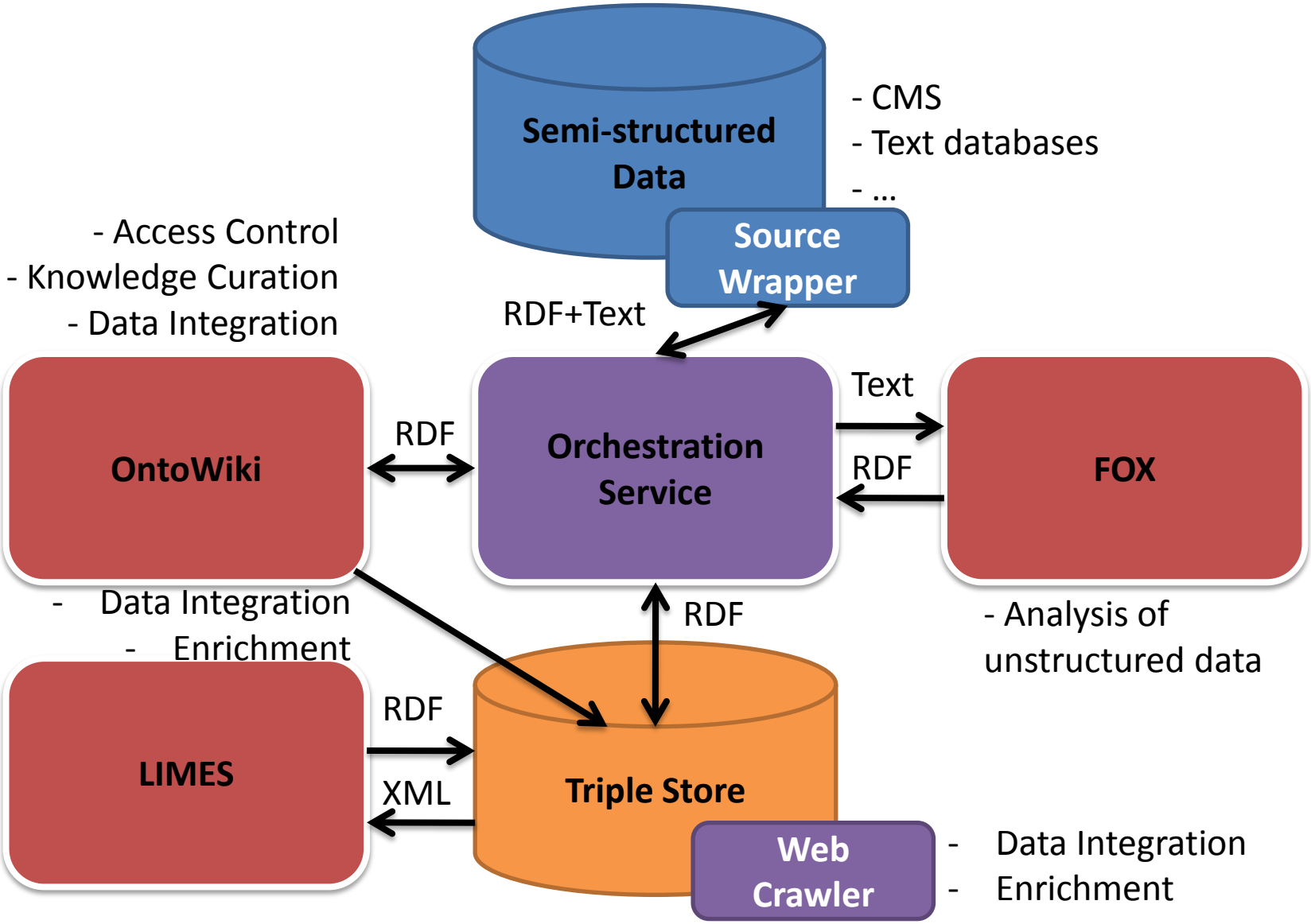


- Need for
  - Keyword Extraction (Faceted Search)
  - Entity Extraction + Data Enrichment (Dossier Mashups)
  - Relation Extraction (Business Questions)

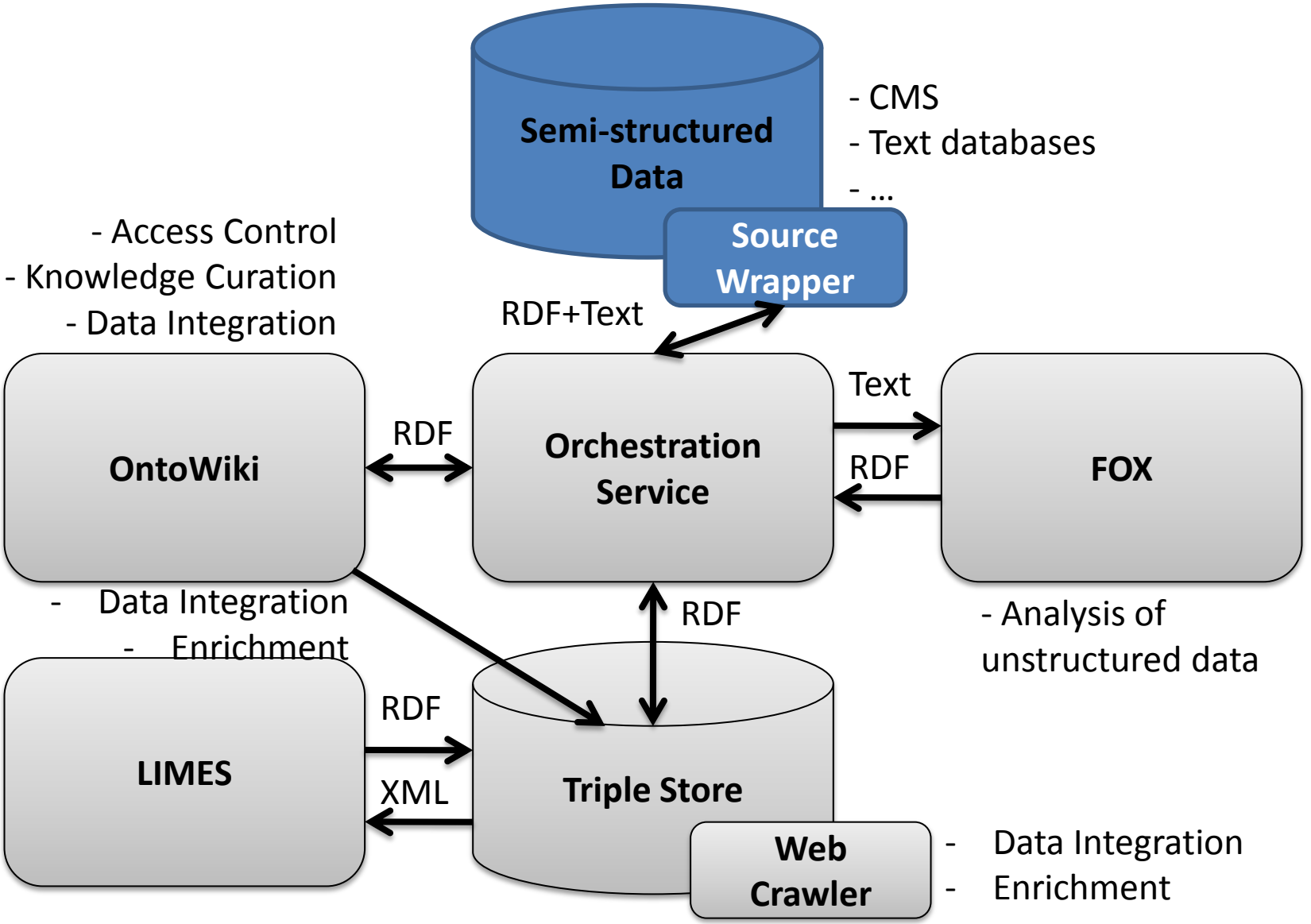
Use Cases

- Tourism
  - Description of cruises
  - Description of touristic attractions
- Need for
  - Keyword extraction (Semantic Search, Automatic update of descriptions & multimedia from the Web)
  - NER (Search enhancement)

Architecture

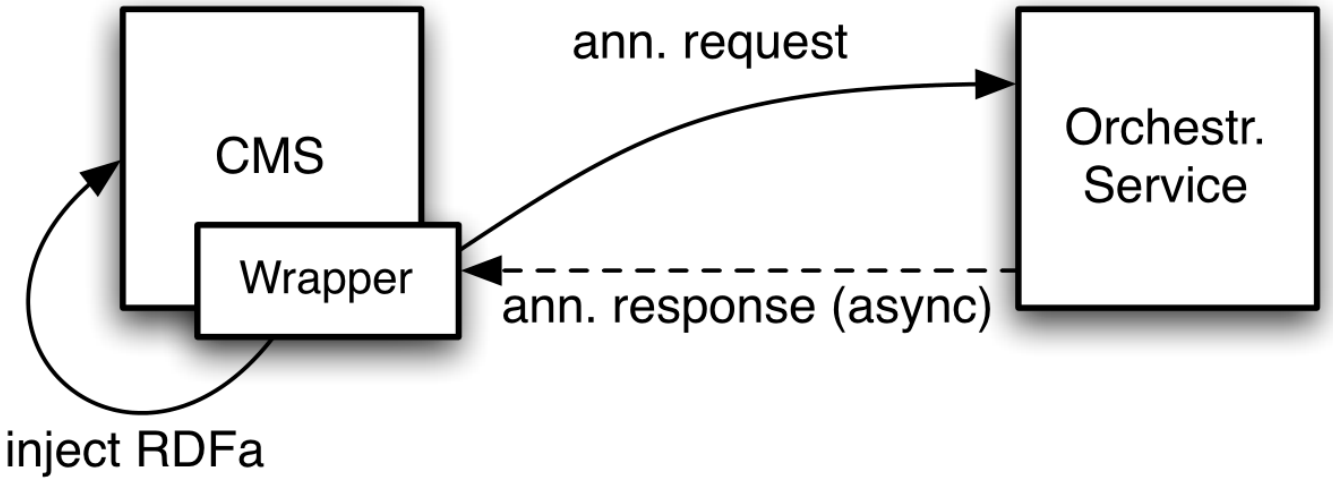


Architecture

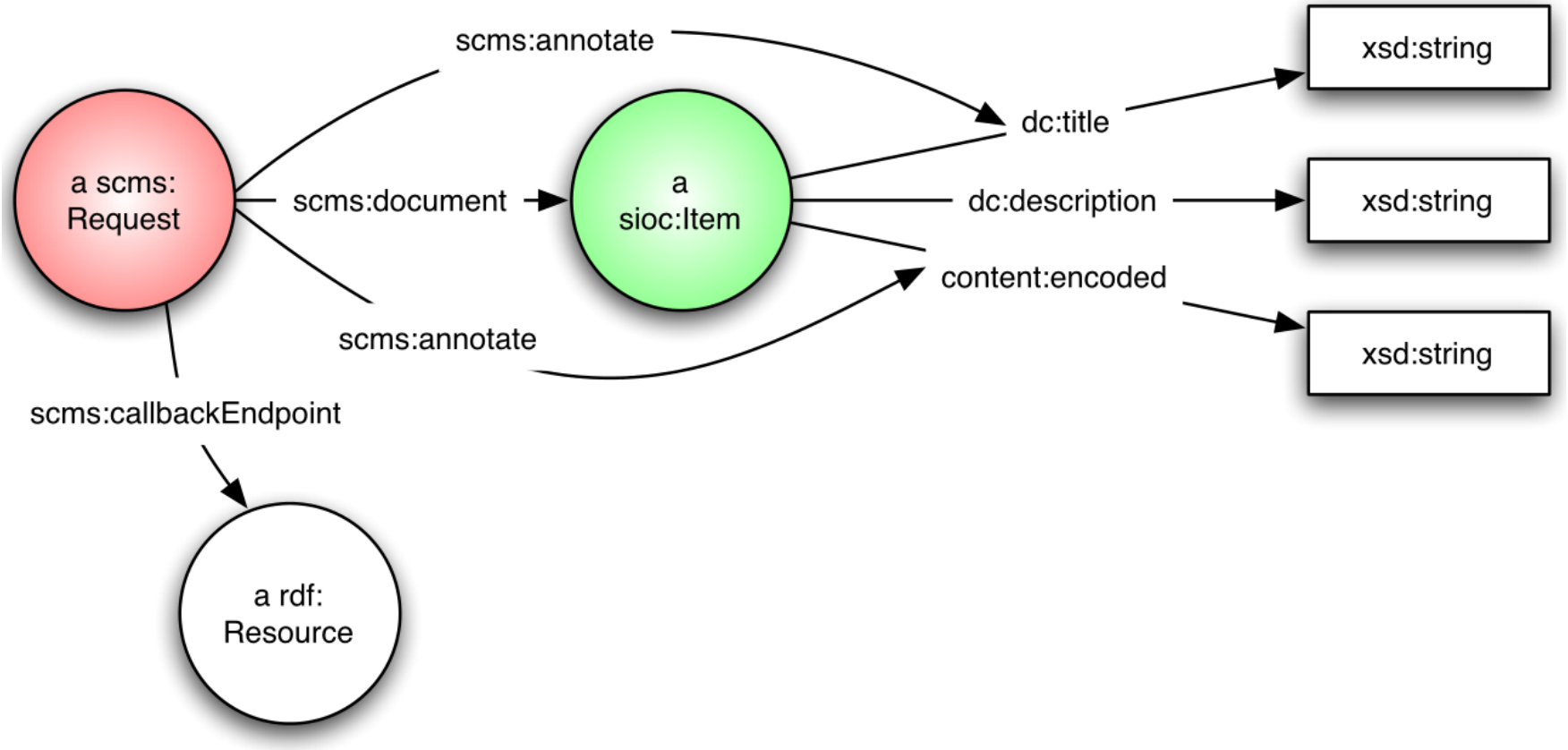


Wrapper

- Main functions
  - Send data to orchestration service
  - Receive data from orchestration service
  - Clean data and write in CMS



Request Vocabulary



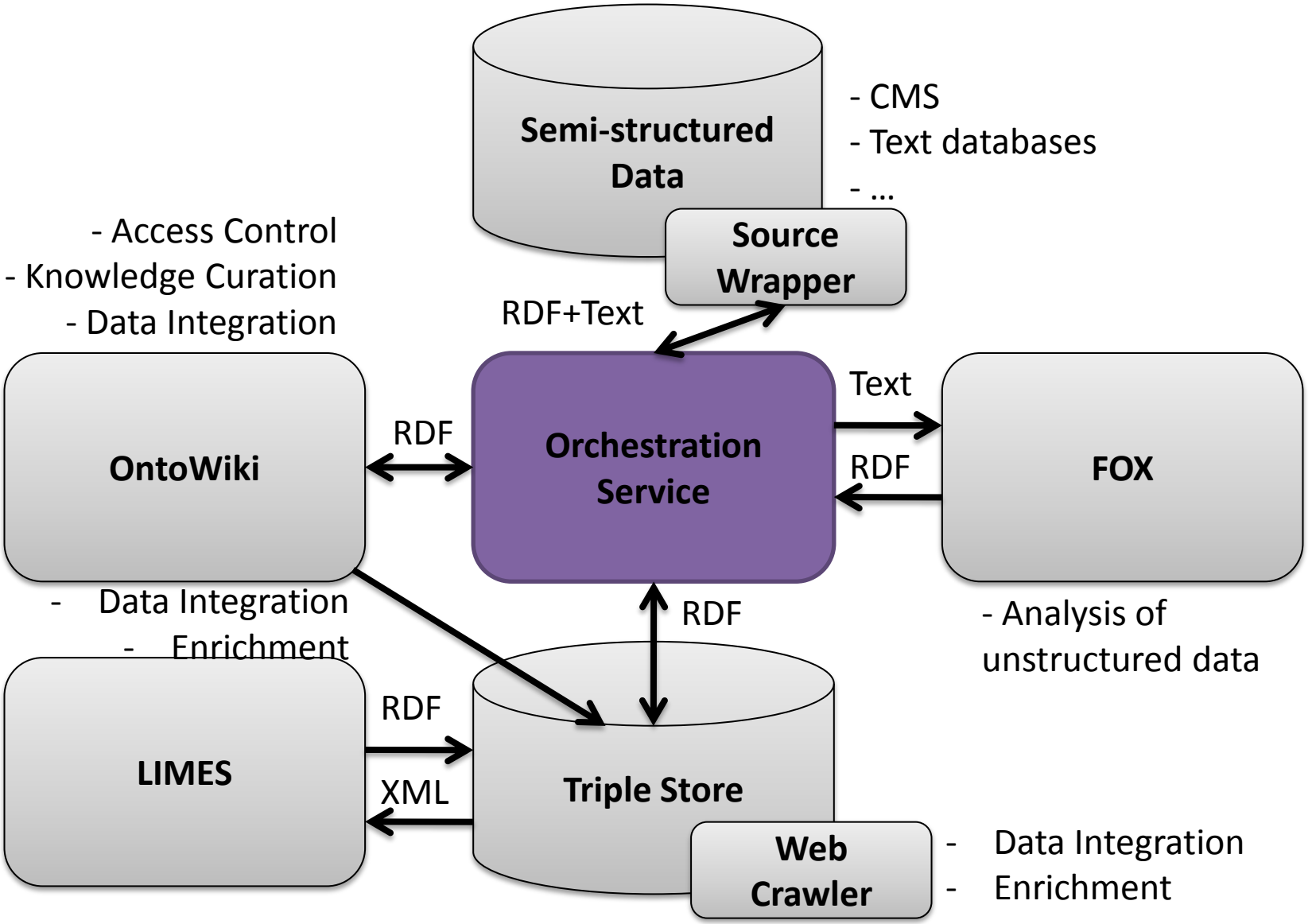
Request Example

```
@prefix content: <http://purl.org/rss/1.0/modules/content/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix sioc: <http://rdfs.org/sioc/ns#> .
@base <http://ns.aksw.org/scms/> .

<http://example.com/wrapperRequest/1> a <Request> ;
    <document> <http://example.com/drupal/node/10> ;
    <callbackEndpoint> <http://example.com/wrapper> ;
    <annotate> content:encoded .

<http://example.com/drupal/node/10> a sioc:Item ;
    dc:title "Prometeus" ;
    content:encoded "The company Prometeus is an energy provider located in
          Budapest." .
```

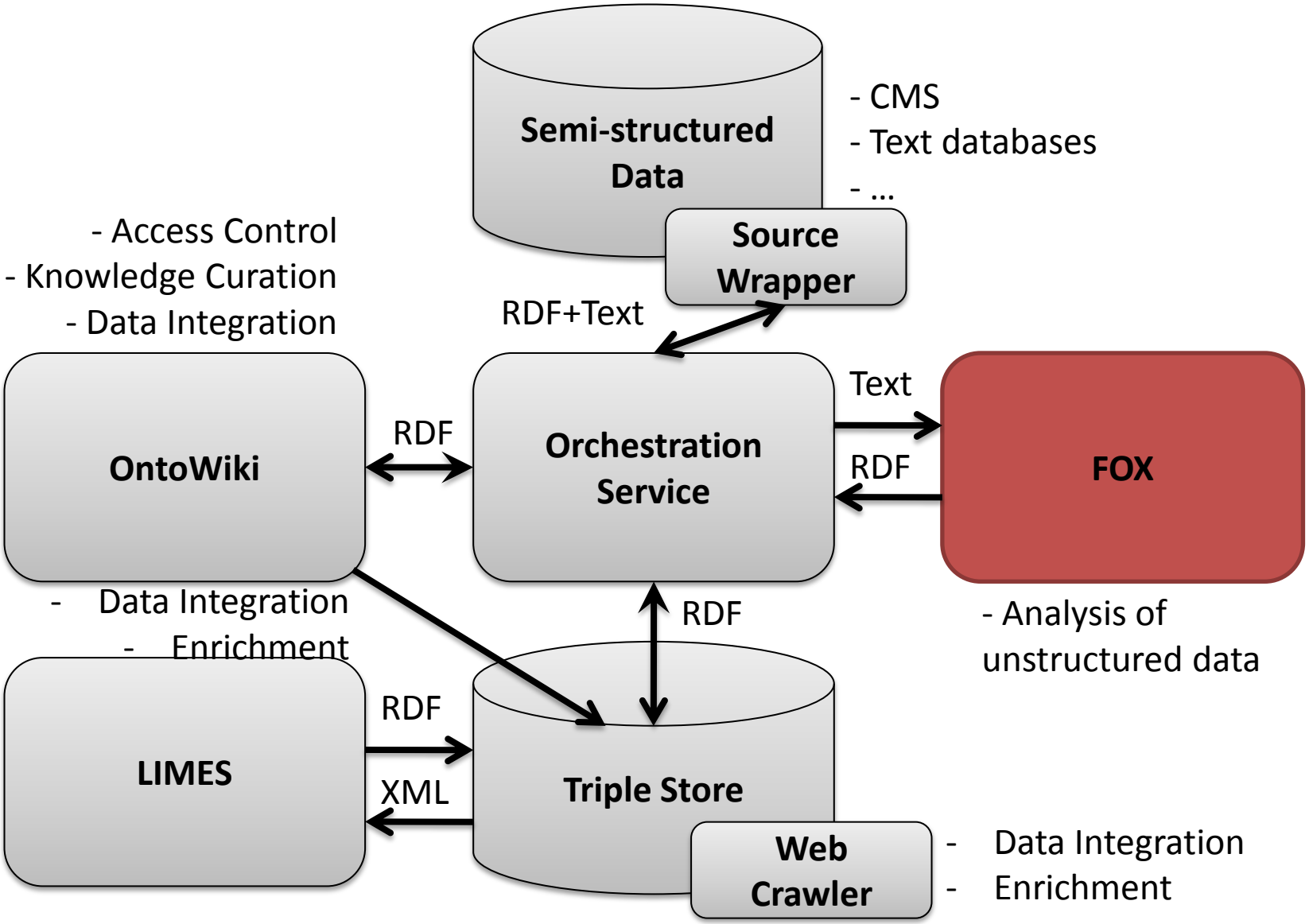
Architecture



Example

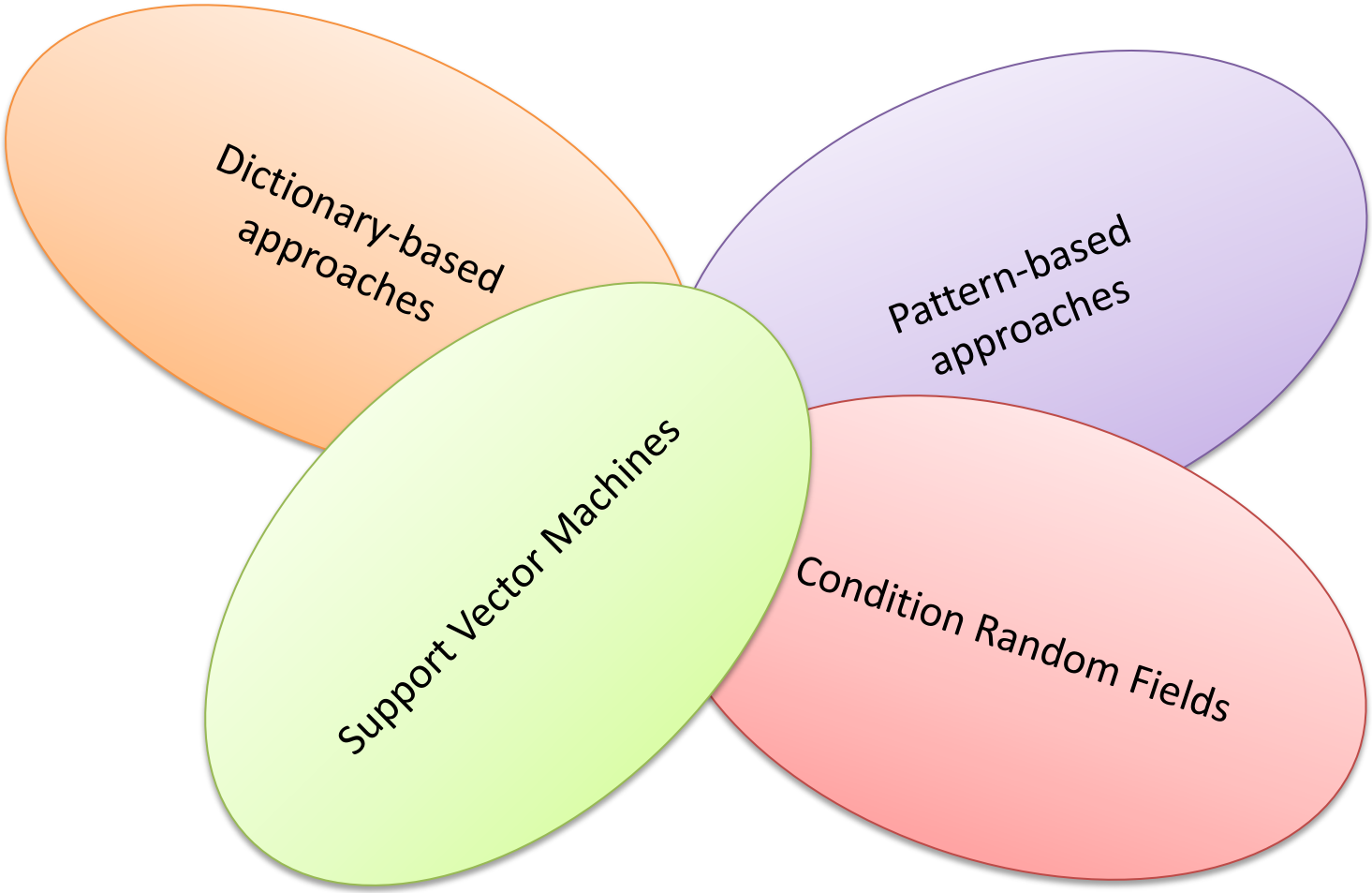
The company Prometeus is an energy provider located in the capital of Hungary, i.e., Budapest.

Architecture



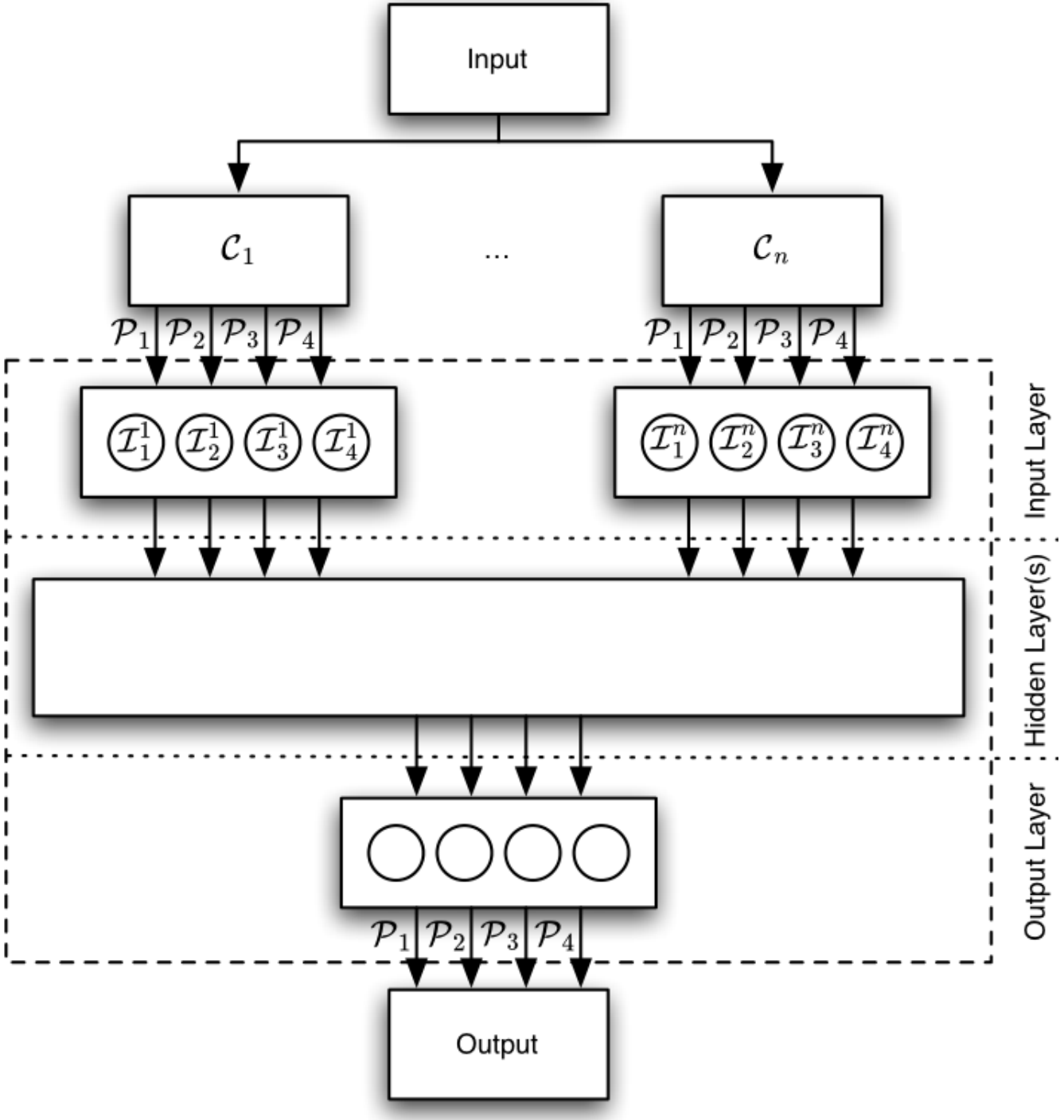
FOX

- Federated knOwledge eXtraction Framework

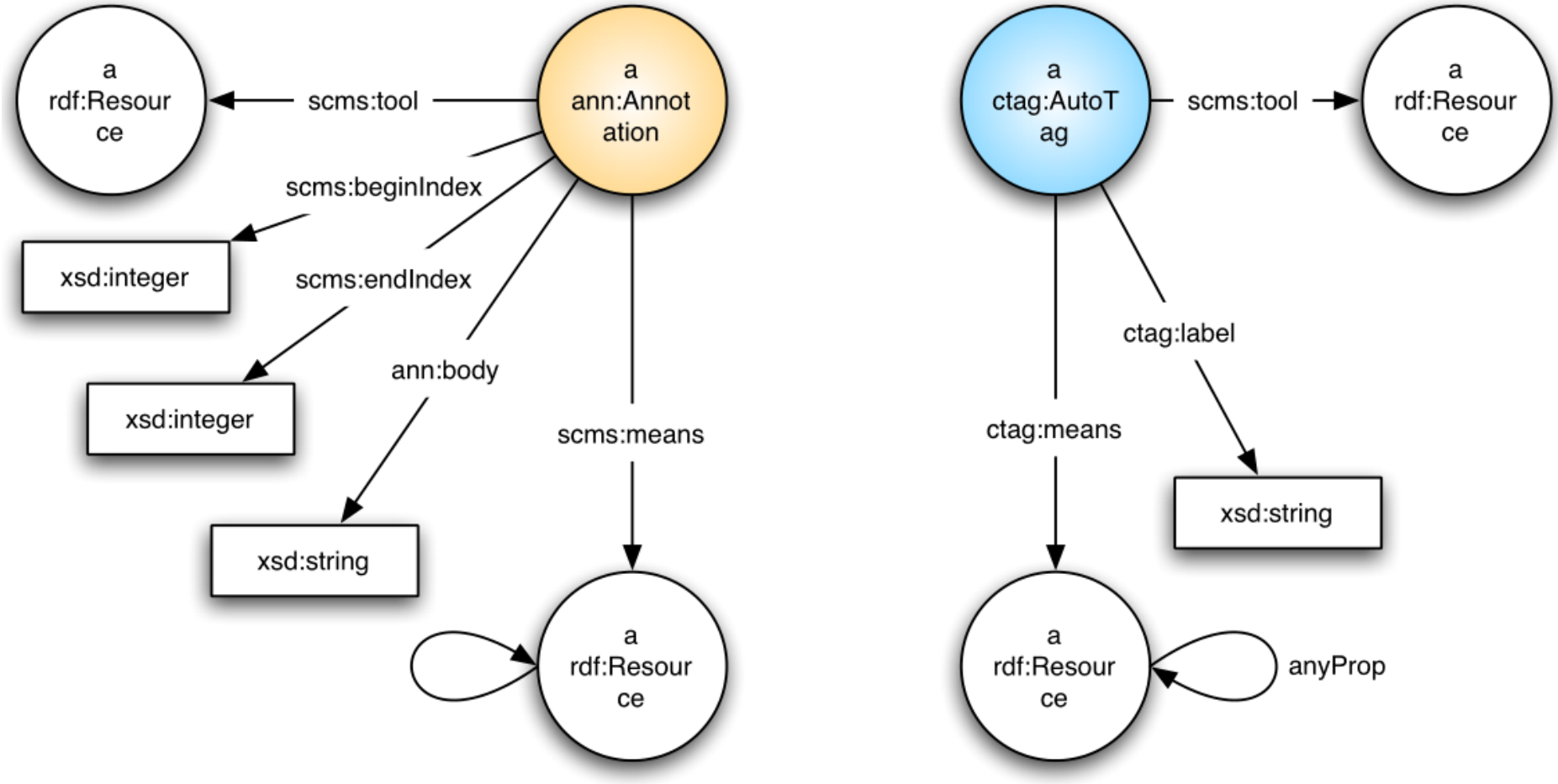


FOX

- Each solution has its strengths and weaknesses
- Apply **ensemble learning** to
  - Combine the tools at hand
  - Compute better results
- Two requirements
  - Accuracy
  - Diversity
- In our case, neural networks


Input
$\mathcal{C}_1$
...
$\mathcal{C}_n$
$\mathcal{P}_1$ $\mathcal{P}_2$ $\mathcal{P}_3$ $\mathcal{P}_4$
$\mathcal{P}_1$ $\mathcal{P}_2$ $\mathcal{P}_3$ $\mathcal{P}_4$
$\mathcal{I}_1^1$ $\mathcal{I}_2^1$ $\mathcal{I}_3^1$ $\mathcal{I}_4^1$
$\mathcal{I}_1^n$ $\mathcal{I}_2^n$ $\mathcal{I}_3^n$ $\mathcal{I}_4^n$
Input Layer
Hidden Layer(s)
Output Layer
$\mathcal{P}_1$ $\mathcal{P}_2$ $\mathcal{P}_3$ $\mathcal{P}_4$
Output

FOX Vocabularies



FOX

http://fox.aksw.org

FOX

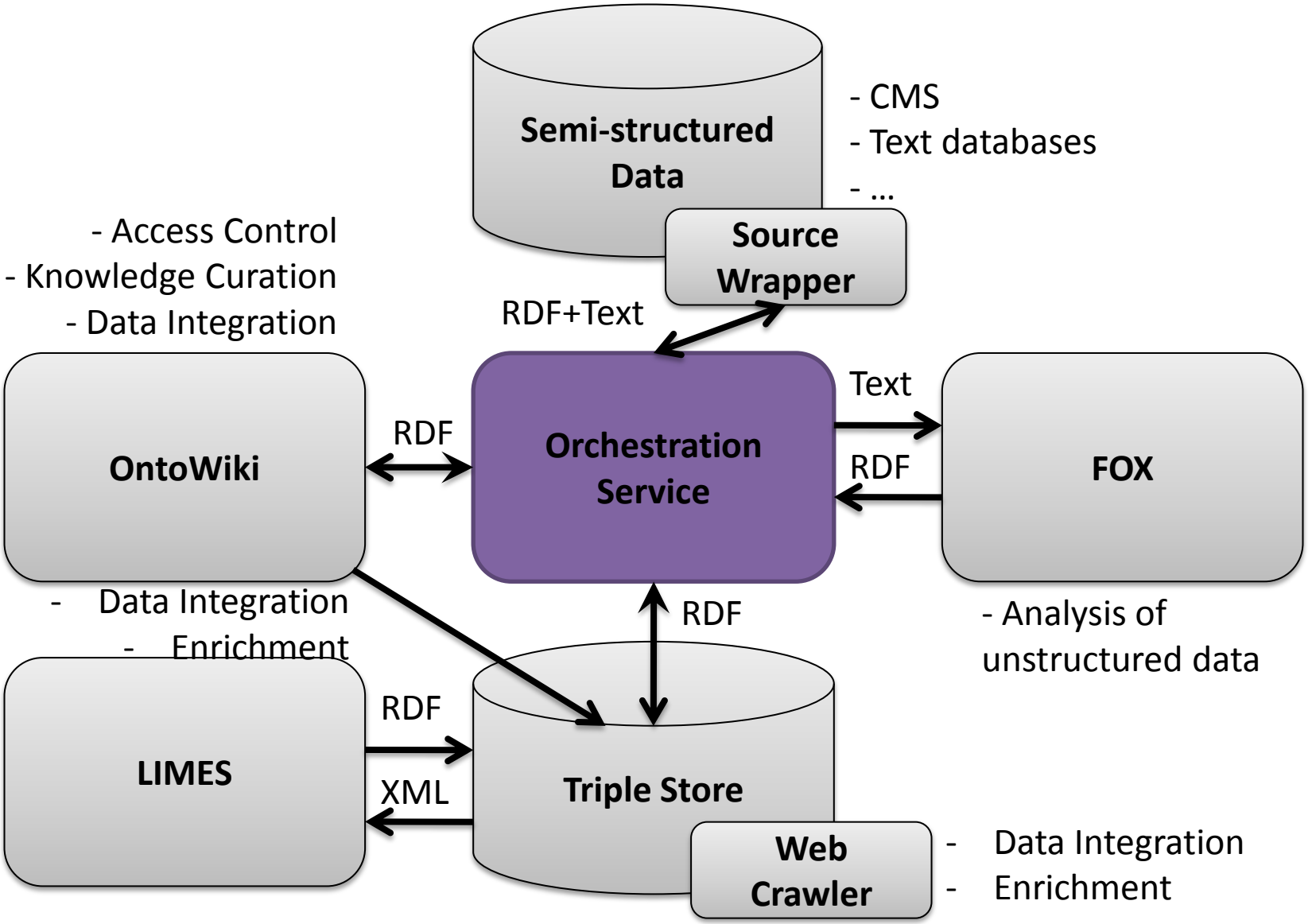
```
@prefix scmsann:  <http://ns.aksw.org/scms/annotations/> .
@prefix ctag:     <http://commontag.org/ns#> .
@prefix xsd:      <http://www.w3.org/2001/XMLSchema#> .
@prefix rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ann:      <http://www.w3.org/2000/10/annotation-ns#> .
@prefix scms:     <http://ns.aksw.org/scms/> .

[]      a         ann:Annotation , scmsann:LOCATION ;
        scms:beginIndex "70"^^xsd:int ;
        scms:endIndex "77"^^xsd:int ;
        scms:means <http://dbpedia.org/resource/Hungary> ;
        scms:source <http://ns.aksw.org/scms/tools/FOX> ;
        ann:body "Hungary"^^xsd:string .

[]      a         ann:Annotation , scmsann:ORGANIZATION ;
        scms:beginIndex "12"^^xsd:int ;
        scms:endIndex "21"^^xsd:int ;
        scms:means <http://scms.eu/Prometeus> ;
        scms:source <http://ns.aksw.org/scms/tools/FOX> ;
        ann:body "Prometeus"^^xsd:string .

[]      a         ann:Annotation , scmsann:LOCATION ;
        scms:beginIndex "85"^^xsd:int ;
        scms:endIndex "93"^^xsd:int ;
        scms:means <http://dbpedia.org/resource/Budapest> ;
        scms:source <http://ns.aksw.org/scms/tools/FOX> ;
        ann:body "Budapest"^^xsd:string .
```

Architecture



Orchestration Service

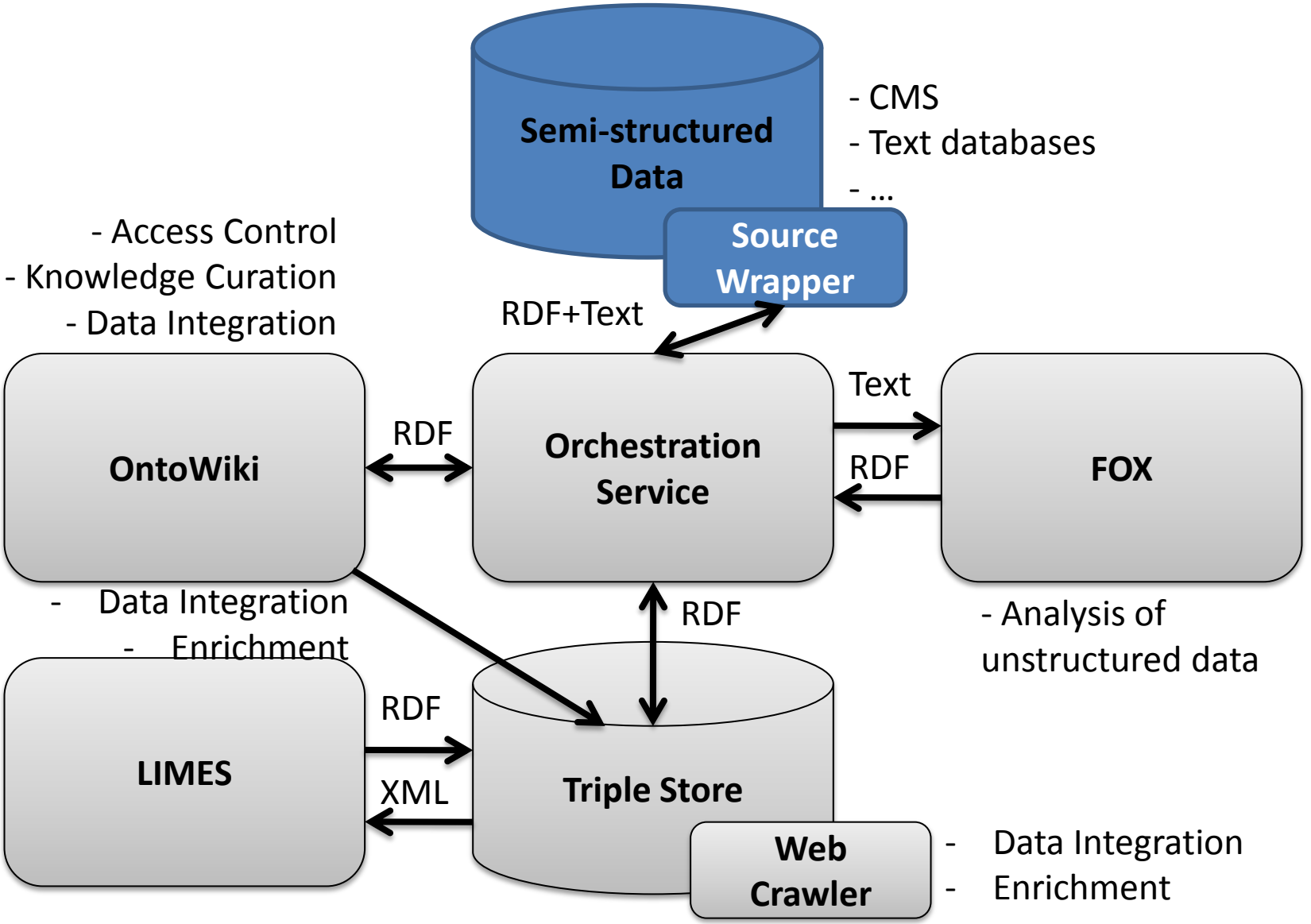
```
@prefix scmsann:  <http://ns.aksw.org/scms/annotations/> .
@prefix ctag:     <http://commontag.org/ns#> .
@prefix xsd:      <http://www.w3.org/2001/XMLSchema#> .
@prefix rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ann:      <http://www.w3.org/2000/10/annotation-ns#> .
@prefix scms:     <http://ns.aksw.org/scms/> .

[]      a         ann:Annotation , scmsann:LOCATION ;
        scms:annotates <http://example.com/drupal/node/10> ;
        scms:property <http://purl.org/rss/1.0/modules/content/encoded> ;
        scms:beginIndex "70"^^xsd:int ;
        scms:endIndex "77"^^xsd:int ;
        scms:means <http://dbpedia.org/resource/Hungary> ;
        scms:source <http://ns.aksw.org/scms/tools/FOX> ;
        ann:body "Hungary"^^xsd:string .

[]      a         ann:Annotation , scmsann:ORGANIZATION ;
        scms:annotates <http://example.com/drupal/node/10> ;
        scms:property <http://purl.org/rss/1.0/modules/content/encoded> ;
        scms:beginIndex "12"^^xsd:int ;
        scms:endIndex "21"^^xsd:int ;
        scms:means <http://scms.eu/Prometeus> ;
        scms:source <http://ns.aksw.org/scms/tools/FOX> ;
        ann:body "Prometeus"^^xsd:string .

[]      a         ann:Annotation , scmsann:LOCATION ;
        scms:annotates <http://example.com/drupal/node/10> ;
        scms:property <http://purl.org/rss/1.0/modules/content/encoded> ;
        scms:beginIndex "85"^^xsd:int ;
        scms:endIndex "93"^^xsd:int ;
        scms:means <http://dbpedia.org/resource/Budapest> ;
        scms:source <http://ns.aksw.org/scms/tools/FOX> ;
        ann:body "Budapest"^^xsd:string .
```

Architecture



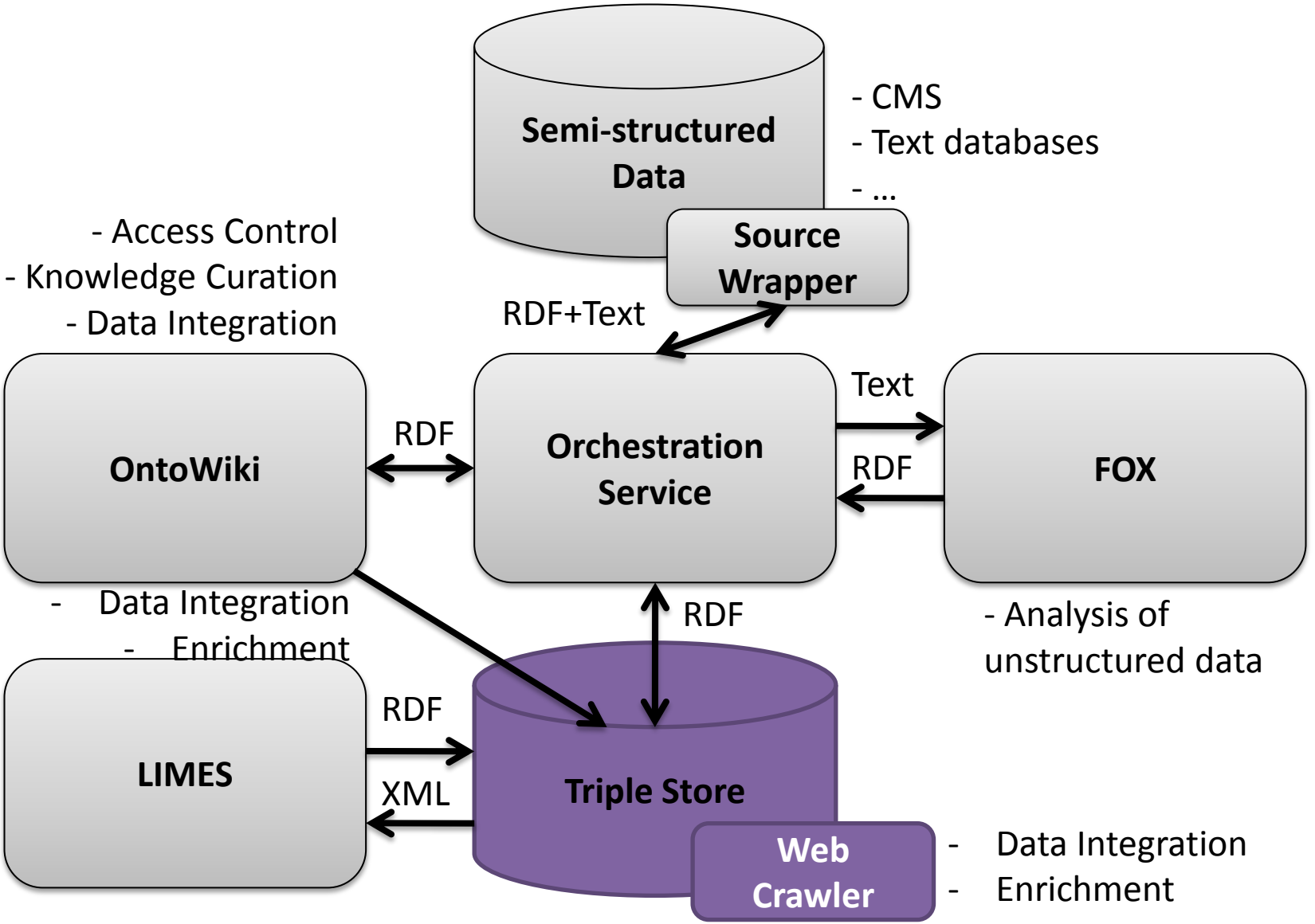
Integration in CMS

- Wrapper can post-process stream, e.g.
  - Only accept curated data
  - Only accept data from certain sources (tools, persons, …)
  - Change URIs
  - …
- Integration in most case as RDFa
  - Completes existing RDF in CMS
  - Finally, structured information

AKSW
SIMBA

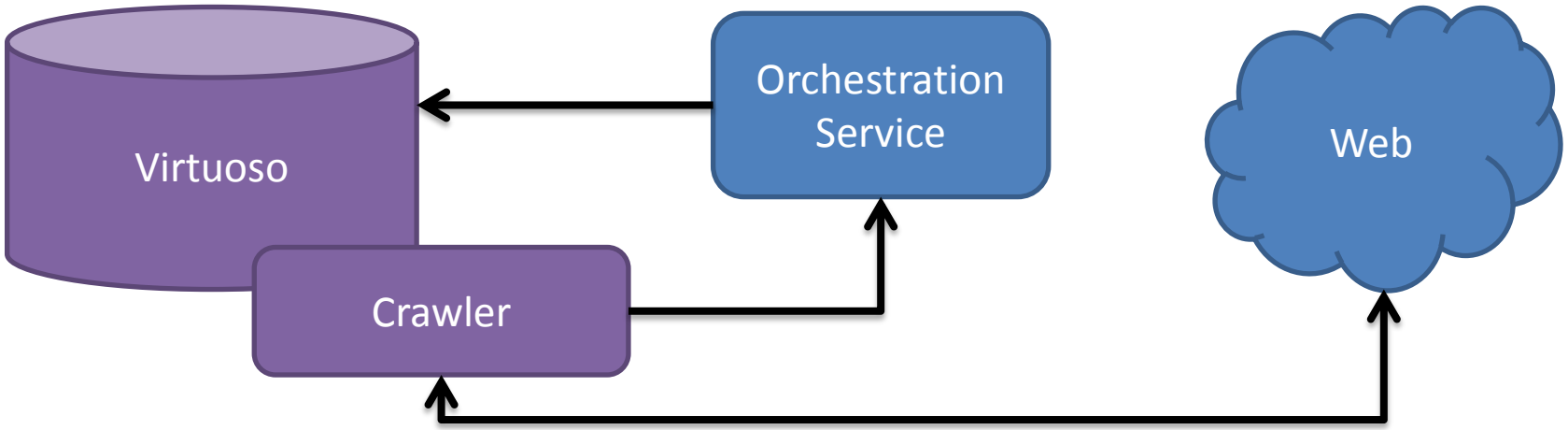
That‘s not the end of the story!

Data Storage

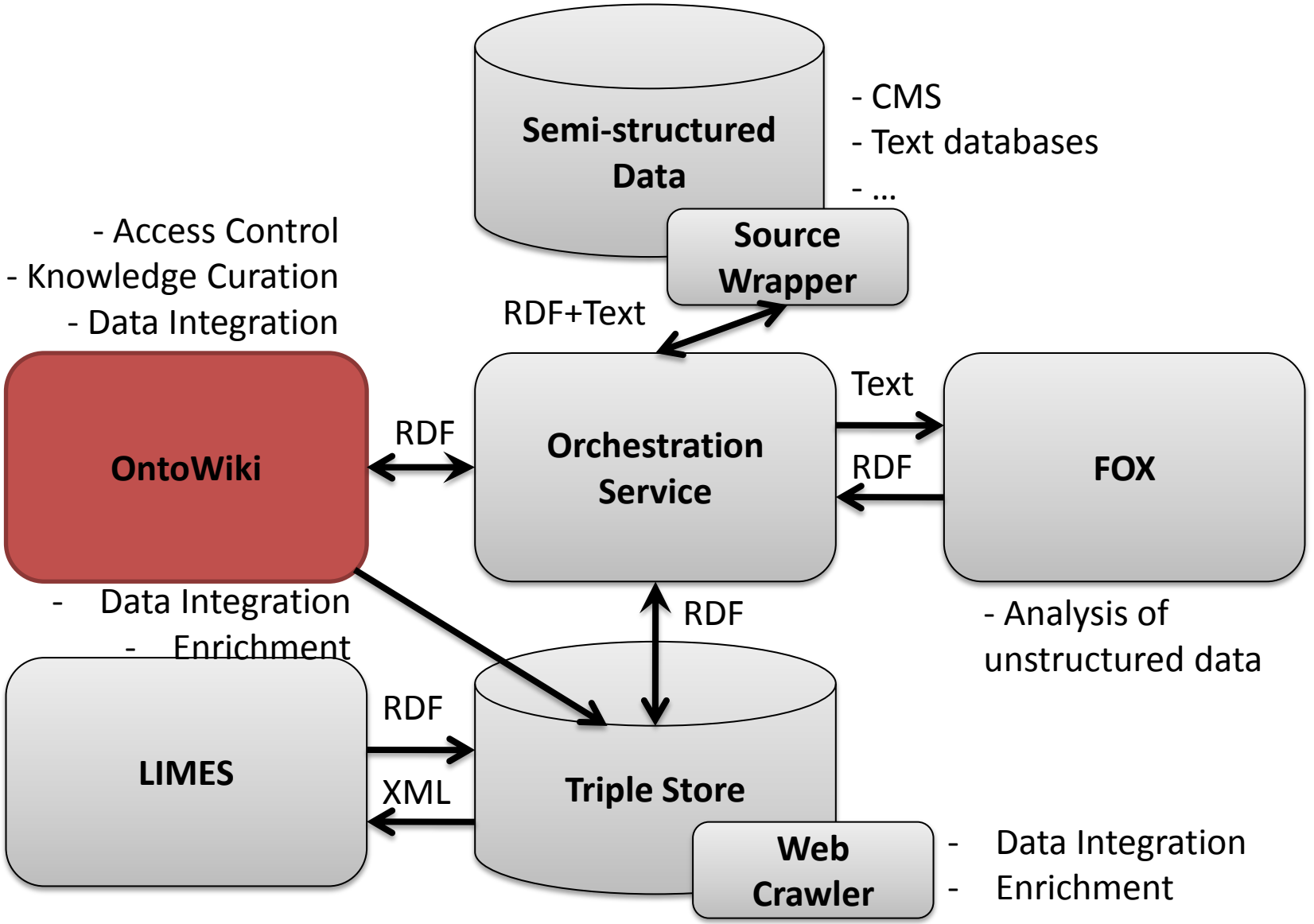


Virtuoso Extension

- Storage for orchestration & OntoWiki
- Provides SPARQL endpoint for LIMES
- **Extension** via Webcrawler
  - Retrieves Web documents based on FOX results or pre-defines sources
  - Implements a wrapper without callback

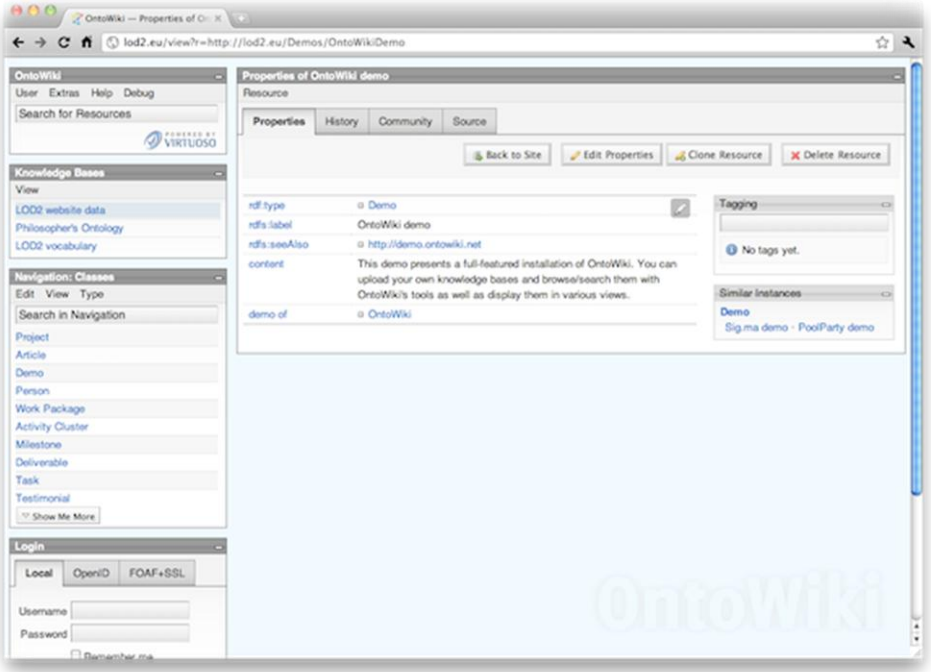


Architecture

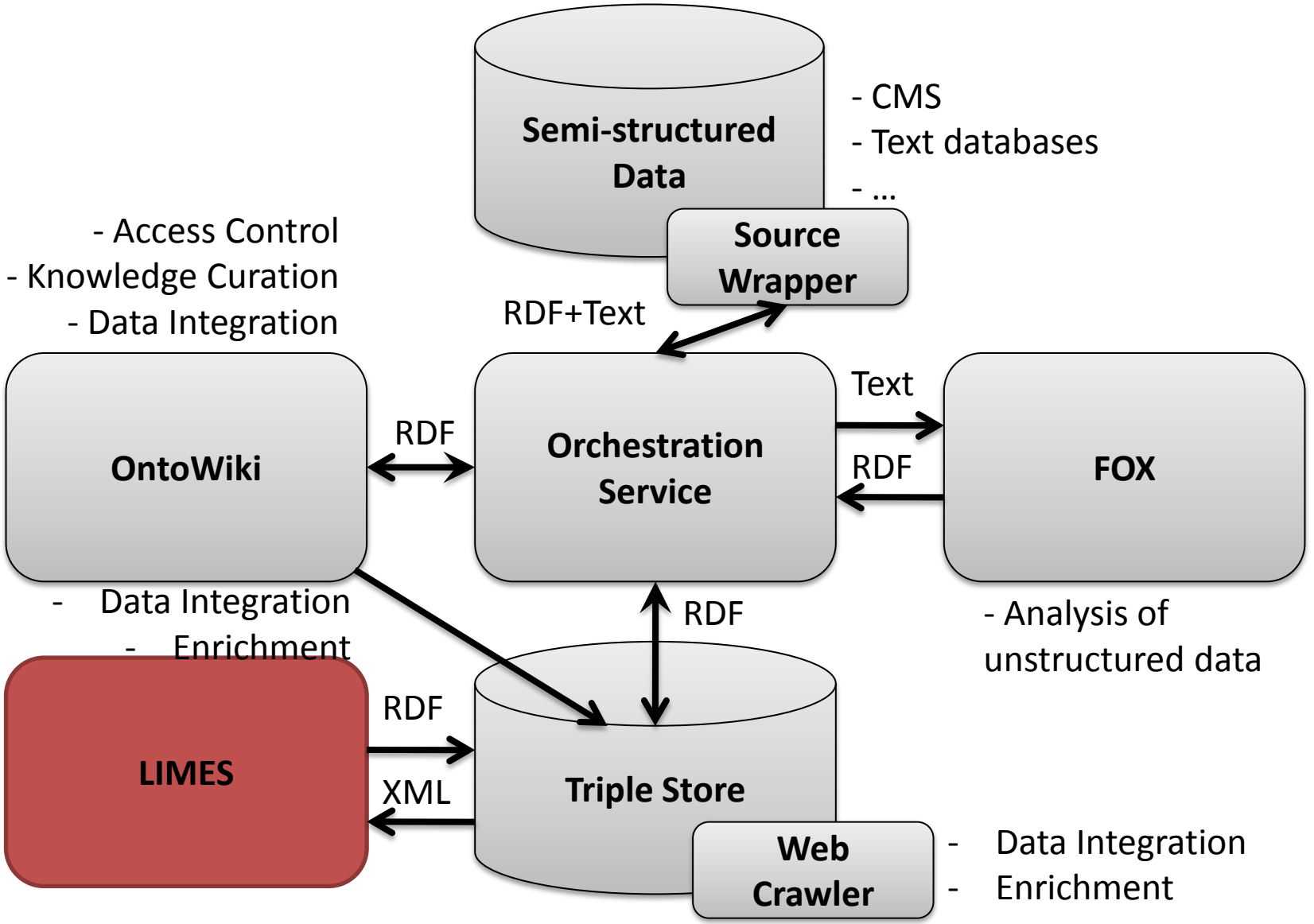


OntoWiki

- **Knowledge curation**
  - Solve access control for manual curation of triples
  - Authorized users can edit FOX results
  - Entries of users are added as triples with source = user name

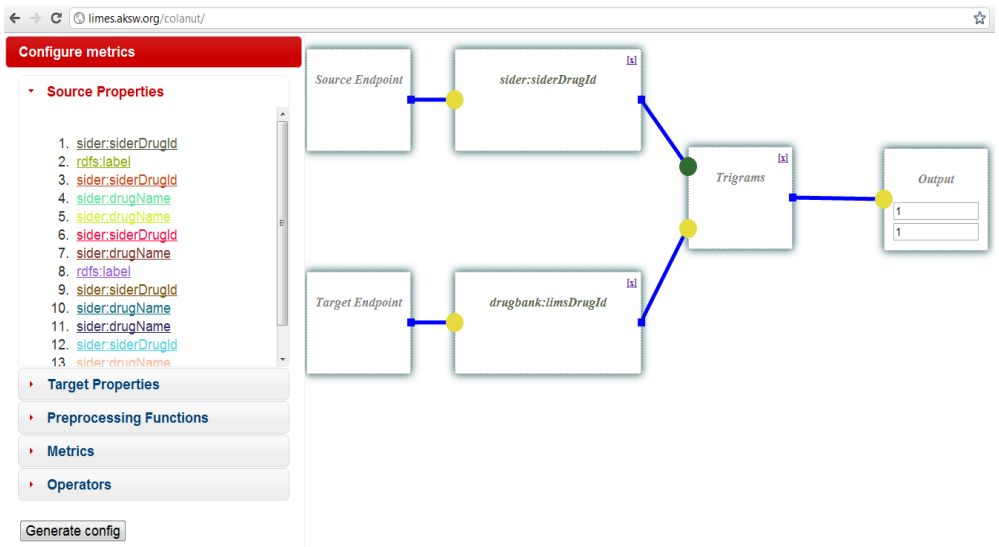


Linking


Semi-structured Data
- CMS
- Text databases
- …
Source Wrapper
RDF+Text
- Access Control
- Knowledge Curation
- Data Integration
OntoWiki
RDF
Orchestration Service
Text
RDF
FOX
- Data Integration
- Enrichment
RDF
- Analysis of unstructured data
LIMES
RDF
XML
Triple Store
Web Crawler
- Data Integration
- Enrichment

LIMES

- Create links between graphs documents via
  - Named Entities
  - Keywords
- Creation of link specs for each case
- Iterative application after update batch

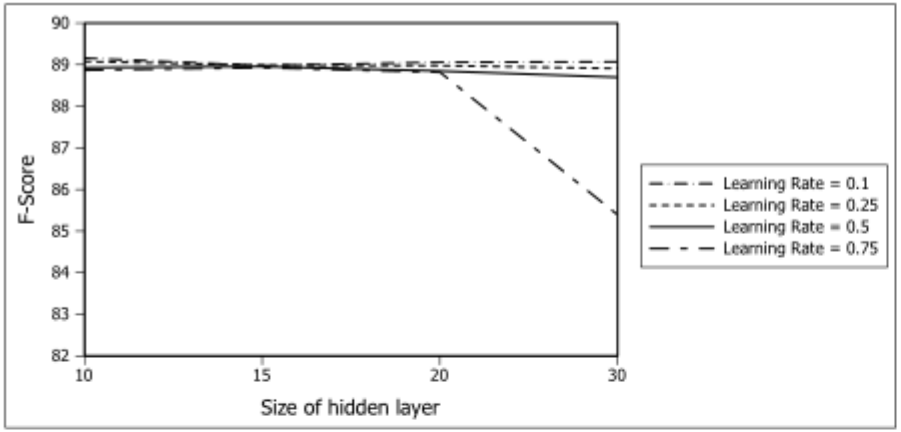


http://limes.aksw.org/colanut

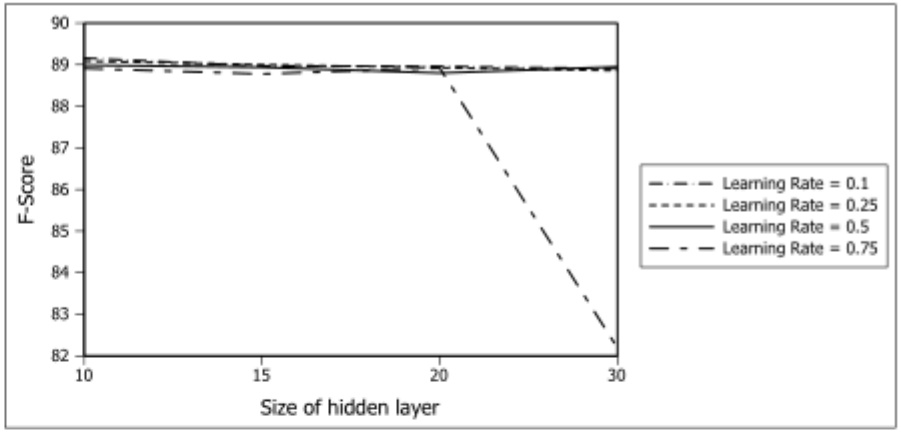
Evaluation

- 4 experimental settings
  - News Data
  - Website Data
  - Actors profiles
  - Country profiles

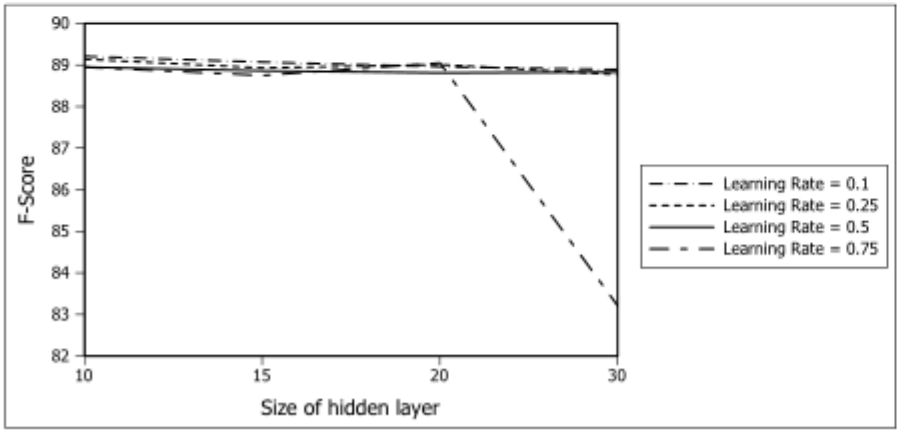
Evaluation (News Data)



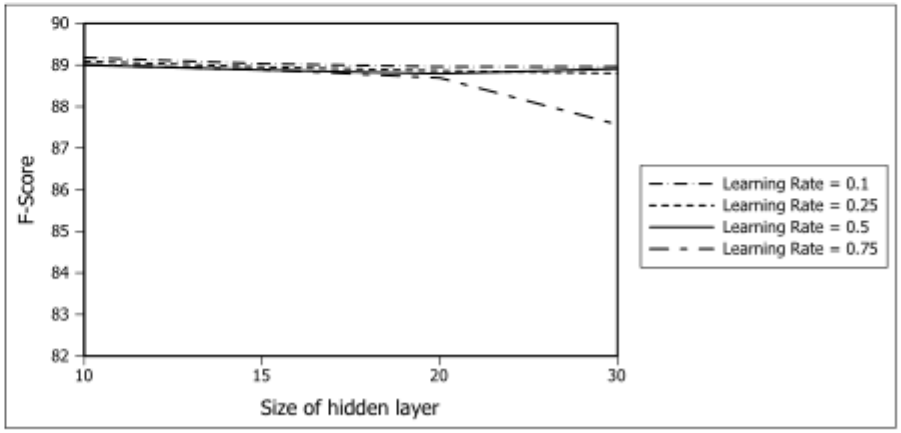
(a) Iterations = 500



(b) Iterations = 1000

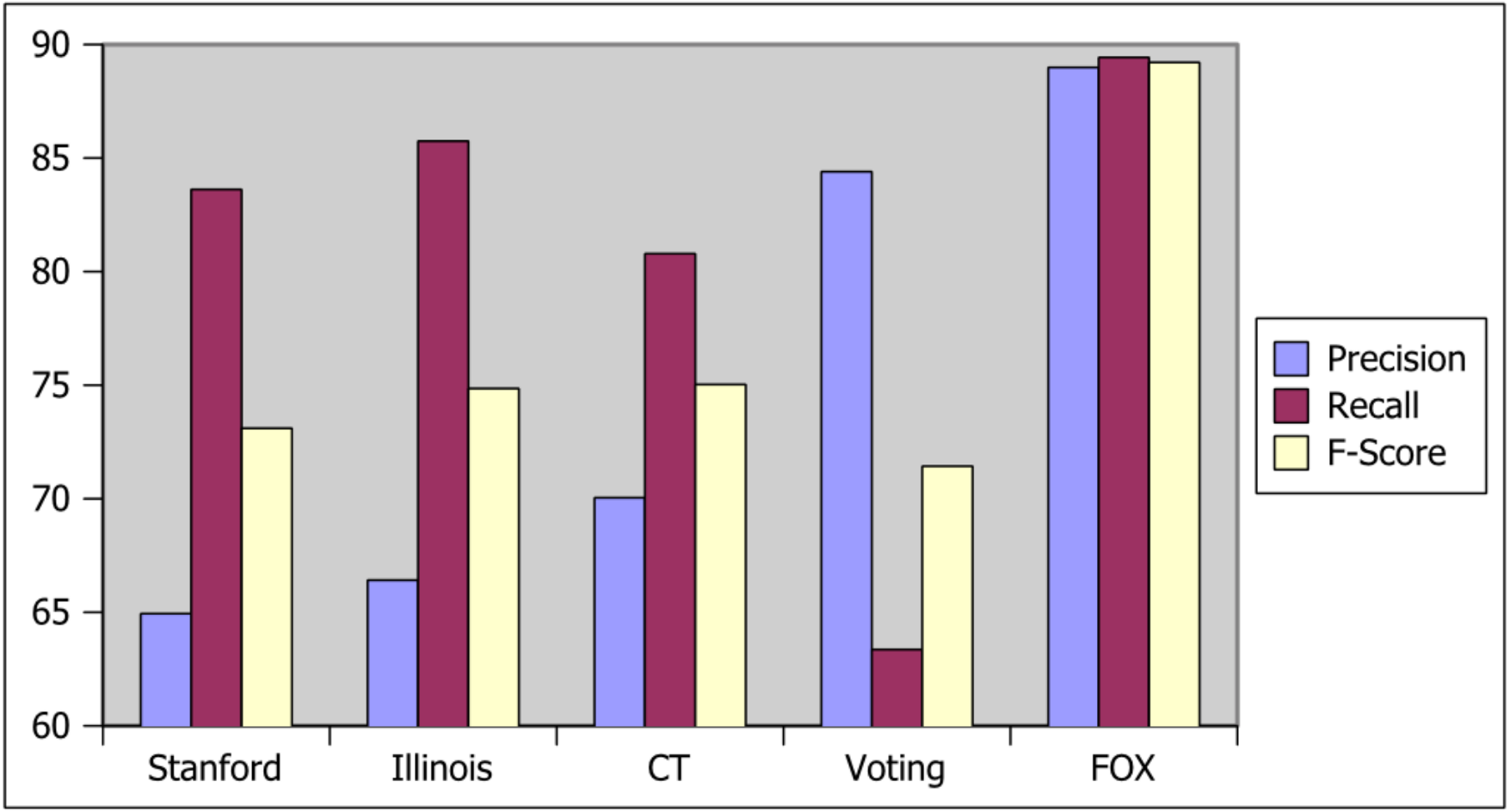


(c) Iterations = 1500

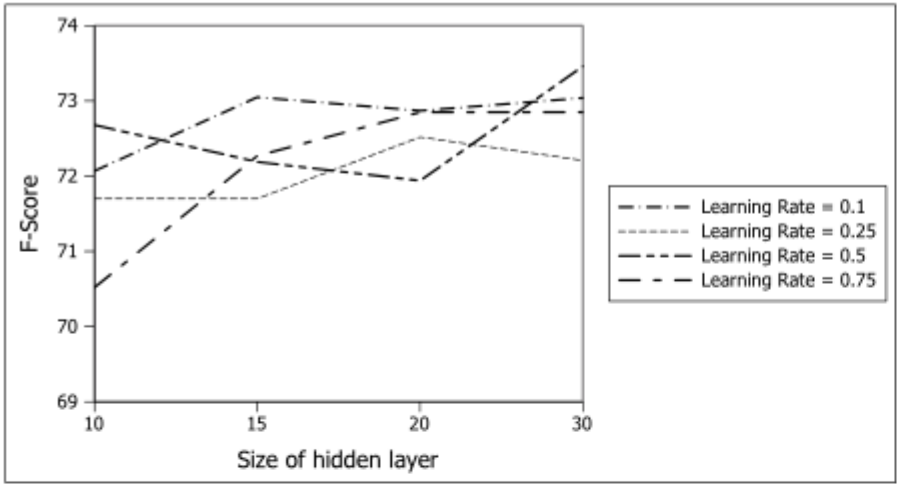


(d) Iterations = 2500

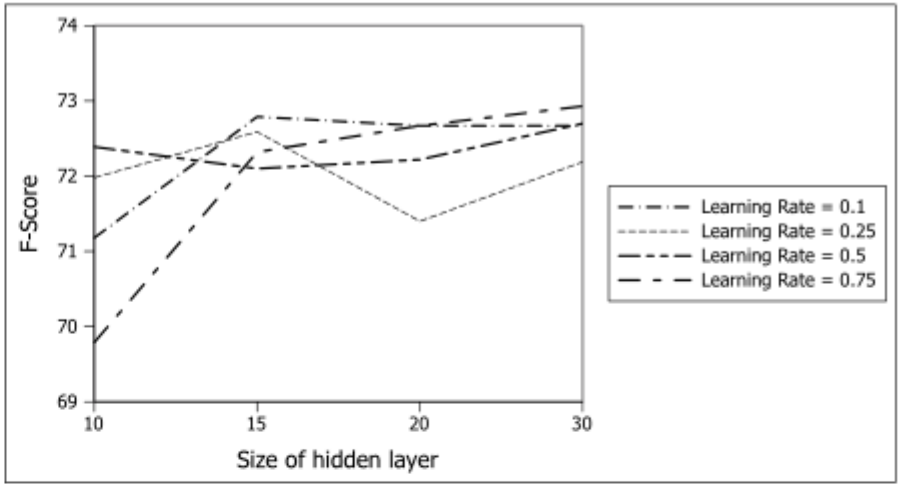
Evaluation (News Data)



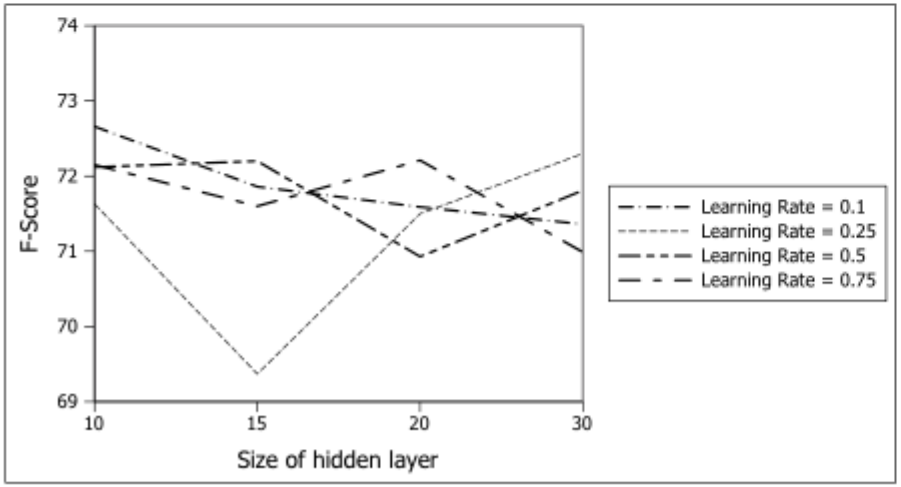
Evaluation (Website Data)



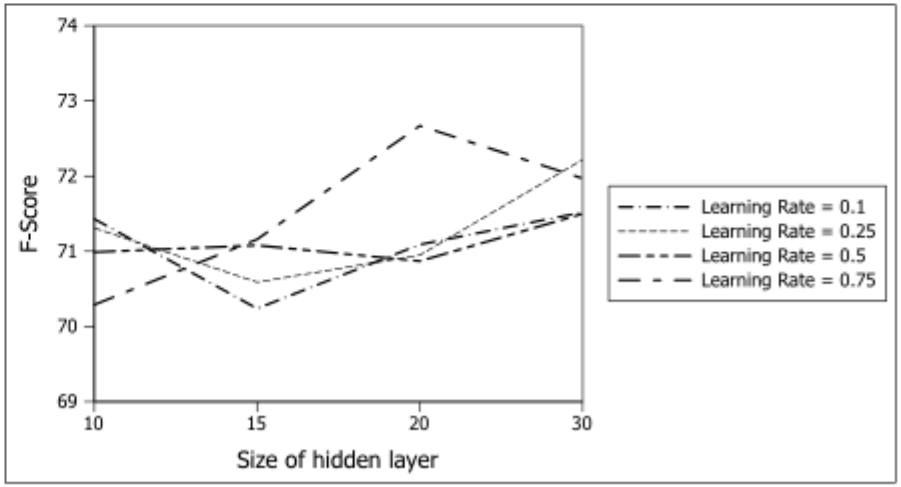
(a) Iterations = 500



(b) Iterations = 1000

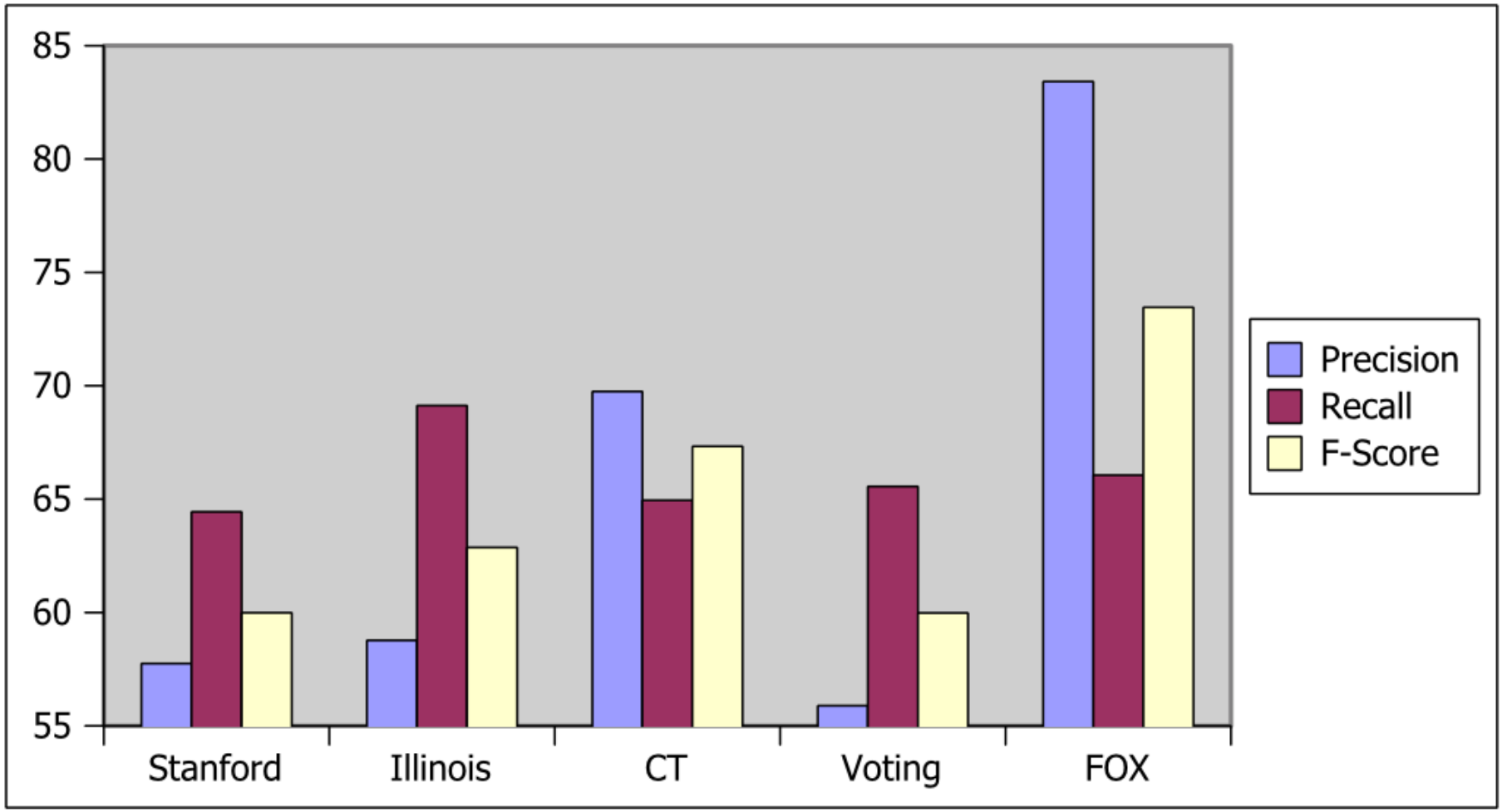


(c) Iterations = 1500



(d) Iterations = 2500

Evaluation (Website Data)



FOX on Reegle + Wikipedia

		Country Profiles		Actors Profiles	
Entity Type	**Measure**	**FOX**	**CS**	**FOX**	**CS**
Location	Precision	98%	100%	83.33%	100%
	Recall	94.23%	78.85%	90%	70%
	F-Score	**96.08%**	88.17%	**86.54%**	82.35%
Organization	Precision	73.33%	100%	57.14%	90.91%
	Recall	68.75%	40%	69.23%	47.44%
	F-Score	**70.97%**	57.14%	**62.72%**	62.35%
Person	Precision	–	–	100%	100%
	Recall	–	–	45.45%	54.55%
	F-Score	–	–	62.5%	**70.59%**
Overall	Precision	93.97%	100%	85.16%	98.2%
	Recall	91.60%	74.79%	70.64%	52.29%
	F-Score	**92.77%**	85.58%	**77.22%**	68.24%

Conclusion

- Presented SCMS Framework
- Combines flexibility and accuracy
  - No direct integration of framework into CMS
  - Communication via RDF
  - Up to 13% more F-Score than traditional frameworks
- Work on IRI disambiguation
- Apply the system in live environment

Thank You!

Questions?

Axel Ngonga
ngonga@informatik.uni-leipzig.de
Johannisgasse 26, Room 5-22
04103 Leipzig, Germany

http://scms.eu

http://fox.aksw.org