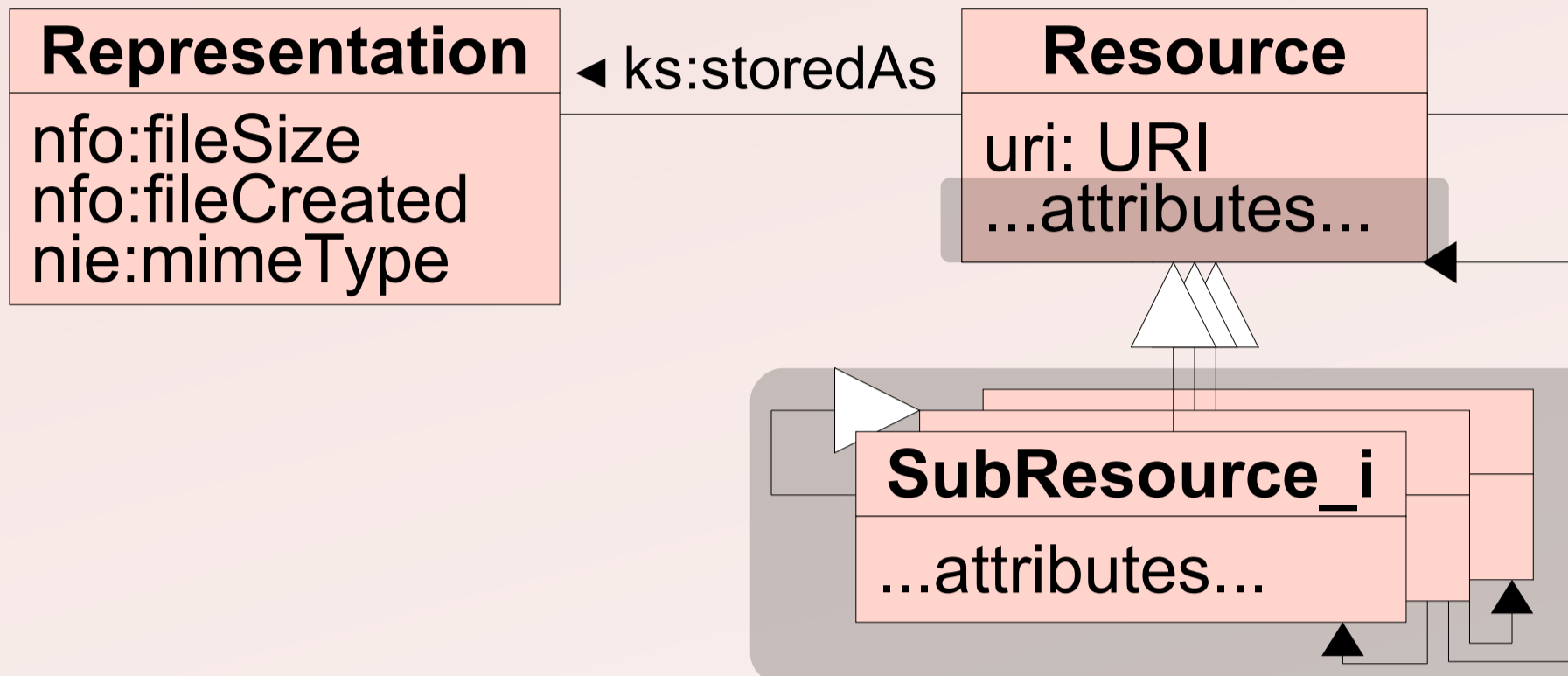


KnowledgeStore

Scalable Framework for Interlinking Text and Knowledge

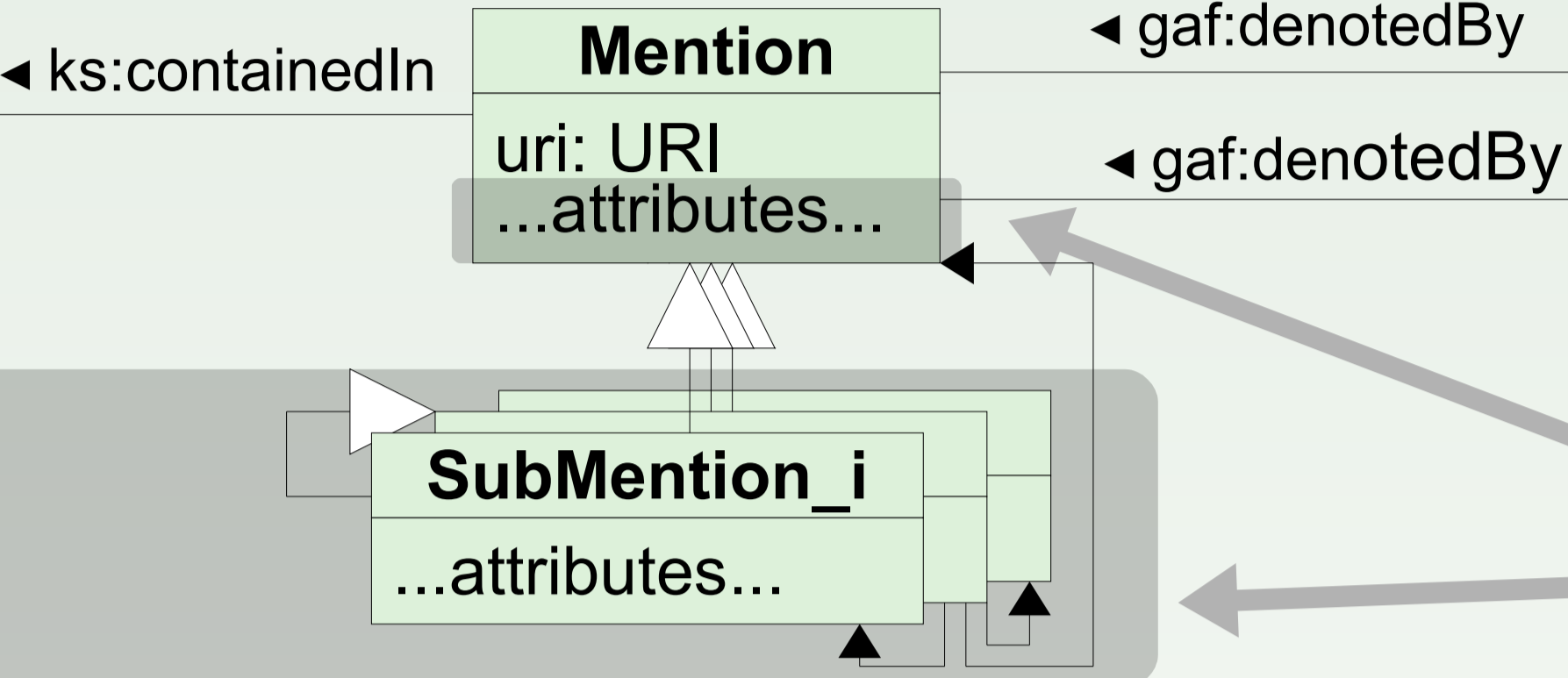
Resources

- arbitrary text documents, e.g., news and associated annotation files
- both digital representation (file) and associated metadata attributes



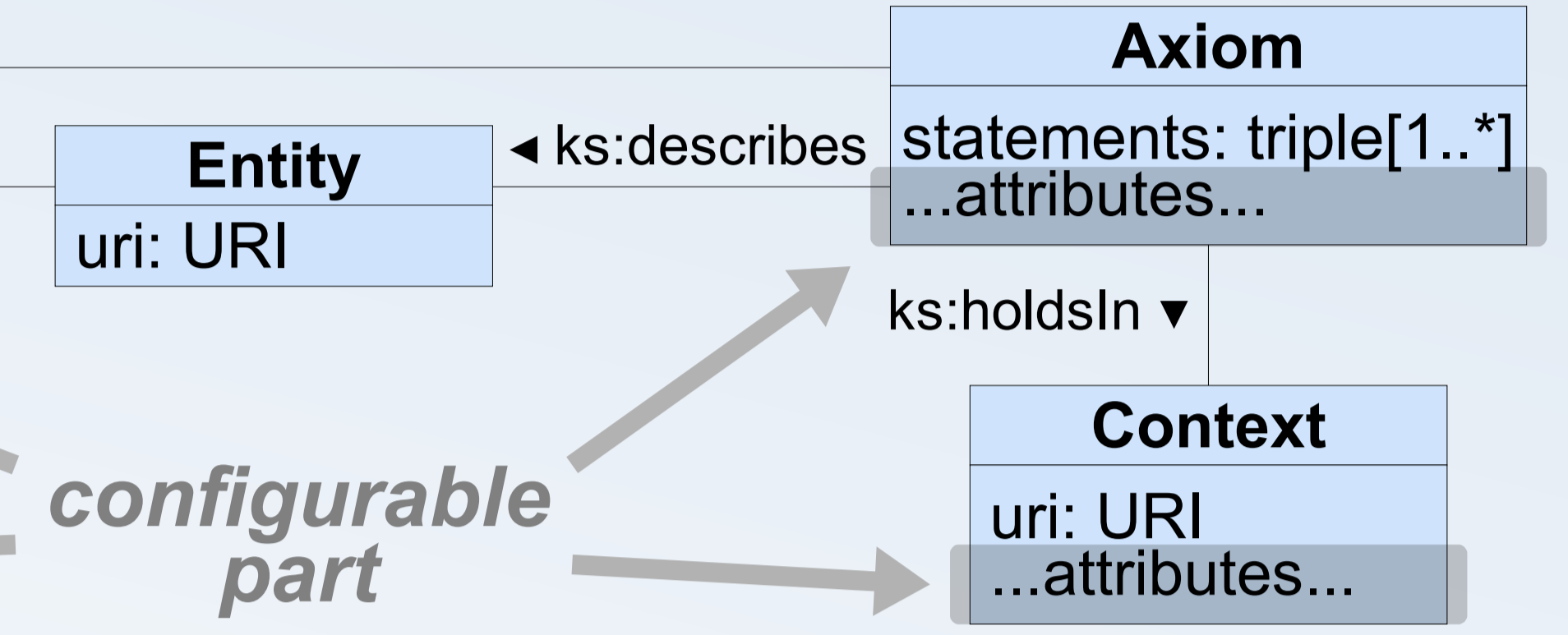
Mentions

- snippets of resources denoting something of interest - entity / axiom
- sign vs referent distinction
- described by linguistic metadata attributes



Entities

- persons, organizations, events, ...
- described by logical axioms (~triples)
- axiom validity restricted to a specific context (e.g., time span, attribution pair)



KnowledgeStore UI

ID http://www.bbc.co.uk/sport/0/football/22251818

Resource text [Select resource metadata](#) [Select mention \(151\)](#)

Alexandra Wrage, an authority on corporate anti-corruption, quit the Independent Governance Committee (IGC) last week, BBC Sport has learned.

World football's governing body set up the IGC to help it become more transparent following several scandals. In a recent interview, Wrage said the proposed reforms had been "neutered".

Fifa launched its reform process almost two years ago amid fierce criticism after Mohamed bin Hammam, an election rival to president Sepp Blatter, was accused of bribery. Bin Hammam was later banned for life by Fifa, but he continues to deny any wrongdoing.

Resource metadata

dct:created	2013-04-22T13:20:19
dct:language	lexvo:eng
dct:title	Fifa adviser resigns in dispute over 'neutered' reform proposals
rdfs:type	nwr:News

KnowledgeStore UI

ID http://www.bbc.co.uk/sport/0/football/22251818#char=741,747&word=w131&term=t131

Mention data

gaf:denotes	<./22251818#banEvent>
ks:mentionOf	<./22251818>
nwr:eventClass	nwr:event_speech_cognitive
nwr:factualityConfidence	0.7226601421959593
nwr:framenetRef	framenet:Prohibiting
nwr:pos	nwr:pos_verb
nwr:pred	ban
nwr:propbankRef	<./ban.01>
nwr:verbnetRef	<./forbid-67>
nif:beginIndex	741
nif:endIndex	747
rdfs:type	nwr:EventMention

KnowledgeStore UI

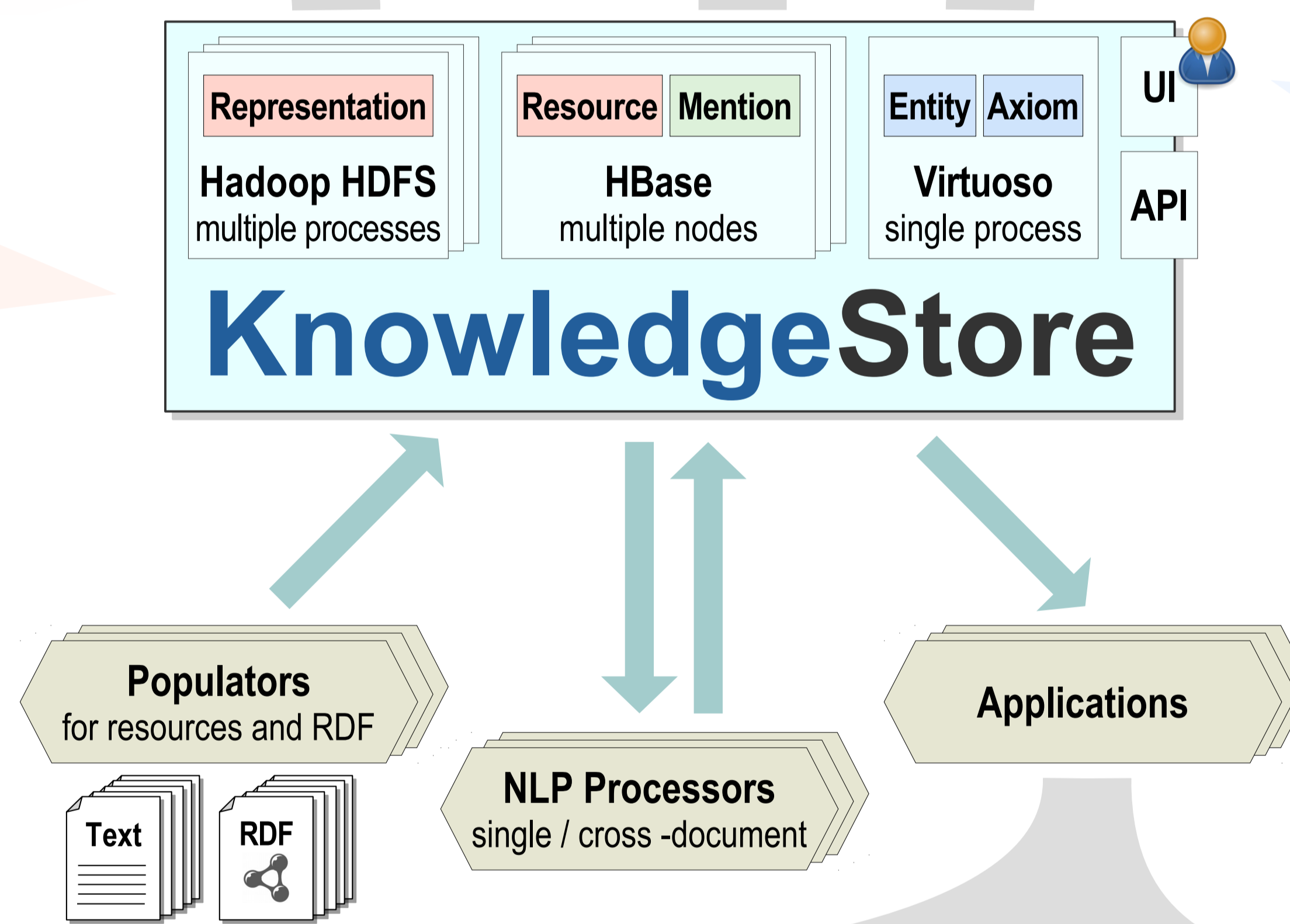
ID http://www.bbc.co.uk/sport/0/football/22251818#banEvent

Entity triples (8, max 1000)

subject	predicate	object
<./22251818#banEvent>	rdfs:type	<./communication>
<./22251818#banEvent>	rdfs:type	framenet:Prohibiting
<./22251818#banEvent>	rdfs:type	sem:Event
<./22251818#banEvent>	rdfs:label	ban
<./22251818#banEvent>	gaf:denotedBy	<./22251818#char=741,747&word=w131&term=t131>
<./22251818#banEvent>	gaf:denotedBy	<./22251818#char=2251,2257&word=w518&term=t518>
<./22251818#banEvent>	sem:hasActor	dbpedia:FIFA
<./22251818#banEvent>	sem:hasActor	dbpedia:Mohammed_bin_Hammam
<./22251818#banEvent>	sem:hasTime	<./22251818#naIHeader_fileDesc_creationTime>

Architecture

- the KnowledgeStore is a scalable storage system built on top of Hadoop, HBase and Virtuoso
- populators feed raw data (resources and RDF triples) in the KnowledgeStore
- an external NLP pipeline extracts mentions and links them to entities
- applications access contents via SPARQL and CRUD API
- a Web UI allows inspecting and querying contents



SPARQL Interface

KnowledgeStore UI

```
SELECT DISTINCT ?event ?datetime ?label
{
  ?event rdfs:label ?label;
  sem:hasTime ?t;
  sem:hasActor dbpedia:FIFA ,
  dbpedia:Mohammed_bin_Hammam .
  ?label bif:contains "ban" .
  ?t rdfs:label ?datetime .
}
```

ORDER BY DESC(?datetime) LIMIT 100

Display results

event	datetime	label
<./22380797#banEvent>	2013-05-02T10:10:49	ban
<./22251818#banEvent>	2013-04-22T13:20:19	ban
<./21944765#banEvent>	2013-03-27T16:34:57	ban
<./21944765#banEvent>	2013-03-27T16:26:15	ban
<./545H-YRRI-DYTG-N238.xml#banEvent>	2011-11-02T11:31:45	ban
<./545H-YRRI-DYTG-N238.xml#banEvent>	2011-11-02T00:00:00	ban
<./53YJ-YHJ1-JBTB>	2011-10-05T18:58:28Z	ban

Decision making in the Automotive Industry

- news aggregated by LexisNexis on car industry (copyrighted), 2004-2013, EN
- LOD data from DBpedia EN, ES, IT, NL
- DSS tool by Synerscope and various analytical visualizations on top of the KnowledgeStore



News	Mentions	Entities	Triples	
			Text	LOD
63.6K	8M	2.2M	33M	270M

Data journalism using FIFA World Cup news

- news from LexisNexis (copyrighted), BBC, The Guardian, 2005-2014, EN
- LOD data from DBpedia EN
- hackathon using data in the KnowledgeStore (40 developers, 30K queries, 20 query/s max)



News	Mentions	Entities	Triples	
			Text	LOD
212K	72M	10.2M	136M	104M

Manual annotation of Wikinews texts

- news on major global events (public), 2004-2014, EN
- LOD data from DBpedia EN
- ongoing manual annotation to produce NewsReader NLP gold standard



News	Mentions	Entities	Triples	
			Text	LOD
17.9K	3.4M	940K	10M	104M



References:

- Cattoni, R., Corcoglioniti, F., Girardi, C., Magnini, B., Serafini, L., and Zanoli, R. *The KnowledgeStore: an entity-based storage system.* In: 8th Int. Conf. on Language Resources and Evaluation (LREC2012)
- Corcoglioniti, F., Rospocher, M., Cattoni, R., Magnini, B. and Serafini, L. *Interlinking unstructured and structured knowledge in an integrated framework.* In: IEEE 7th Int. Conf. On Semantic Computing (ICSC2013)

Web site: <http://knowledgestore.fbk.eu/>

Powered By:

