

On the Diversity and Availability of Temporal Information in Linked Open Data

Anisa Rula¹, Matteo Palmonari¹, Andreas Harth²,
Steffen Stadtmueller², and Andrea Maurino¹

1. University of Milano-Bicocca

2. Karlsruhe Institute of Technology (KIT)

Outline

- Motivation & Research Question
- Temporal Information in the LOD cloud
 - General analysis of temporal information
 - Temporal meta-information
- Temporal Meta-information Analysis
 - Systematic review of proposed models
 - Quantitative analysis based on large-scale experiment
- Conclusion
 - Guidelines for consumers and publishers
 - Future Works

Outline

- Motivation & Research Question
- Temporal Information in the LOD cloud
 - General analysis of temporal information
 - Temporal meta-information
- Temporal Meta-information Analysis
 - Systematic review of proposed models
 - Quantitative analysis based on large-scale experiment
- Conclusion
 - Guidelines for consumers and publishers
 - Future Works

Temporal metadata on the Web

Search bar containing "rihanna news" with a microphone icon and a search button.

Navigation tabs: **Web**, Images, Maps, Shopping, News, More ▾, Search tools

About 261,000,000 results (0.24 seconds)

November, 9th 2012

News for [rihanna news](#)



[Rihanna and Chris Brown act like a couple, but both are in denial](#)

[Examiner.com](#) - 2 hours ago

Rihanna and Chris Brown have both come out separately recently and ...
And we're working on our friendship now", according to **MTV News**.

Us Magazine

[Rihanna](#) Previews 'Unapologetic' at Benefit for Hurricane Sandy Victims

[RollingStone.com](#) - 4 hours ago

Fashion shocker of the week: [Rihanna](#) and Kate's S&M photo shoot for V Magazine

[Examiner.com](#) - 2 hours ago

[Rihanna Breaking News](#) and Photos | [PerezHilton.com](#) 📶

[perezhilton.com](#) › [Rihanna](#)

1 hour ago – See what **Rihanna** has to say about Madonna here!

[Rihanna Breaking News](#) and Photos | [Just Jared](#) 📶

[www.justjared.com/tags/rihanna/](#)

1 day ago – "Nobodies Business' is basically the way I look at everything regarding my personal life," **Rihanna** shared. "Even though you have to witness it, ...

[Katy Perry, Rihanna](#) Rift: Singers Have Fallout Over Chris Brown ... 📶

[www.huffingtonpost.com/.../katy-perry-rihanna-rift-singers-have-fall...](#)

5 days ago – According to the New York Daily **News**, **Rihanna's** relationship with Chris Brown has put a strain on her friendship with the "Part of Me" singer, ...

Web of Data is dynamic

How much up-to-date are linked open data?

The screenshot shows the SIG.MA Semantic Information Mashup interface. The main content area displays the profile for Antonio Cassano, including his given name, family name, and a detailed comment. A red box highlights the 'currentclub' field, which is 'Associazione Calcio Milan S.p.A.'. The right sidebar shows a list of 13 sources, with the first source being 'Antonio Cassano' with 12 facts, dated 2012-11-03, and a URL from rdf.basekb.com.

SIG.MA
SEMANTIC INFORMATION MASHUP

Help About Support

Version: 2.0.8

Antonio Cassano Add More Info Start New Order Options Use it

Antonio Cassano

given name: Antonio [2]
family name: Cassano [2]

comment: Antonio Cassano (born 12 July 1982) is an Italian professional footballer who plays for Serie A club Milan as a striker. Nicknamed Il Gioiello di Bari Vecchia ("the jewel of Old Bari"), he is known for his short temper as much as his ability on the pitch, which led to the coining of the neologism Cassanata by his former coach, Fabio Capello, in November 2002. The word is regularly used by Italian journalists as a euphemism for any behavior incompatible with team spirit in football. Cassano was born in Bari and his father left the family shortly thereafter. He was raised in the Bari Vecchia neighbourhood, and began playing football at an early age. Cassano was spotted by an A.S. Bari scout and brought up through the team's youth system, and he made his Serie A debut for Bari against local rivals U.S. Lecce in 1999. At the age of nineteen, Cassano signed with reigning Serie A champions A.S. Roma, for a transfer fee of 60billion Italian lire (about €30 million). His first season produced five goals, and he attracted media attention after openly clashing with coach Fabio Capello after he was left out of a practice match a few days after his international debut. In the 2003 Italian Cup [1]

Antonio Cassano (born 12 July 1982 in Bari) is an Italian footballer who plays for Serie A club Milan, mainly as a forward. Nicknamed Il Gioiello di Bari Vecchia (the jewel of Old Bari), he is known for his short temper as much as his ability on the pitch, which led to the coining of the neologism Cassanata by his former coach, Fabio Capello, in November 2002. The word is regularly used by Italian journalists as a euphemism for any behavior incompatible with team spirit in football. [2]

Antonio cassano [7]
Passionnement football [8]
Canale Juve - Notizie, curiosità e approfondimenti sul mondo della Juventus [13]
MILANO - C [20]

currentclub: Associazione Calcio Milan S.p.A. [2]

club: http://dbpedia.org/resource/AS_Rom [2]
[Unione Calcio Sampdoria_SpA](http://dbpedia.org/resource/Sampdoria_SpA) [2]
http://dbpedia.org/resource/Sampdoria_Genua [2]
[Real Madrid](http://dbpedia.org/resource/Real_Madrid) [2]
http://dbpedia.org/resource/AC_Mailand [2]
[AS Bari](http://dbpedia.org/resource/AS_Bari) [2]

Sources (20) Approved (0) Rejected (0)

- 1 Antonio Cassano 12 facts | 2012-11-03
<http://rdf.basekb.com/ns/m.036wtl>
- 2 Antonio Cassano 345 facts | 2012-06-05
http://dbpedia.org/resource/Antonio_Cassa...
- 3 Antonio Cassano 2 facts | 2012-06-05
<http://dbpedia.org/class/yago/AntonioCass...>
- 4 Photos for DBpedia.org r... 4 facts | 2011-05-30
<http://www4.wiwiss.fu-berlin.de/flickrwra...>
- 5 Sanremo: fuori D'Angelo... 5 facts | 2010-03-02
<http://www.ansa.it/web/enrichednews/topne...>
- 6 Prandelli convoca Cassan... 6 facts | 2010-08-06
<http://www.ansa.it/web/enrichednews/topne...>
- 7 Antonio cassano 11 facts | 2010-07-16
<http://it.bestshopping.com/prezzi/Antonio...>
- 8 Antonio Cassano 13 facts | 2011-03-04
<http://www.footiz.com/soccer/antonio-cass...>
- 9 Cassano 10 facts | 2012-06-05
<http://dbpedia.org/resource/Cassano>
- 10 antonio cassano 13 facts | 2011-07-05
<http://www.youradfree.com/wordpress/tag/a...>
- 11 Pictures of Antonio Cass... 20 facts | 2012-06-21
<http://ookaboo.com/o/pictures/topic/25057...>
- 12 Untitled document 1 facts | 2012-07-13
<http://www.voetbal.be/nieuws-tags/antonio...>
- 13 Antonio Cassano | Canal... 18 facts | 2012-05-09
<http://www.canalejuve.it/tags/antonio-cas...>

<- 1 [2] -> reject all approve all

November, 2nd 2012

Web of Data is dynamic

How much up-to-date are linked open data?


SIG.MA
SEMANTIC INFORMATION MASHUP

Help About Support


Version: 2.0.8

Antonio Cassano [Add More Info](#) [Start New](#) [Order](#) [Options](#) [Use it](#)

Antonio Cassano



born 12 July 1982) is an Italian professional footballer who plays for Serie A club A.S. Roma of Bari Vecchia ("the jewel of Old Bari"), he is known for his short temper and for the coining of the neologism Cassanata by his former coach, Fabio Capello. The word is regularly used by Italian journalists as a euphemism for any behavior incompatible with the club's interests. He was raised in Bari and his father left the family shortly thereafter. He began playing football at an early age. Cassano was spotted by an A.S. Roma scout while playing for a local rivals U.S. Grotto. He signed for a transfer fee of €10 million. His first season was successful, earning him media attention after he was left out of a practice match a few days before his international debut.



born 12 July 1982 in Bari) is an Italian footballer who plays for Serie A club A.S. Roma of Bari Vecchia (the jewel of Old Bari), he is known for his short temper and for the coining of the neologism Cassanata by his former coach, Fabio Capello. The word is regularly used by Italian journalists as a euphemism for any behavior incompatible with the club's interests. He was raised in Bari and his father left the family shortly thereafter. He began playing football at an early age. Cassano was spotted by an A.S. Roma scout while playing for a local rivals U.S. Grotto. He signed for a transfer fee of €10 million. His first season was successful, earning him media attention after he was left out of a practice match a few days before his international debut.

Summer 2012

MILANO - C [20]

currentclub: [Associazione Calcio Milan S.p.A.](#) [2]

club: [http://dbpedia.org/resource/AS_Roma](#) [2]
[Unione Calcio Sampdoria_SpA](#) [2]
[http://dbpedia.org/resource/Sampdoria_Genua](#) [2]
[Real Madrid](#) [2]
[http://dbpedia.org/resource/AC_Milan](#) [2]
[AS Bari](#) [2]

Sources (20) Approved (0) Rejected (0)

- 1 [Antonio Cassano](#) 12 facts | 2012-11-03
[http://rdf.basekb.com/ns/m.036wtl](#)
- 2 [Antonio Cassano](#) 345 facts | 2012-06-05
[http://dbpedia.org/resource/Antonio_Cassano](#)
- 11 [Pictures of Antonio Cassano](#) 20 facts | 2012-06-21
[http://ookaboo.com/o/pictures/topic/25057...](#)
- 12 [Untitled document](#) 1 facts | 2012-07-13
[http://www.voetbal.be/nieuws-tags/antonio...](#)
- 13 [Antonio Cassano | Canal+](#) 18 facts | 2012-05-09
[http://www.canalejuve.it/tags/antonio-cas...](#)

<- 1 2 -> [↑ reject all](#) [↑ approve all](#)

November, 2nd 2012

Temporal information, applications and research areas

Collecting and processing temporal information is crucial for several applications and research areas

- Temporal Query Answering and Search [Alonso&2011]
- Temporal Validity of Statements [Wang&2010]
- Data Fusion and Integration [Mendes&2012]
- Temporal Data Exploration [Alonso&2011]
- Temporal Entity Matching [Li&2011]
- Currency of Statements and Documents [Rula&2012]
- Analysis of provenance of information [Hartig&2009]

Research questions

- To which extent are temporal metadata and temporal information available in the LOD cloud?
- How is the temporal information represented in the LOD cloud?
- To which extent we can collect and interpret such information?

Outline

- Motivation & Research Question
- Temporal Information in the LOD cloud
 - General analysis of temporal information
 - Temporal meta-information
- Temporal Meta-information Analysis
 - Systematic review of proposed models
 - Quantitative analysis based on large-scale experiment
- Conclusion
 - Guidelines for consumers and publishers
 - Future Works

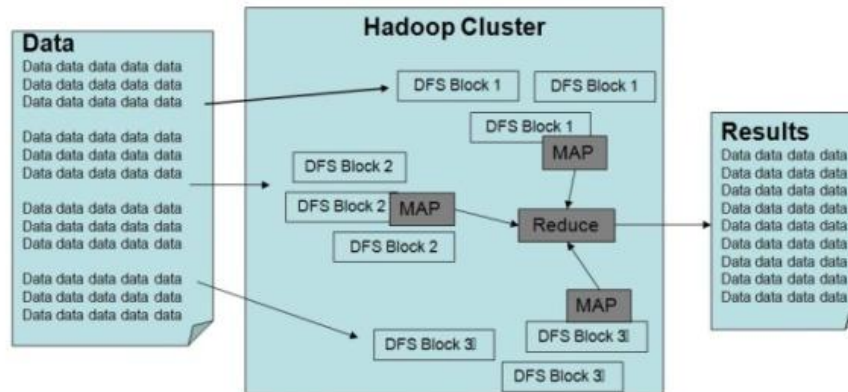
Large-scale experimental analysis

Billion Triple Challenge (BTC) 2011 dataset

- 2.1 billion statements in N-Quads format
- 47K unique predicates
- collected from 7.4M RDF documents

We use Apache Hadoop in our experiments

- parallel and distributed processing of large datasets across clusters of computers (54 work nodes)



Distribution of temporal information

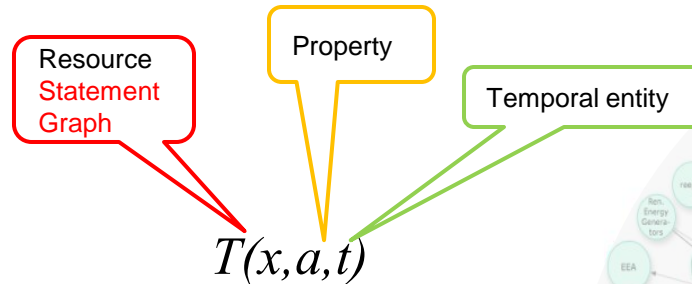
Pay-Level Domain	quads (K)	Tquads. (K)	doc (K)	Tdoc (K)	Temporal Property	quads (K)	doc (K)
scinets.org	56200	3391	51.9	44.3	dcterms:#modified	3400	44
legislation.gov.uk	33100	1249	246.4	246.4	dcterms:modified	2300	842
ontologycentral.com	55300	1029	4.6	4.4	dcterms:date	1500	247
bibsonomy.org	34500	881	234.7	177.3	dc:date	1400	188
loc.gov	7800	854	345.3	302.9	dcterms:created	600	450
bbc.co.uk	6300	679	173.5	83.6	dcterms:issued	200	222
livejournal.com	169800	530	239.2	238.9	lj:dateCreated	200	238
rdfize.com	37600	495	204.7	204.6	swivt:#creationDate	200	197
data.gov.uk	13800	479	178.8	91.9	lj:dateLastUpdated	220	225
dbpedia.org	28400	423	596.6	124.1	wiki:Attribute3ANRHP	180	53
musicbrainz.org	2500	359	0.3	0.3	certificationdate	170	31
tfri.gov.tw	153300	272	154.4	78.2	tl:timeline.owl#start	150	24
archiplanet.org	16300	186	79.2	53.5	tl:timeline.owl#end	140	143
freebase.com	27800	173	572.9	109.1	bio:date	140	15
vu.nl	6800	156	294.2	26.7	po:scheduledate	96	37
fu-berlin.de	5700	139	291.6	37.4	swrc:ontology#value	78	0.002
bio2rdf.org	20200	129	744.7	71.6	cordis:endDate	76	26
blogspace.com	900	124	0.2	0.2	nl:currentLocationDateStart	74	10
opera.com	24100	124	160.3	124.1	po:startofmediaavailability	68	68
myexperiment.org	1500	114	26.1	13.7	foaf:dateOfBirth	62	62
					liteco:dateTime		

Extracted using standard date formats and variations

Temporal Information vs Temporal meta-information (Abstract Definition)

Temporal information

a ternary relation $T(x,a,t)$, where x is a resource, a statement, or a graph, a is a property, and t is a temporal entity

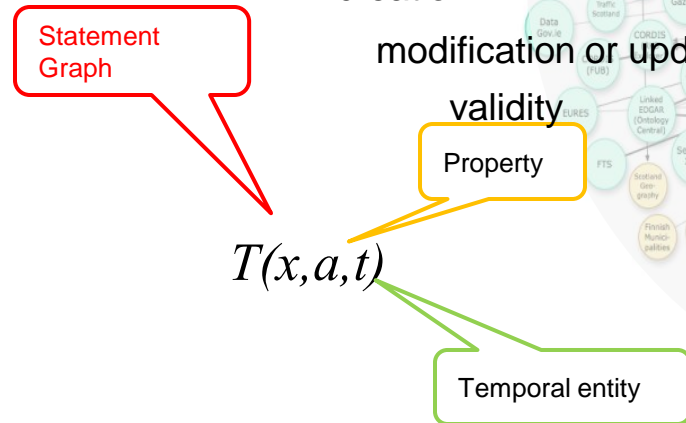


$T(:Rihanna,foaf:dateOfBirth,1980-09-04)$

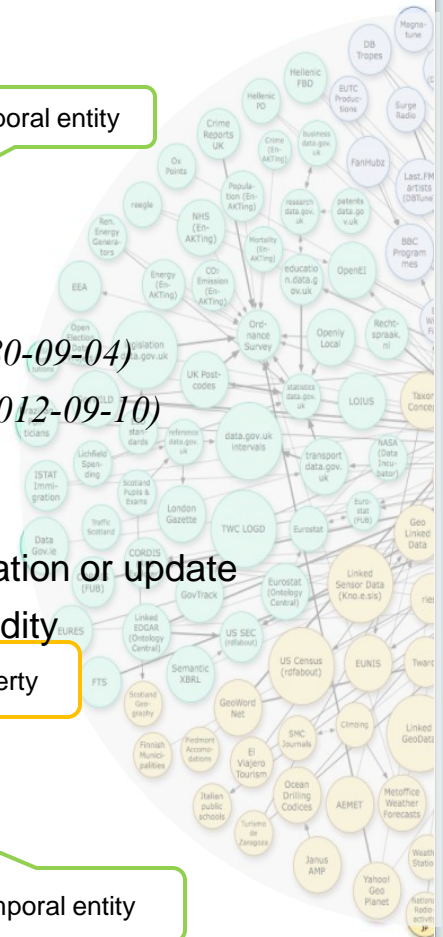
$T(foafRihanna.rdf,dc:modified,2012-09-10)$

Temporal meta-information (TMI)

a temporal information $T(x,a,t)$ is interpreted as a temporal meta-information if and only if x is interpreted as a truth-valued RDF element (statement or graph)



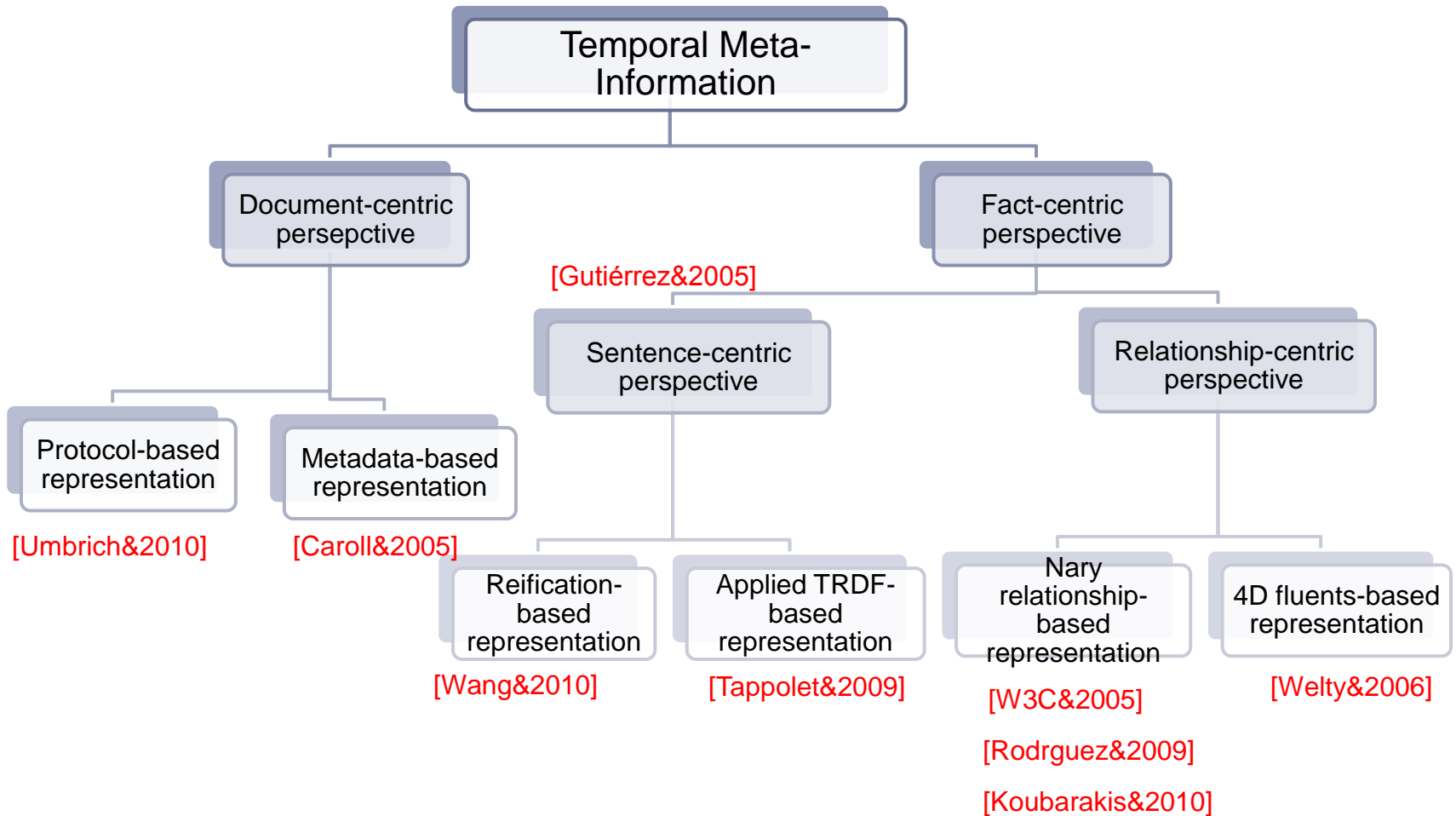
$T(foafRihanna.rdf,dc:modified,2012-09-10)$



Outline

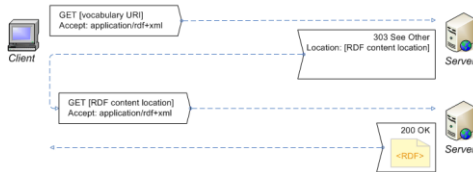
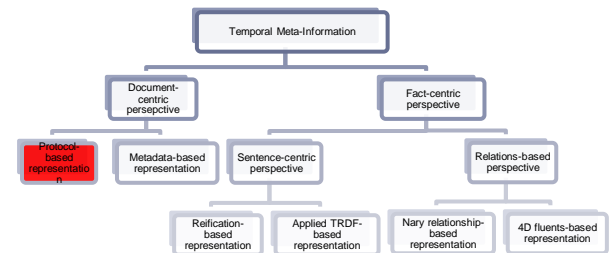
- Motivation & Research Question
- Temporal Information in the LOD cloud
 - General analysis of temporal information
 - Temporal meta-information
- Temporal Meta-information Analysis
 - Systematic review of proposed models
 - Quantitative analysis based on large-scale experiment
- Conclusion
 - Guidelines for consumers and publishers
 - Future Works

Models for representing TMI



Document-centric perspective

A) Protocol-based representation



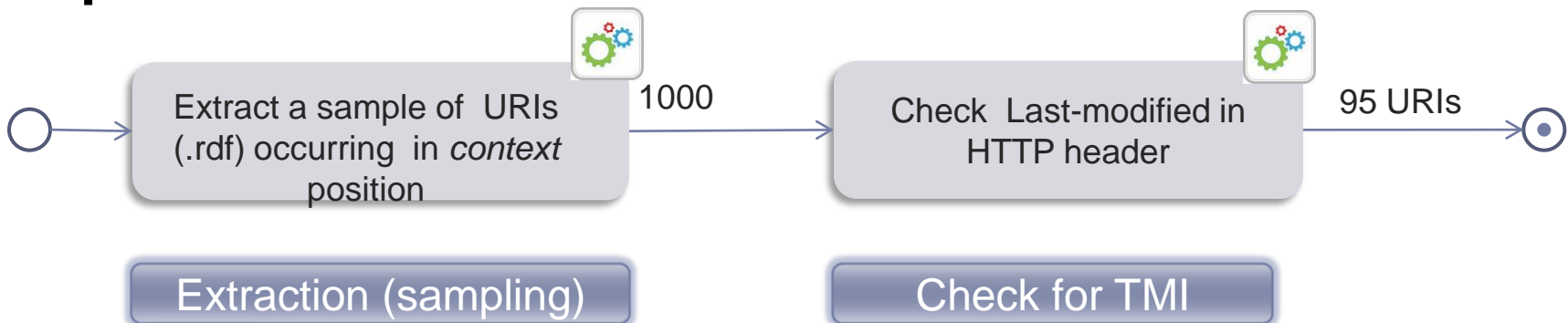
Look up: **<http://ex/data/Antonio_Cassano>**

Check: **HTTP Response Header.**

Status: **HTTP/1.1 200 OK**

Select: **Last-Modified: Tue, 01 January 2012**

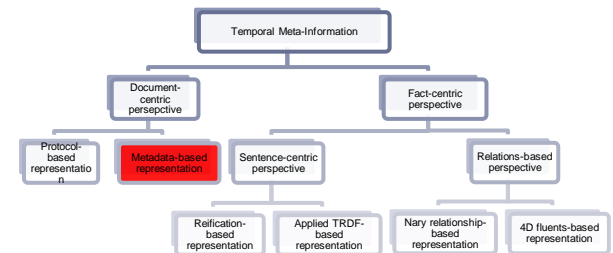
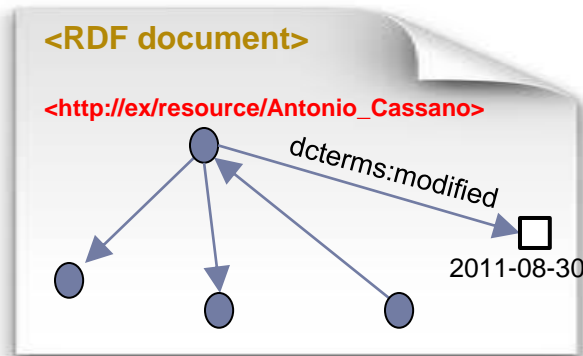
Experiment:



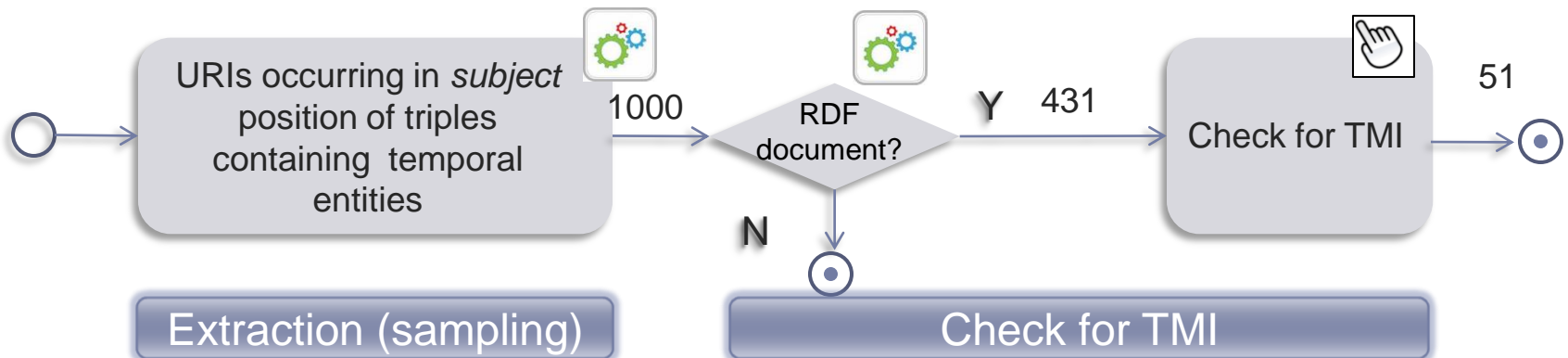
Document-centric perspective

B) Metadata-based representation

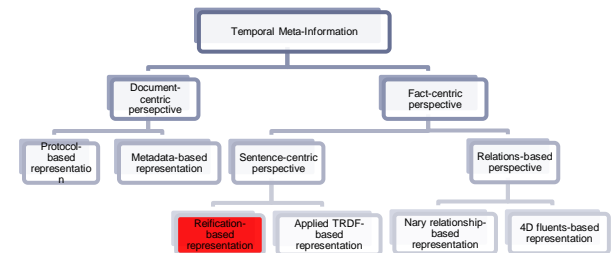
http://ex/data/Antonio_Cassano



Experiment:

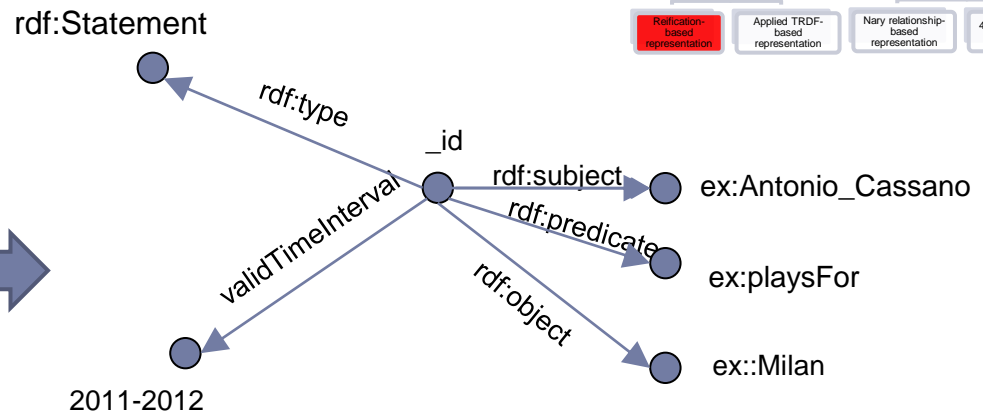
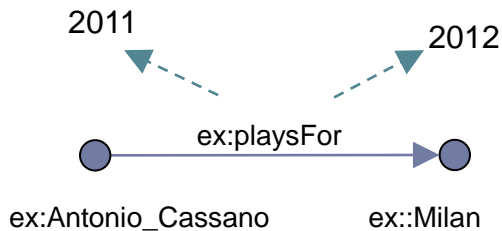


Sentence-centric perspective

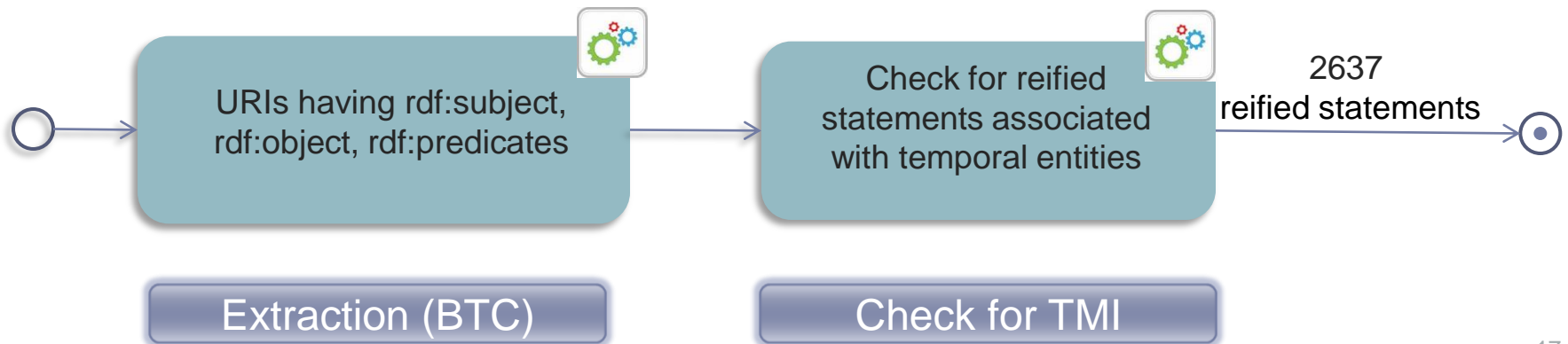


A) Reification-based representation

Temporal RDF (model)

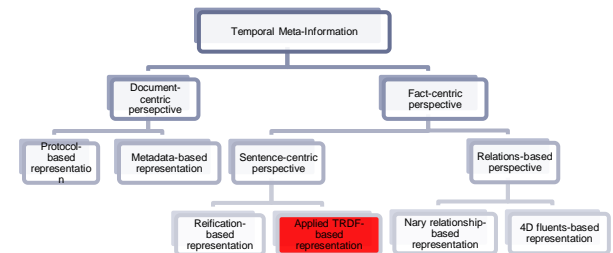


Experiment:

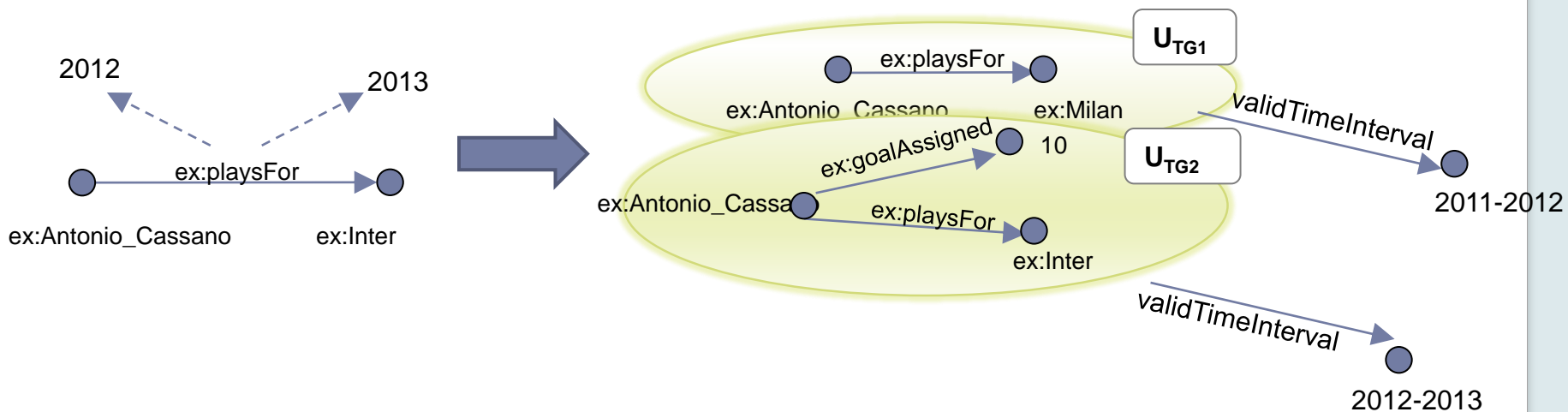


Sentence-centric perspective

B) Applied temporal RDF-based representation



Temporal RDF (model)

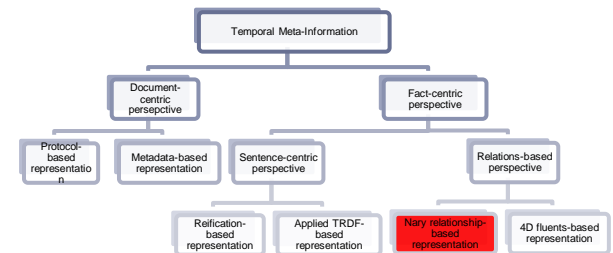
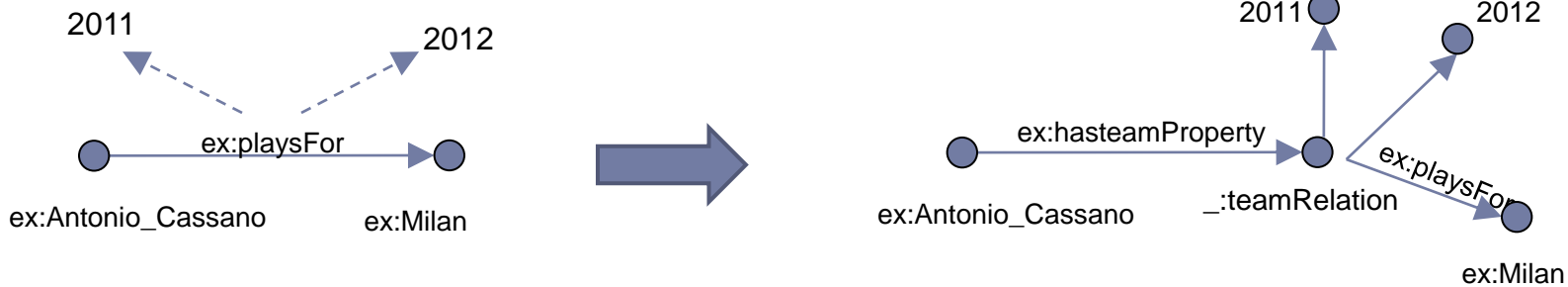


Experiment:

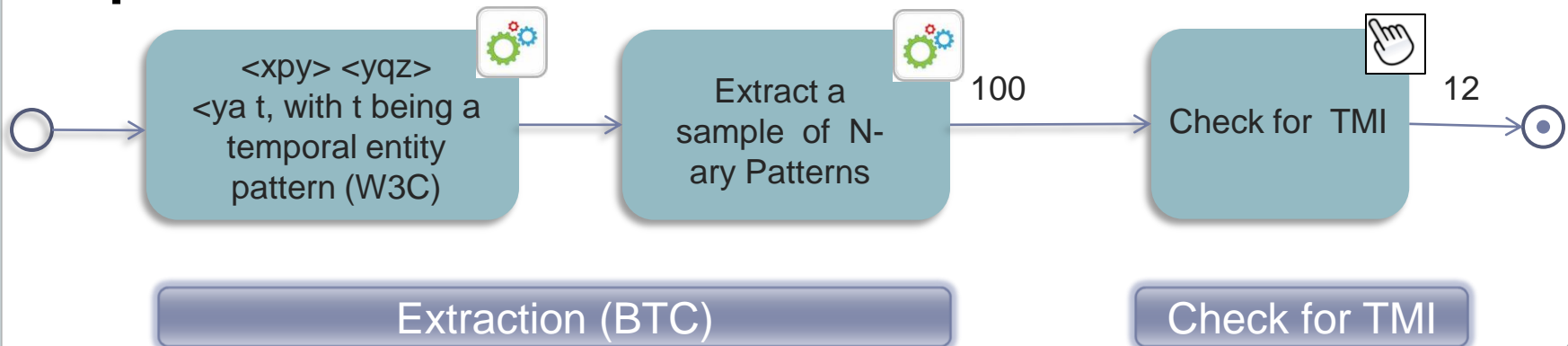
Not found in the BTC

Relationship-centric perspective

A) *N*-ary relationship-based representation

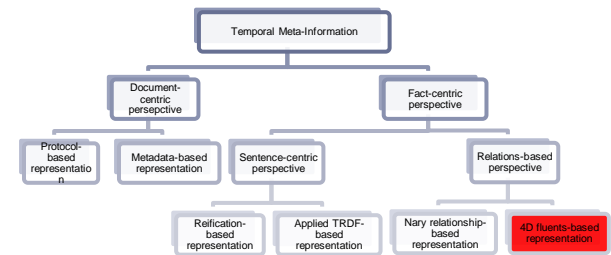
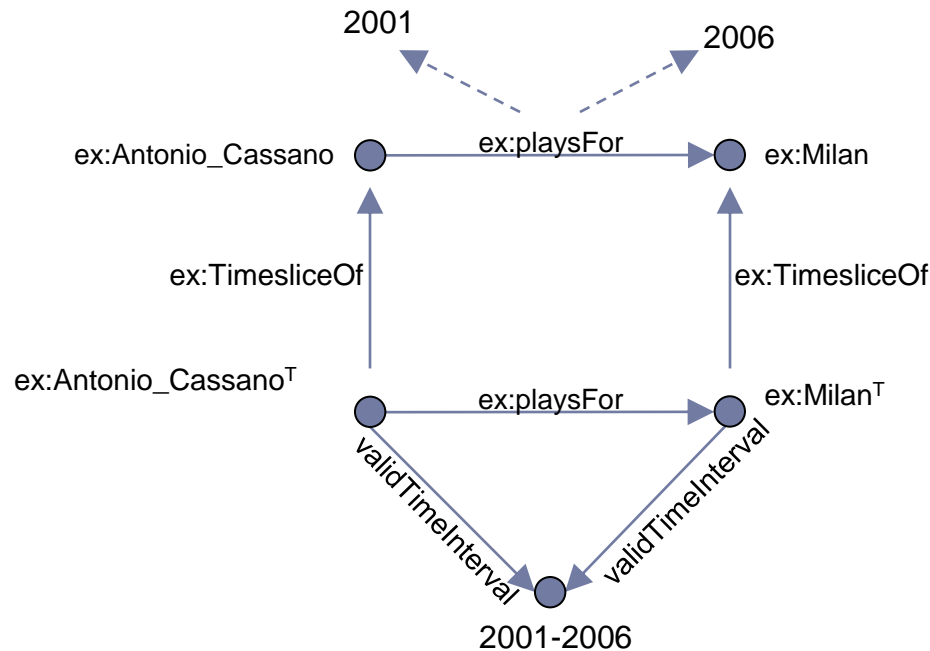


Experiment:



Relationship-centric perspective

B) 4D-fluents-based representation



Experiment:

Not found in the BTC

Availability of TMI

Perspective	Approach	Occurrence temp. quads (%)	Occurrence overall quads (%)	Occurrence overall docs (%)
Document	Protocol	n/a	n/a	9.500
	Metadata	5.10	0.0001900	0.560
Fact	Reification	0.02	0.0000008	0.006
	Applied temporal RDF	-	-	-
	N-ary relationship	12.24	0.0005000	0.600
	4D fluents	-	-	-

- Low availability (in the BTC)
- High diversity (models and vocabularies)
- Automatic interpretation/processing is not trivial (e.g., metadata-based and N-ary relationship)

Outline

- Motivation & Research Question
- Temporal Information in the LOD cloud
 - General analysis of temporal information
 - Temporal meta-information
- Temporal Meta-information Analysis
 - Systematic review of proposed models
 - Quantitative analysis based on large-scale experiment
- **Conclusion**
 - **Guidelines for consumers and publishers**
 - **Future Works**

Guidelines (I)



Perspective	Approach	Consumers	Publishers
Document	Protocol	<ul style="list-style-type: none">- retrieve TMI in the Protocol-based rep	<ul style="list-style-type: none">- provide Last-modified field in the HTTP header- update TMI whenever the data in the document is changed
	Metadata	<p><i>otherwise</i></p> <ul style="list-style-type: none">- retrieve TMI from the Metadata-based rep.	<ul style="list-style-type: none">- check for TMI to be consistent in both Protocol- and Metadata-based rep.

Guidelines (II)



Perspective	Approach	Consumers	Publishers
Fact	Reification	<ul style="list-style-type: none">- retrieve TMI by using the predicates in the RDF reification vocabulary	<ul style="list-style-type: none">- evaluate based on the application scenario if it is possible to avoid such rep.- avoid its use since it is cumbersome with SPARQL queries
	Applied temporal RDF		<ul style="list-style-type: none">- use the Applied temporal RDF rep. rather than pure Reification-based repr.- avoid the worst case when using the applied temporal RDF
	N-ary relationship	<ul style="list-style-type: none">- difficult to be identified	<ul style="list-style-type: none">- use N-ary relationship-based rep. for complex modelling tasks because of its flexibility

Future works

- Model-independent techniques
 - for tracking the modification of documents or triples (extending [Umbrich&2010], [Rula&2012])
 - for the assessment of temporal data qualities in Linked Open Data (data currency and timeliness)
 - for temporal based query answering and search
- Extended version of the paper (journal publication)

Thank you for your attention.
Questions?

References

- [Rula&2012]** A. Rula, M. Palmonari, and A. Maurino. Capturing the Age of Linked Open Data: Towards a Dataset-independent Framework. 1st International Workshop on Data Quality Management and Semantic Technologies at IEEE ICSC, 2012
- [Umbrich&2010]** J. Umbrich, M. Hausenblas, A. Hogan, A. Polleres, and S. Decker. Towards Dataset Dynamics: Change Frequency of Linked Open Data Sources. In 3rd Linked Data on the Web Workshop at WWW, 2010
- [Caroll&2005]** J. J. Carroll, C. Bizer, P. Hayes, and P. Stickler. Namedgraphs. Journal of Web Semantics, 3, 2005
- [Gutiérrez&2005]** C. Gutierrez, C. A. Hurtado, and A. A. Vaisman. Temporal RDF. In The 2ndESWC, pages 93-107, 2005
- [Tappolet&2009]** J. Tappolet and A. Bernstein. Applied Temporal RDF: Efficient Temporal Querying of RDF Data with SPARQL. In The 6th ESWC, pages 308-322, 2009
- [W3C&2006]** <http://www.w3.org/TR/swbp-n-aryRelations/>
- [Welty&2006]** C. Welty, R. Fikes, and S. Makarios. A Reusable Ontology for Fluents in OWL. Frontiers in Artificial Intelligence and Applications, page 226, 2006
- [Hartig&2009]** O. Hartig. Provenance Information in the Web of Data. LDOW2009, April 20, 2009, Spain.
- [Koubarakis&2010]** M. Koubarakis and K. Kyzirakos. Modeling and Querying Metadata in the Semantic Sensor Web: the Model stRDF and the Query Language stSPARQL. The 7th ESWC, pages 425-439, 2010.
- [Rodríguez,&2009]** Rodríguez, R. McGrath, Y. Liu, and J. Myers. Semantic Management of Streaming Data. 2nd International Workshop on Semantic Sensor Networks at ISWC, 2009.