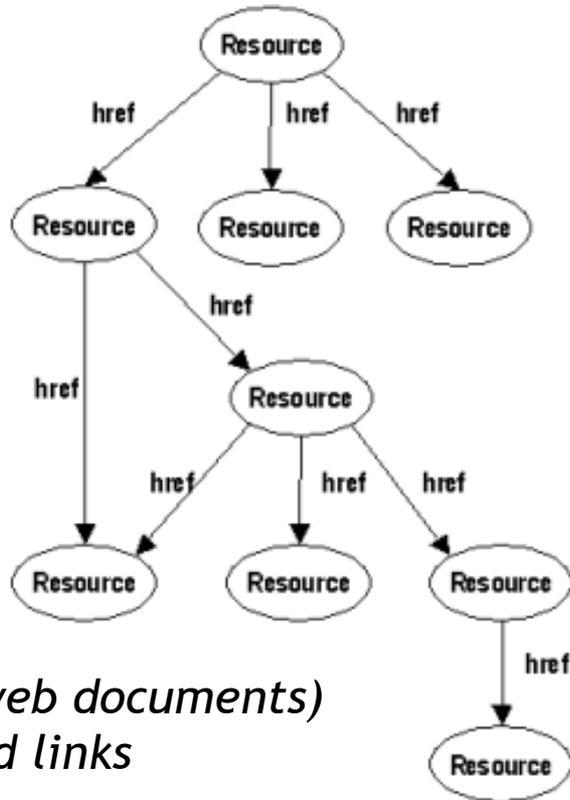


From Hyperlinks to Semantic Web Properties using Open Knowledge Extraction

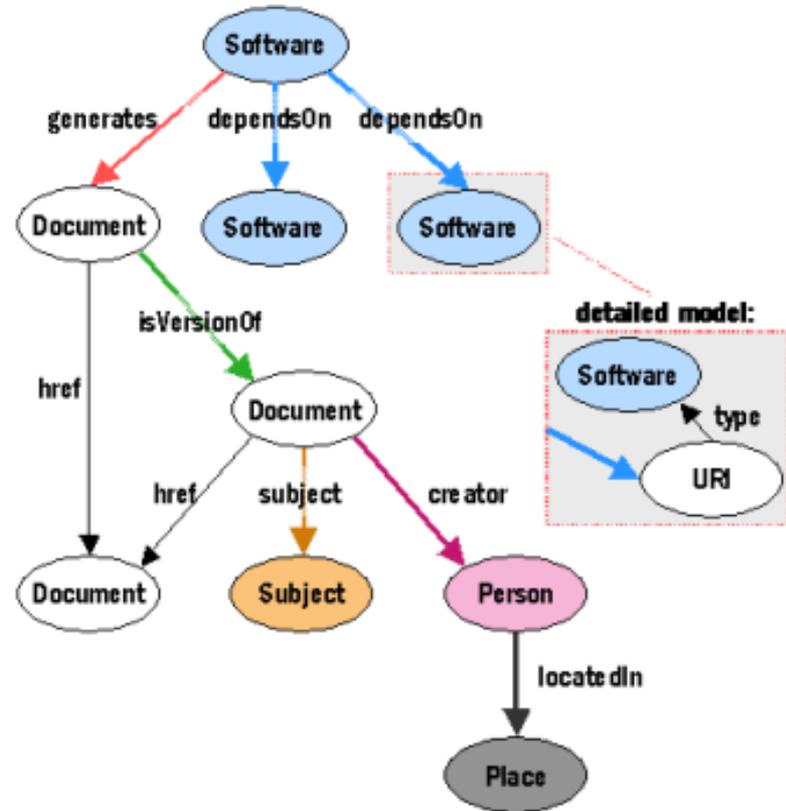
Valentina Presutti

Aldo Gangemi, Andrea Nuzzolese,
Diego Reforgiato, Sergio Consoli
STLab-ISTC, CNR, Italy

The current Web vs the Semantic Web (Marja-Riitta Koivunen and Eric Miller 2001)



a) Current Web



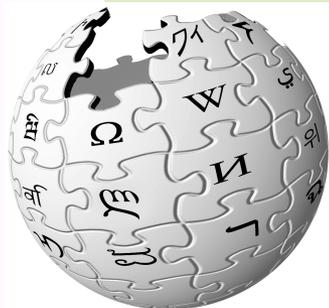
b) Semantic Web

*resources (web documents)
and links*

*Resources and links
that can have explicit types*

Hyperlinks

Hyperlinks are *pragmatic traces* of semantic relations between two entities



“McCarthy often commented on world affairs on the [Usenet forums](#)”

John McCarthy



John McCarthy at a conference in 2006

Born	September 4, 1927 Boston, Massachusetts, U.S.
Died	October 24, 2011 (aged 84) Stanford, California, U.S.
Residence	United States
Nationality	American
Fields	Computer technology
Institutions	Stanford University; Massachusetts Institute of Technology; Dartmouth College; Princeton University
Alma mater	Princeton University; California Institute of Technology

Texts surrounding hyperlinks are *linguistic traces* of semantic relations between two entities

Pragmatic trace: presence of a hyperlink in a web page

John McCarthy (computer scientist)

From Wikipedia, the free encyclopedia

John McCarthy (September 4, 1927 – October 24, 2011)^[1] was an American [computer scientist](#) and [cognitive scientist](#). McCarthy was one of the founders of the discipline of artificial intelligence.^[2] He coined the term "[artificial intelligence](#)" (AI), developed the [Lisp programming language](#) family, significantly influenced the design of the [ALGOL programming language](#), popularized [timesharing](#), and was very influential in the early development of AI.

McCarthy received many accolades and honors, such as the [Turing Award](#) for his contributions to the topic of AI, the United States [National Medal of Science](#), and the [Kyoto Prize](#).

Contents [hide]

- [1 Personal life and education](#)
- [2 Career in computer science](#)
- [3 Awards and honors](#)
- [4 Major publications](#)
- [5 See also](#)
- [6 References](#)
- [7 Further reading](#)
- [8 External links](#)

Personal life and education

John McCarthy was born in Boston, Massachusetts, a Lithuanian Jewish immigrant family.

John McCarthy



John McCarthy at a conference in 2006

Born	September 4, 1927 Boston, Massachusetts, U.S.
Died	October 24, 2011 (aged 84) Stanford, California, U.S.
Residence	United States
Nationality	American
Fields	Computer technology

He coined the term "[artificial intelligence](#)" (AI), developed the [Lisp programming language](#) family, significantly influenced the design of the [ALGOL programming language](#), popularized timesharing, and was very influential in the early development of AI.



Pragmatic trace: presence of a hyperlink in a web page

John McCarthy (computer scientist)

From Wikipedia, the free encyclopedia

John McCarthy (September 4, 1927 – October 24, 2011)^[1] was an American [computer scientist](#) and [cognitive scientist](#). McCarthy was one of the founders of the discipline of artificial intelligence.^[2] He coined the term "[artificial intelligence](#)" (AI), developed the [Lisp programming language](#) family, significantly influenced the design of the [ALGOL programming language](#), popularized [timesharing](#), and was very influential in the early development of AI.

McCarthy received many accolades and honors, such as the [Turing Award](#) for his contributions to the topic of AI, the United States [National Medal of Science](#), and the [Kyoto Prize](#).

Contents [hide]

- [1 Personal life and education](#)
- [2 Career in computer science](#)
- [3 Awards and honors](#)
- [4 Major publications](#)
- [5 See also](#)
- [6 References](#)
- [7 Further reading](#)
- [8 External links](#)

Personal life and education

John McCarthy was born in Boston, Massachusetts, a Lithuanian Jewish immigrant family.

John McCarthy



John McCarthy at a conference in 2006

Born	September 4, 1927 Boston, Massachusetts, U.S.
Died	October 24, 2011 (aged 84) Stanford, California, U.S.
Residence	United States
Nationality	American
Fields	Computer technology

He coined the term "[artificial intelligence](#)" (AI), developed the [Lisp programming language](#) family, significantly influenced the design of the [ALGOL programming language](#), popularized timesharing, and was very influential in the early development of AI.



Linguistic trace: text surrounding a hyperlink

John McCarthy

?

He coined the term "[artificial intelligence](#)" (AI), developed the [Lisp programming language](#) family, significantly influenced the design of the [ALGOL programming language](#), popularized timesharing, and was very influential in the early development of AI.

Linguistic trace: text surrounding a hyperlink

John McCarthy

dbpedia:John_McCarthy dbpo:wikiPageWikiLink dbpedia:Lisp

(DBpedia)

dbpo:wikiPageWikiLink

?

He coined the term "[artificial intelligence](#)" (AI), developed the [Lisp programming language](#) family, significantly influenced the design of the [ALGOL programming language](#), popularized timesharing, and was very influential in the early development of AI.

Linguistic trace: text surrounding a hyperlink

John McCarthy

dbpedia:John_McCarthy myont:developed dbpedia:Lisp

?

(Legalo)

legalo:developProgrammingLanguage

He coined the term "[artificial intelligence](#)" (AI), developed the [Lisp programming language](#) family, significantly influenced the design of the [ALGOL programming language](#), popularized timesharing, and was very influential in the early development of AI.

Linguistic trace: text surrounding a hyperlink

John MCCarthy

dbpedia:John_McCarthy myont:developed dbpedia:Lisp

?

(Legalo)

legalo:developProgrammingLanguage

rdfs:subPropertyOf <http://swrc.ontoware.org/ontology#develops>

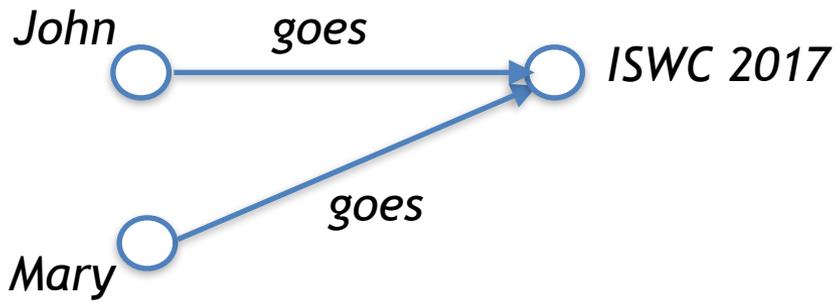
He coined the term "[artificial intelligence](#)" (AI), developed the [Lisp programming language](#) family, significantly influenced the design of the [ALGOL programming language](#), popularized timesharing, and was very influential in the early development of AI.

John and Mary went to ISWC 2017, together

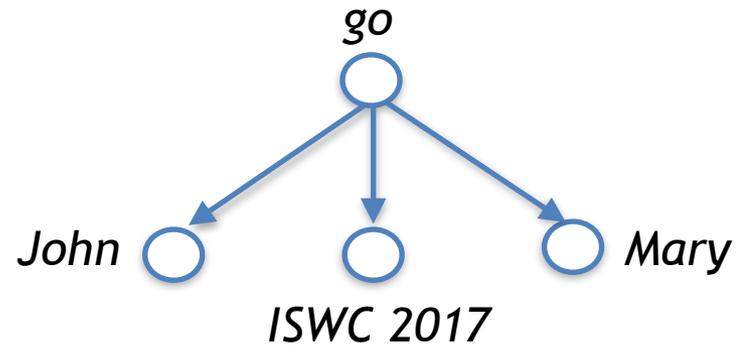
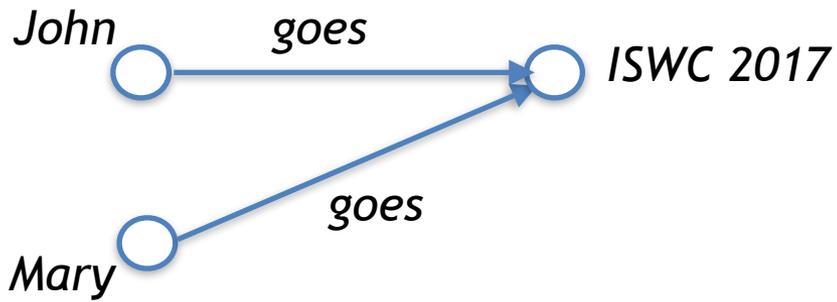
John and Mary went to ISWC 2017, together



John and Mary went to ISWC 2017, together



John and Mary went to ISWC 2017, together



however

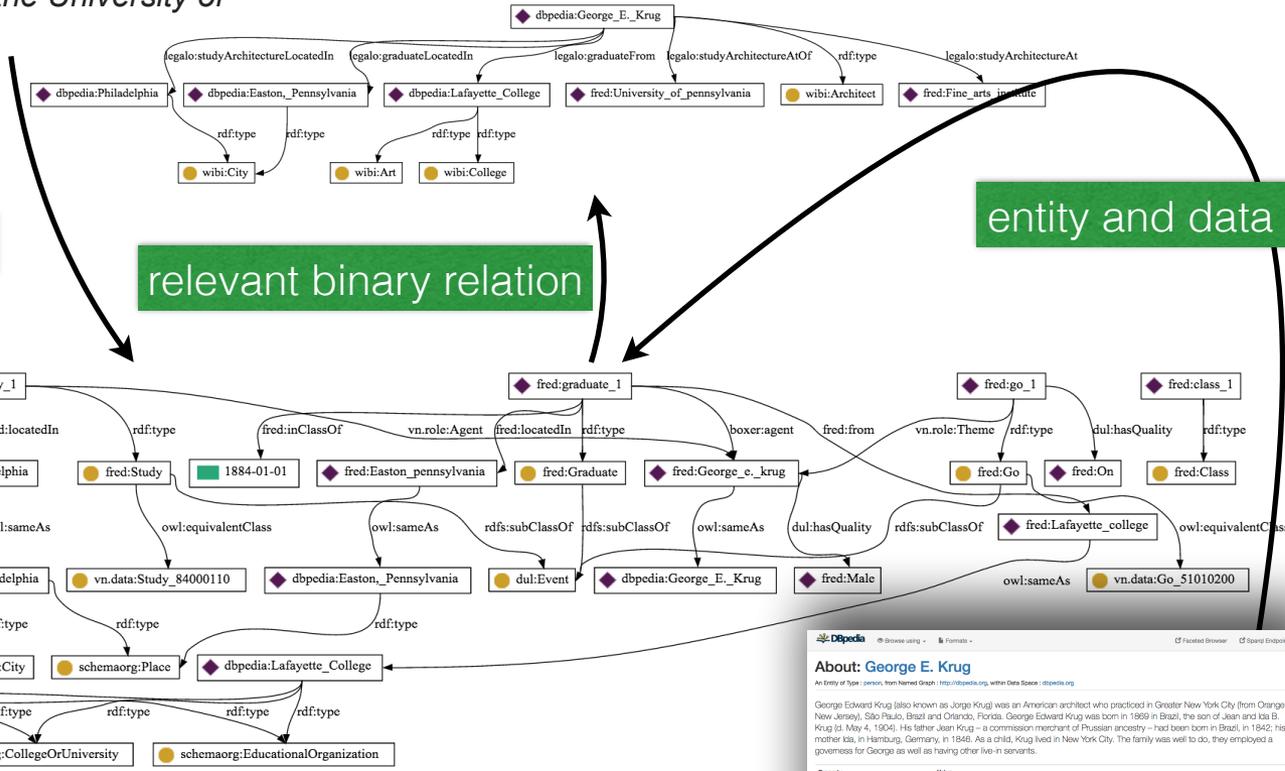
- binary relations are good for interaction
- querying: developers don't want to struggle with n-ary relations
- robots: to answer questions they need strategies for *de-reifying* n-ary relations
 - and generate language out of them
- similar intuition in more recent FrameBase ReDer rules

George E. Krug graduated from Lafayette College in Easton, Pennsylvania, in the class of 1884. He went on to study architecture in Philadelphia, at the Fine Arts Institute of the University of Pennsylvania.

knowledge extraction

relevant binary relation

entity and data linking



data enrichment

About: George E. Krug
 An Entity of type: person, from named graph: http://dbpedia.org, within Data Source: dbpedia

George Edward Krug (also known as Jorge Krug) was an American architect who practiced in Greater New York City (from Orange, New Jersey), São Paulo, Brazil and Orlando, Florida. George Edward Krug was born in 1884 in Brazil, the son of Jean and Ida B. Krug (d. May 4, 1904). His father, Jean Krug – a commission merchant of Prussian ancestry – had been born in Brazil, in 1842; his mother, Ida, in Hamburg, Germany, in 1846. As a child, Krug lived in New York City. The family was well to do; they employed a governess for George as well as having other live-in servants. George Krug graduated from Lafayette College in Easton, Pennsylvania, in the class of 1884. He went on to study architecture in Philadelphia, at the Fine Arts Institute of the University of Pennsylvania. Thereafter he spent most of his decade in São Paulo, Brazil, starting in 1888. There, he collaborated with other architects including Maximilian Ernst Herzl, Pauler Arns Mariani (1888–1866) was his niece. Upon his return to the United States, Krug maintained an architectural practice with offices in Brooklyn and in Orange, NJ, designing buildings in the greater New York City area. Krug was the architect of the Hyde Park Club houses and many residential properties in Hyde Park, NJ. He was also one of the select group of architects who designed buildings for the general suburban community of "The Linton Moore" in Hyde Park, NJ. This development consisted of architect-designed homes in various styles: Queen Anne houses, Burglows, Four-squares, and Colonial Revival houses with embellishments typical of the Dutchman style vernacular, which emphasized the use of the louver of pilasters who showed pride in their address. By 1913 George E. Krug relocated to Florida. He is listed as an architect in Orlando, Florida, in the report of the Florida Office of the Secretary of State of that year. Krug designed numerous great houses and mansions in the Lake Wales area, wherever such residences conveyed academic qualities, while having unique characteristics. Fine examples of Krug's Lake Wales style are Forest, Georgian, Greek, and American. Later, he worked through downtown Orlando, particularly in The Lake Wales and Lake Wales-Davenport-Deshaugh historic districts, and the vicinity of Winter Park, brick houses, some fluted columns, Gothic Revival and Federal style mansions and town row, and many other mansions of the Krug design. Noteworthy houses are a single story to their site plan was also a common theme of his. As such, Krug was among those that a dozen architects in Orlando at that time. The others include Frank B. Brooks, Fred L. Hess, David Froy Mary B. Krug, Howard B. Reynolds, Frederick B. Tivolis, Froy and Roberts Jay Auman Froy and Isabel Reardon and Henry B. Turner. George E. Krug and his wife Clara A. Krug were associated with the St. John's Episcopal Church in Apopka, Florida, where Mrs. Krug was for some time the superintendent of the church school. George E. Krug died in Orlando, Florida in 1959. [\[w\]](#)

Property	Value
dbpedia:birth	<ul style="list-style-type: none"> George Edward Krug (also known as Jorge Krug) was an American architect who practiced in Greater New York City (from Orange, New Jersey), São Paulo, Brazil and Orlando, Florida. George Edward Krug was born in 1884 in Brazil, the son of Jean and Ida B. Krug (d. May 4, 1904). His father, Jean Krug – a commission merchant of Prussian ancestry – had been born in Brazil, in 1842; his mother, Ida, in Hamburg, Germany, in 1846. As a child, Krug lived in New York City. The family was well to do; they employed a governess for George as well as having other live-in servants. George Krug graduated from Lafayette College in Easton, Pennsylvania, in the class of 1884. He went on to study architecture in Philadelphia, at the Fine Arts Institute of the University of Pennsylvania. Thereafter he spent most of his decade in São Paulo, Brazil, starting in 1888. There, he collaborated with other architects including Maximilian Ernst Herzl, Pauler Arns Mariani (1888–1866) was his niece. Upon his return to the United States, Krug maintained an architectural practice with offices in Brooklyn and in Orange, NJ, designing buildings in the greater New York City area. Krug was the architect of the Hyde Park Club houses and many residential properties in Hyde Park, NJ. He was also one of the select group of architects who designed buildings for the general suburban community of "The Linton Moore" in Hyde Park, NJ. This development consisted of architect-designed homes in various styles: Queen Anne houses, Burglows, Four-squares, and Colonial Revival houses with embellishments typical of the Dutchman style vernacular, which emphasized the use of the louver of pilasters who showed pride in their address. By 1913 George E. Krug relocated to Florida. He is listed as an architect in Orlando, Florida, in the report of the Florida Office of the Secretary of State of that year. Krug designed numerous great houses and mansions in the Lake Wales area, wherever such residences conveyed academic qualities, while having unique characteristics. Fine examples of Krug's Lake Wales style are Forest, Georgian, Greek, and American. Later, he worked through downtown Orlando, particularly in The Lake Wales and Lake Wales-Davenport-Deshaugh historic districts, and the vicinity of Winter Park, brick houses, some fluted columns, Gothic Revival and Federal style mansions and town row, and many other mansions of the Krug design. Noteworthy houses are a single story to their site plan was also a common theme of his. As such, Krug was among those that a dozen architects in Orlando at that time. The others include Frank B. Brooks, Fred L. Hess, David Froy Mary B. Krug, Howard B. Reynolds, Frederick B. Tivolis, Froy and Roberts Jay Auman Froy and Isabel Reardon and Henry B. Turner. George E. Krug and his wife Clara A. Krug were associated with the St. John's Episcopal Church in Apopka, Florida, where Mrs. Krug was for some time the superintendent of the church school. George E. Krug died in Orlando, Florida in 1959. [w]
dbpedia:Date	1884-1-1
dbpedia:Date	1959-1-1
dbpedia:PageID	1973876 (outgoing)
dbpedia:PageRevisionsID	7366160 (incoming)
db:description	American architect [w]
db:subject	<ul style="list-style-type: none"> db:Brazilian_architect db:Brazilian_people_of_German_descent db:People_from_Orlando,Florida db:1884_births db:1959_deaths db:American_architects db:Architecture_in_São_Paulo,_Florida
http://owl.org/Inq/ontology/gdm/sym	db:Architect
db:type	<ul style="list-style-type: none"> db:Person http://dbpedia.org/ontology/1884/01/01 http://dbpedia.org/ontology/1959/01/01 http://dbpedia.org/ontology/1959/01/01



Relation extraction

- Distant supervision: maps textual sentences to linked data triples
 - no support for n -ary relation
 - bound to already defined relation types

“Paris is the capital city of France”

dbpedia:Paris dbpedia-owl:country dbpedia:France

“John Stigall received a Bachelor of arts”

?

- Open Information Extraction: creates resources of triples composed of text fragments (subject, relational phrase, object)
 - not directly reusable as linked data

“John Stigall received a Bachelor of arts”

(John Stigall, received, a Bachelor of arts)

Open Challenges:

- > discovery of *new* relations (both facts and relation types)
- > methods for automatically formalising discovered knowledge
- > usable label generation for new discovered relations (abstractive summarisation)

OIE vs OKE

“John Stigall received a Bachelor of arts from the State University of New York at Cortland “

<i>Subject</i>	<i>Predicate</i>	<i>Object</i>	<i>Approach</i>
<i>John Stigall</i>	<i>received</i>	<i>a Bachelor of arts</i>	<i>extractive</i>
<i>John Stigall</i>	<i>received</i>	<i>from the State University of New York at Cortland</i>	<i>extractive</i>
<i>dbpedia:John_Stigall</i>	<i>myprop:receives AcademicDegree</i>	<i>dbpedia:Bachelor_of_Arts</i>	<i>abstractive</i>
<i>dbpedia:John_Stigall</i>	<i>myprop:receives AcademicDegree From</i>	<i>dbpedia:State_University_of_New_York</i>	<i>abstractive</i>

OIE vs OKE

“John Stigall received a Bachelor of arts from the State University of New York at Cortland “

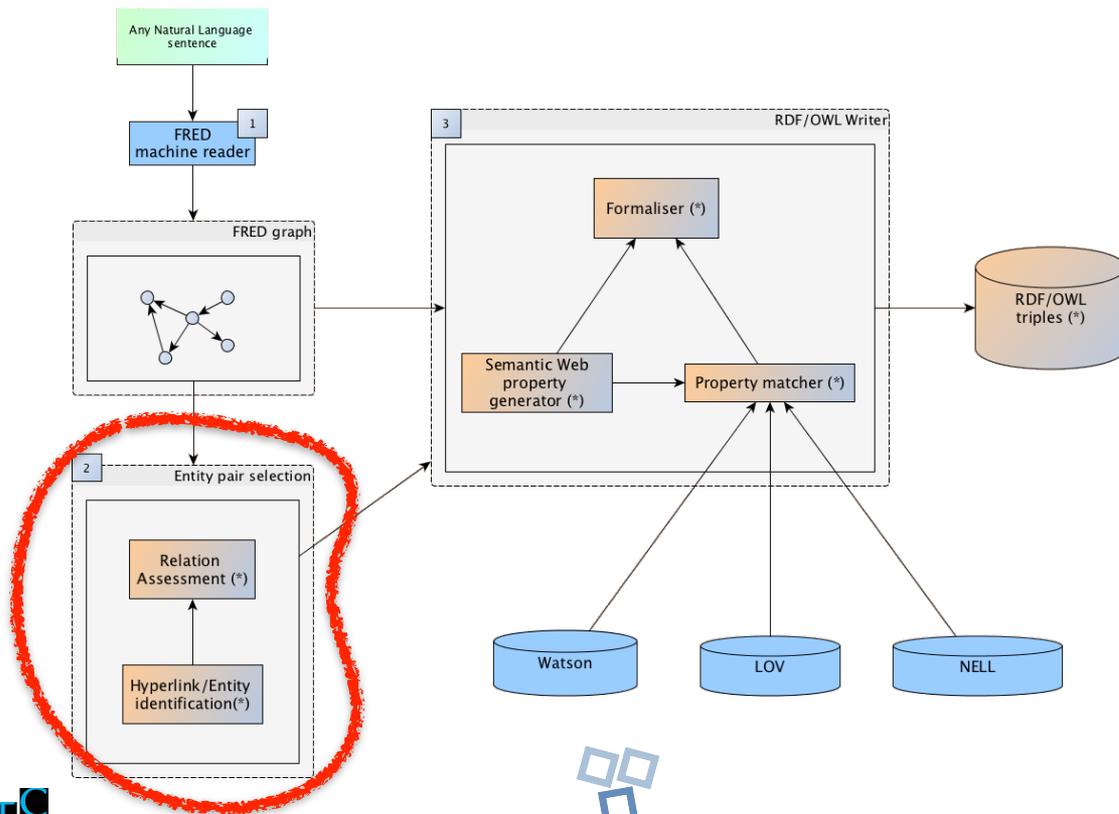
<i>Subject</i>	<i>Predicate</i>	<i>Object</i>	<i>Approach</i>
<i>John Stigall</i>	<i>received</i>	<i>a Bachelor of arts</i>	<i>extractive</i>
<i>John Stigall</i>	<i>received</i>	<i>from the State University of New York at Cortland</i>	<i>extractive</i>
<i>myprop:receive_1</i>	<i>role:Agent</i>	<i>dbpedia:John_Stigall</i>	<i>abstractive</i>
<i>myprop:receive_1</i>	<i>role:Theme</i>	<i>dbpedia:Bachelor_of_Arts</i>	<i>abstractive</i>
<i>myprop:receive_1</i>	<i>role:Source</i>	<i>dbpedia:Univ_NYC_Cortland</i>	<i>abstractive</i>

Let's see how it works in detail...



Relevant relations assessment

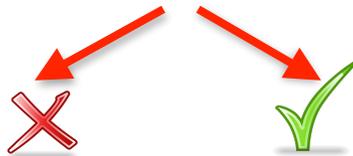
“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



Relevant Relation Assessment

“Joey Foster Ellis has published on The New York Times, and The Wall Street Journal.”

? $\varphi_s(e_{subj}, e_{obj})_i$?

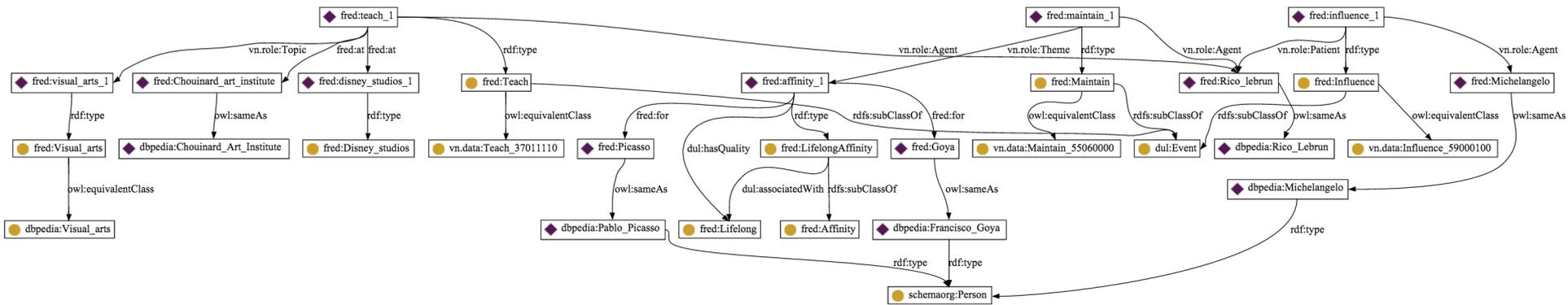


$\varphi_s(\text{Joey Foster Ellis}, \text{The New York Times})$ ✓

$\varphi_s(\text{The Wall Street Journal}, \text{The New York Times})$ ✗

Relevant relation assessment

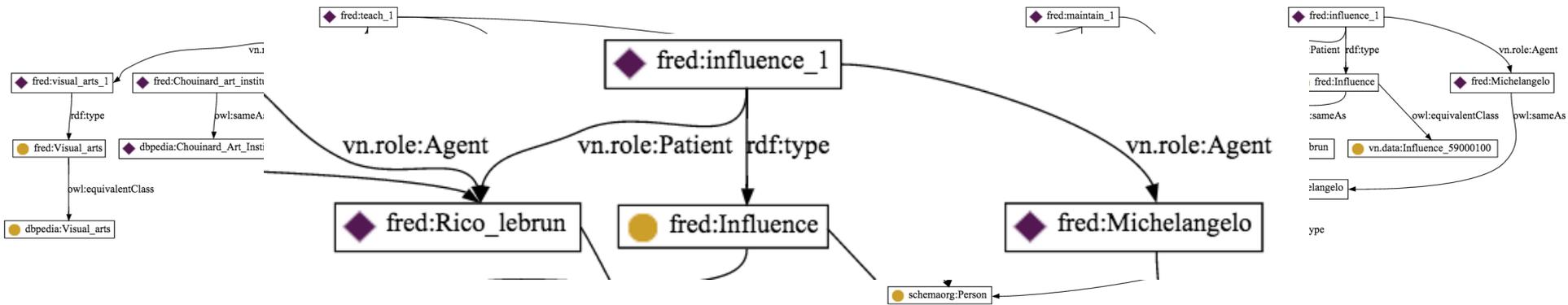
“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



Necessary condition: $\varphi_s(e_{\text{subj}}, e_{\text{obj}}) \Rightarrow \exists P(v_{\text{subj}}, v_{\text{obj}})$

Relevant relation assessment

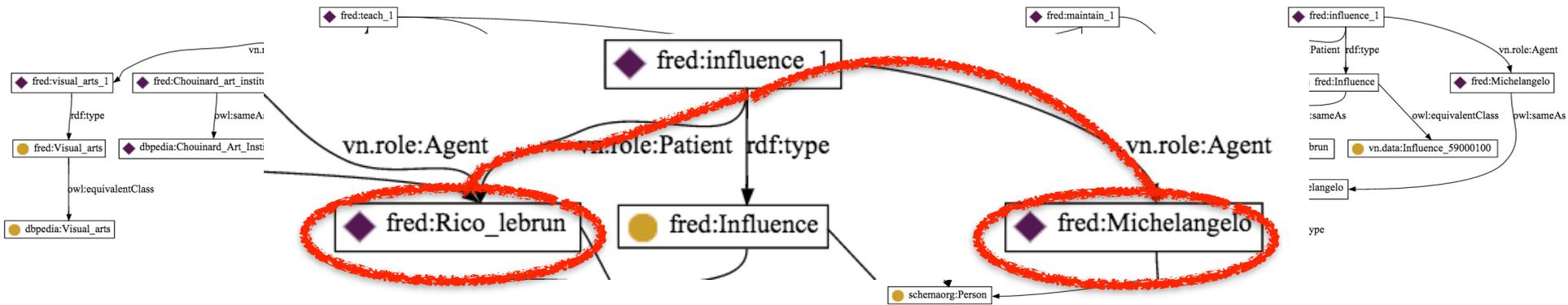
“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



Necessary condition: $\varphi_s(e_{\text{subj}}, e_{\text{obj}}) \Rightarrow \exists P(v_{\text{subj}}, v_{\text{obj}})$

Relevant relation assessment

“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”

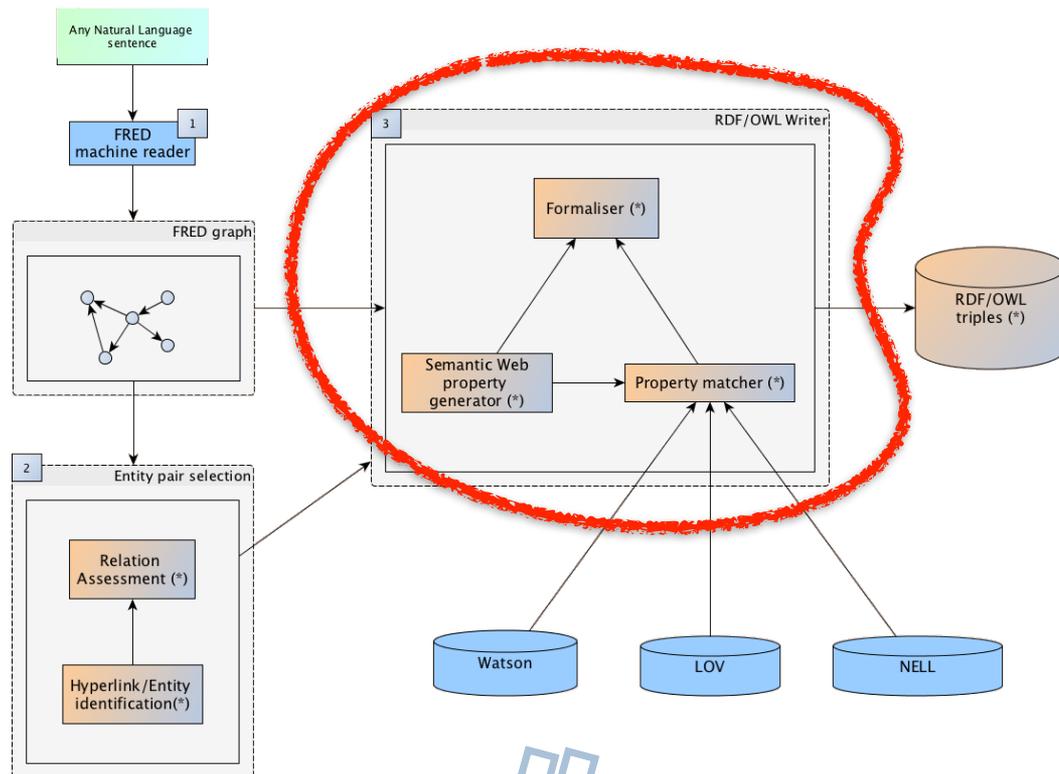


Necessary condition: $\varphi_s(e_{\text{subj}}, e_{\text{obj}}) \Rightarrow \exists P(v_{\text{subj}}, v_{\text{obj}})$

$\varphi_s(\text{Rico Lebrun}, \text{Michelangelo}) \Rightarrow \exists P(\text{fred:Rico_Lebrun}, \text{fred:Michelangelo})$

Semantic Web triples and properties generation

“[Rico Lebrun](#) taught [visual arts](#) at the [Chouinard Art Institute](#) and at the [Disney Studios](#). He was influenced by [Michelangelo](#) and maintained a lifelong affinity for [Goya](#) and [Picasso](#).”



Usable predicate generation

“Joey Foster Ellis has published on The New York Times, and The Wall Street Journal.”

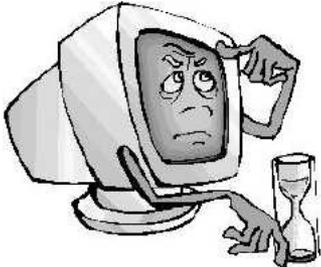
$\varphi_s(\text{Joey Foster Ellis, The New York Times})$



$$\Lambda \equiv \{\lambda_1, \dots, \lambda_n\}$$

labels for $\varphi_s(e_{subj}, e_{obj})_i$
defined for a LOD vocabulary

(has) published on
(has) published on journal
(is) author for
writes for

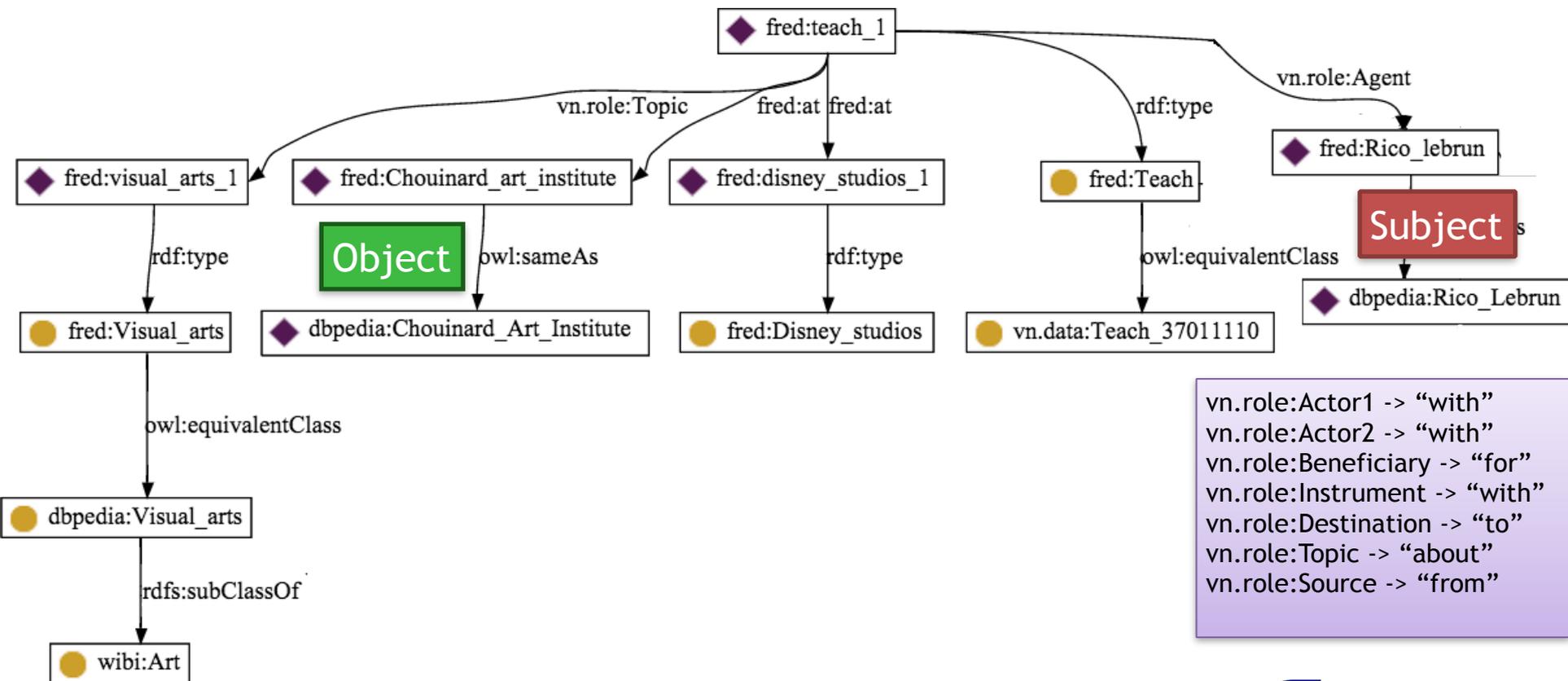


$$\lambda \sim \lambda_i, \lambda_i \in \Lambda$$

publish on

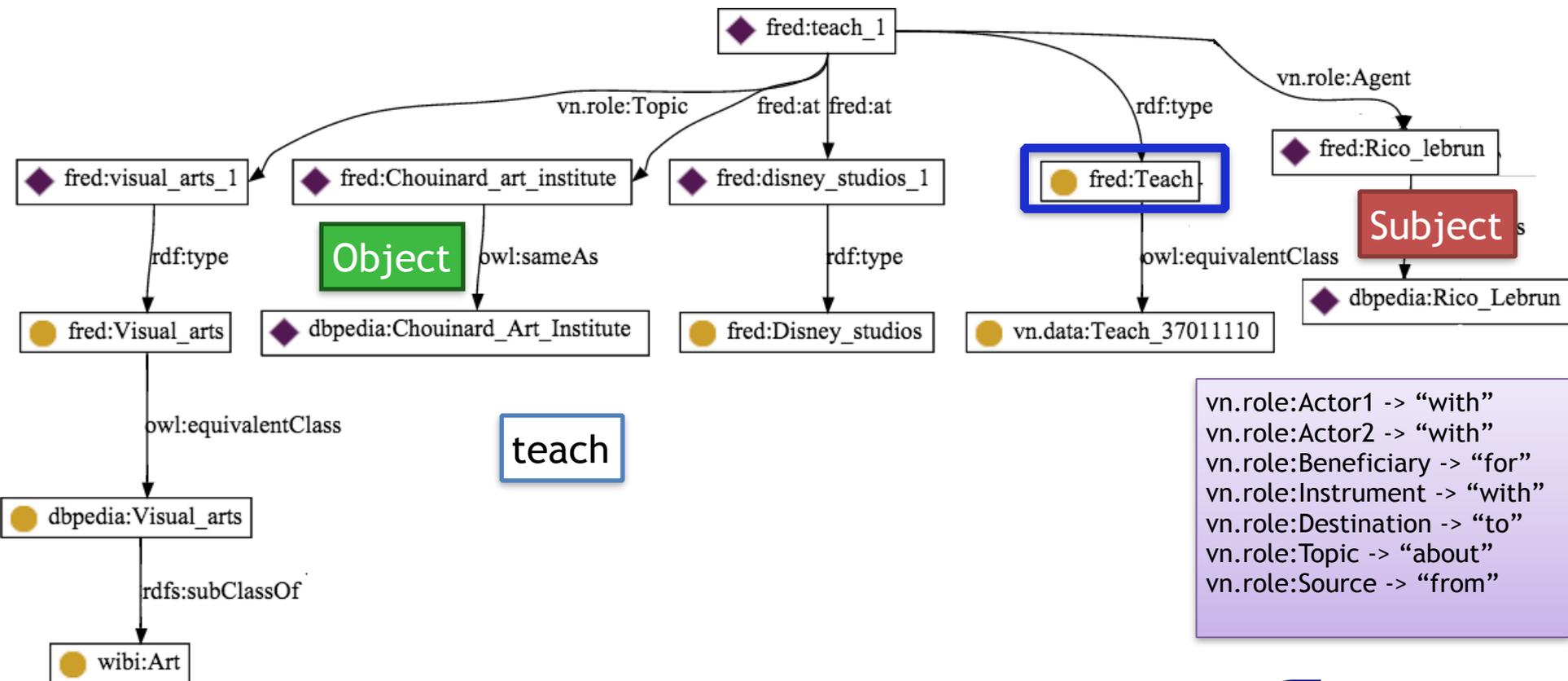
Semantic Web triples and properties generation

“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



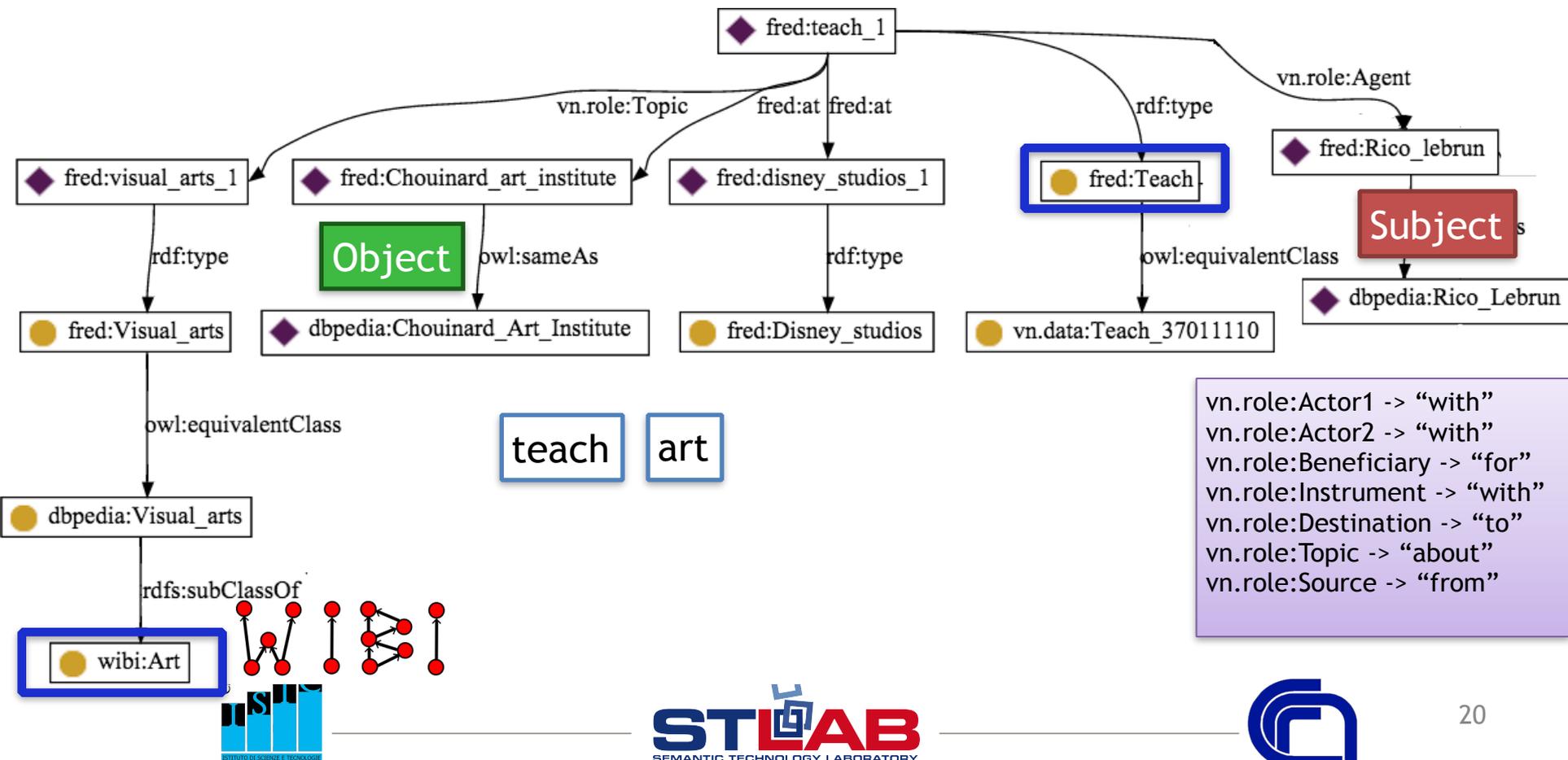
Semantic Web triples and properties generation

“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



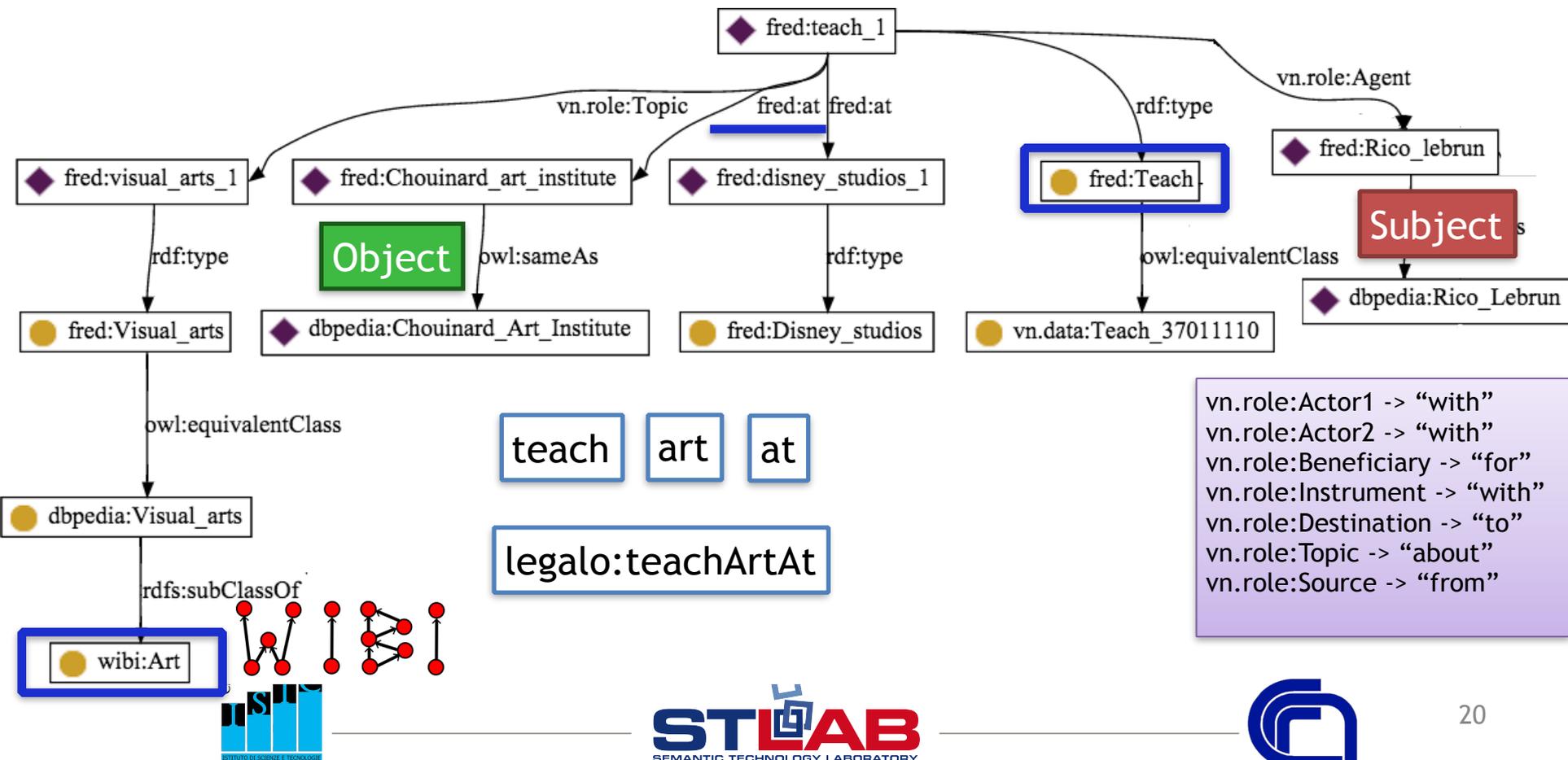
Semantic Web triples and properties generation

“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



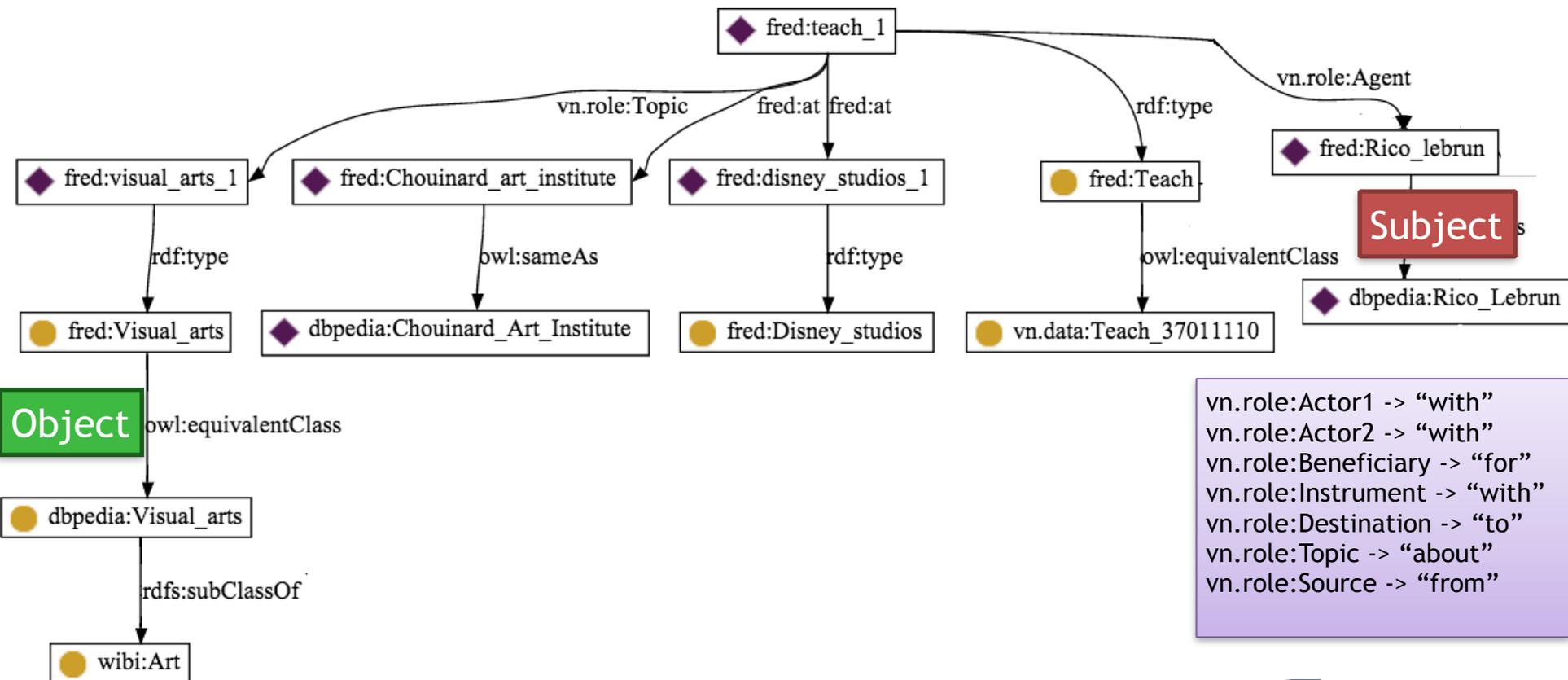
Semantic Web triples and properties generation

“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



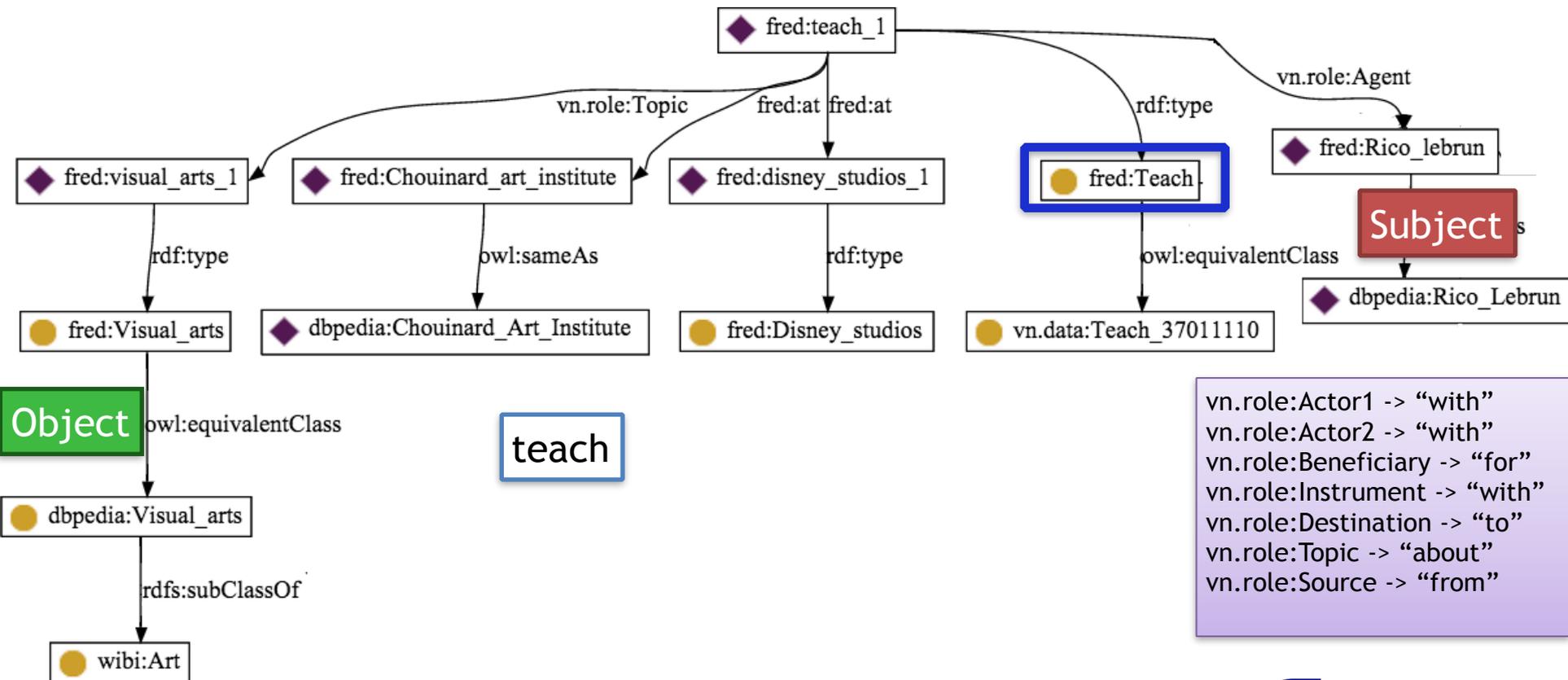
Semantic Web triples and properties generation

“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



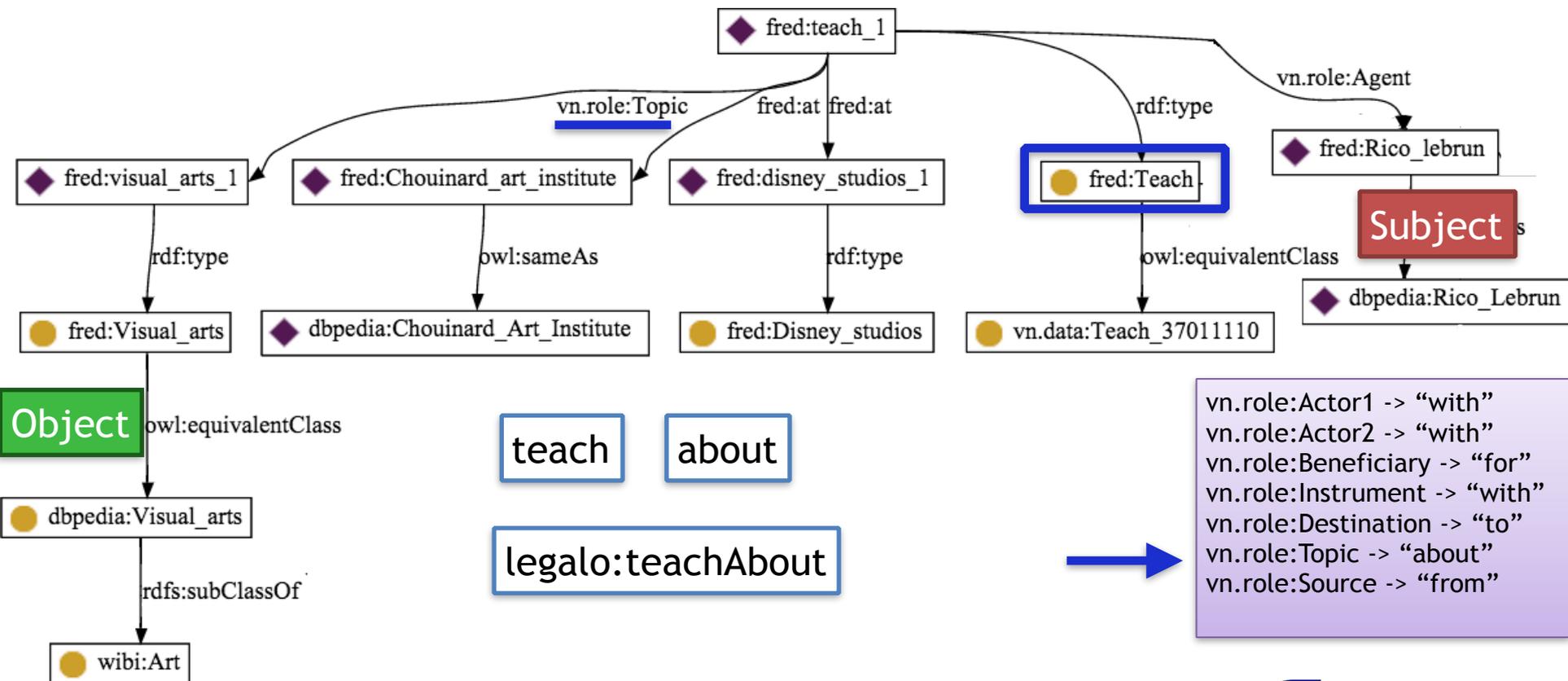
Semantic Web triples and properties generation

“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



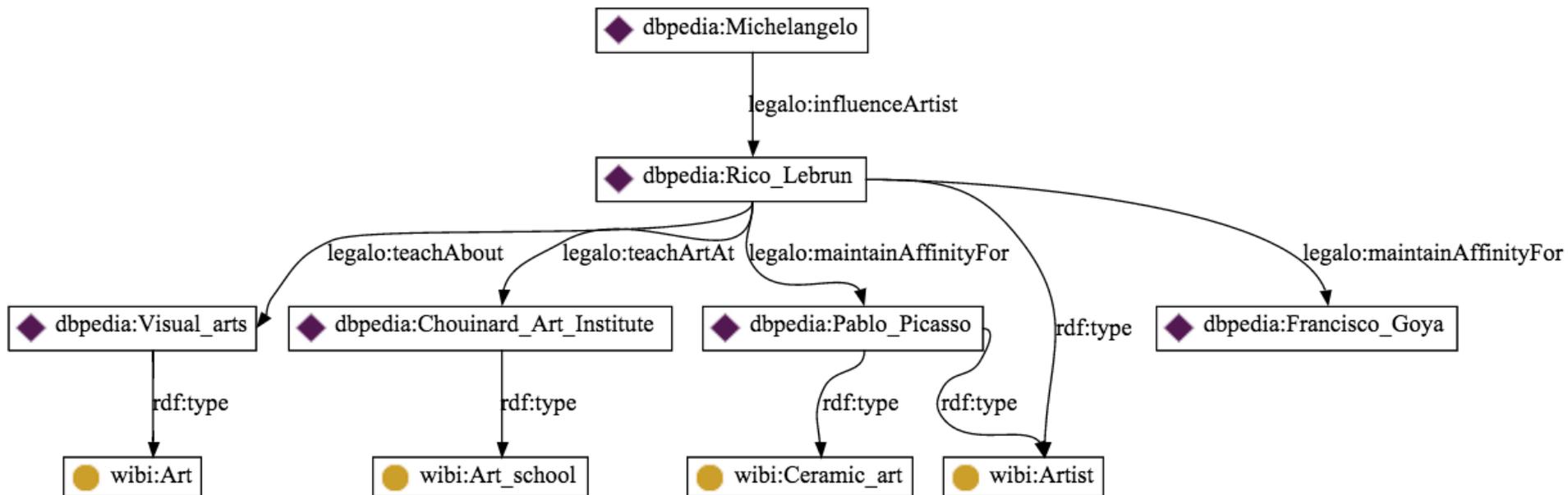
Semantic Web triples and properties generation

“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



Semantic Web triples and properties generation

“[Rico Lebrun](#) taught [visual arts](#) at the [Chouinard Art Institute](#) and at the [Disney Studios](#). He was influenced by [Michelangelo](#) and maintained a lifelong affinity for [Goya](#) and [Picasso](#).”



Relation formalisation

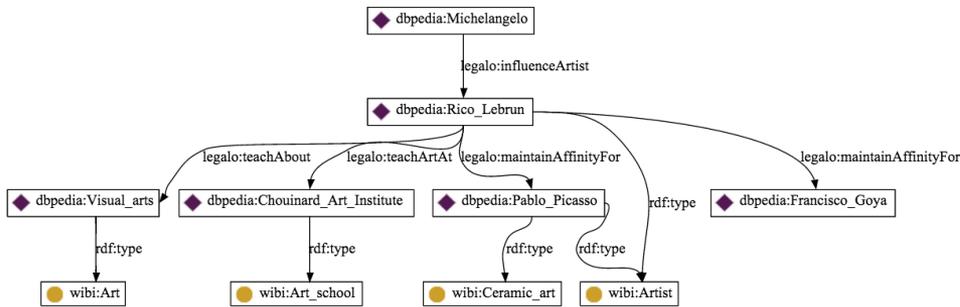
“Joey Foster Ellis has published on The New York Times, and The Wall Street Journal.”

$\varphi_s(\text{Joey Foster Ellis}, \text{The New York Times})$

```
dbpedia:Joey_Foster_Ellis  
  legalo:publishOn dbpedia:The_New_York_Times .
```

```
legalo:publishOn a owl:ObjectProperty ;  
  rdfs:range ... ;  
  rdfs:domain ... ;  
  ... .
```

Semantic Web triples and properties generation



“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”

```
dbpedia:Rico_Lebrun s:teachAbout dbpedia:Visual_arts .
```

```
s:teachAbout a owl:ObjectProperty ;
  rdfs:subPropertyOf fred:Teach ;
  rdfs:domain wibi:Artist ;
  rdfs:range wibi:Art ;
```

domain, range, subsumption

```
grounding:definedFromFormalRepresentation
```

```
  fred-graph:a6705cedbf9b53d10bbcdedaa3be9791da0a9e94 ;
```

```
grounding:derivedFromLinguisticEvidence s:linguisticEvidence ;
```

```
owl:propertyChainAxiom([ owl:inverseOf s:AgentTeach ] s:TopicTeach) .
```

linguistic and formal scope

```
_:b2 a alignment:Cell ;
```

```
  alignment:entity1 s:teachAbout ;
```

```
  alignment:entity2 <http://purl.org/vocab/aaiso/schema#teaches> ;
```

```
  alignment:measure "0.846"^^xsd:float ;
```

```
  alignment:relation "equivalence" .
```

alignment to existing LOD vocabularies



Evaluation

Legalo: Evaluation sample

C_{rel-extraction}: 5 datasets each dedicated to a specific relation*

```
{“pred”:  
“/people/person/education./education/education/institution”,  
“sub”：“/m/0g9zjv3”,  
“obj”：“/m/0ckrjh”,  
“evidences”:  
[{"url":"http://en.wikipedia.org/wiki/Dmitry_Chernyakov",  
“snippet”:  
“Dmitry was born in Moscow. In 1993 he graduated from Russian Academy of Theatre Arts as stage director. He started his career in the Russian Drama Theatre of Lithuania in Vilnius. Then he directed opera and drama in many major Russian cities: Moscow, Saint Petersburg, Novosibirsk, Omsk, Samara, Kazan and others. He usually creates scenic design and stage clothes for his plays.”}],  
“judgments”:  
[{"rater":“1264223381988340244”,“judgment”：“yes”},{“rater”：“6968726908160095830”,“judgment”：“yes”},  
{"rater”：“6063741883945424276”,“judgment”：“yes”},{“rater”：“5346153820624061638”,“judgment”：“yes”},  
{"rater”：“12022408018620867151”,“judgment”：“yes”}]}
```

**<https://code.google.com/p/relation-extraction-corpus/downloads/list>
corpus for relation extraction*

Legalo: Evaluation sample

$C_{\text{institution}}$: random sample of 130 snippets (one entity pair) from $C_{\text{rel-extraction}}$ for **“attending or graduating from an institution”**
Legalo produced always an output, either p or *“no relation”*

$C_{\text{education}}$: random sample of 130 snippets (one entity pair) from $C_{\text{rel-extraction}}$ for **“obtaining a degree of education”**
Legalo produced always an output, either p or *“no relation”*

C_{general} : random sample of 60 snippets from all $C_{\text{rel-extraction}}$ datasets then broken into 186 single sentences with at least one entity pair.
Legalo produced 867 results, 262 p and 605 *“no relation”*

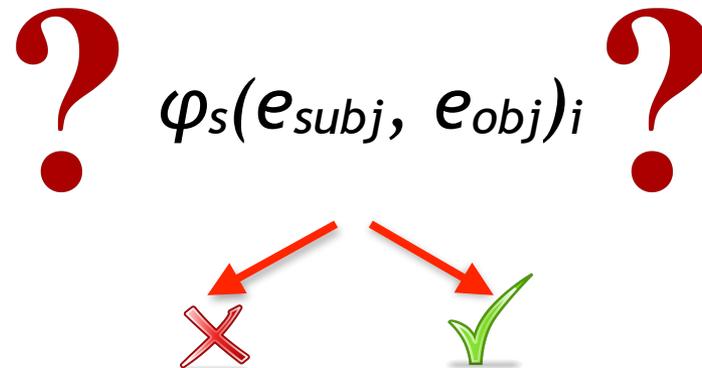


Hypothesis	Task	#workers
Hypothesis 1	Task 1,2	35
Hypothesis 2	Task 3 (institution)	10
Hypothesis 2	Task 3 (education)	18
Hypothesis 2	Task 4	19
Hypothesis 2	Task 5	12

Hypothesis 1: Relevant relation assessment

Legalo is able to assess if, given a sentence s , a relevant relation exists which holds between two entities, according to the content of s :

$$\exists \varphi. \varphi_s (e_{subj} , e_{obj})$$



Crowdsourcing tasks for Relevant Relation Assessment

Task 1:

Assessing if a sentence s is an evidence of a referenced relation (i.e. either “institution” or “education”) between two entities, mentioned in s .

Based on data from $C_{\text{institution}}$ and $C_{\text{education}}$, respectively

Task 2:

Assessing if a sentence s is an evidence of any relation between two given entities mentioned in s .

Based on data from C_{general}

At least 3 raters with $t > 0.70$

Answer could be “yes” or “no”

Evaluation results for Relevant Relation Assessment

Task	Relation	Precision	Recall	F-measure	Accuracy	Confidence
2	Any	0.83	0.92	0.87	0.82	0.82
1	Education	0.95	0.91	0.93	0.87	0.96
1	Institution	0.93	0.90	0.91	0.84	0.94

The confidence value expresses a weighted value for inter-rater agreement, by rater trust scores

Hypothesis 2: Usable predicate generation

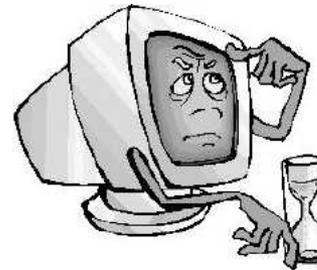
Legalo is able to generate a usable predicate λ for a relevant relation φ_s between to entities, expressed in a sentence s : given λ , a label generated by Legalo for φ_s , and λ_i a label generated by a human for φ_s the following holds:

$$\lambda \sim \lambda_i, \lambda_i \in \Lambda$$



$$\Lambda \equiv \{\lambda_1, \dots, \lambda_n\}$$

*labels for $\varphi_s(e_{subj}, e_{obj})_i$
defined for a LOD vocabulary*



$$\lambda \sim \lambda_i, \lambda_i \in \Lambda$$

Crowdsourcing tasks for Usable predicate generation

Task 3:

Judging if λ generated by a machine expresses is a good summarisation of a specific relation (i.e. either “institution” or “education”) between two given entities mentioned in s , according to the evidence provided by s .

Based on data from $C_{\text{institution}}$ and $C_{\text{education}}$, respectively

Task 4:

judging if a predicate λ is a good summarisation of a relation expressed by the content of s , between two given entities mentioned in s , according to the evidence provided by s .

Based on data from C_{general}

At least 3 raters with $t > 0.70$

Answer could be Agree, Partly Agree, Disagree

Evaluation results for Usable Predicate Generation

Task	Relation	Precision	Recall	F-measure	Accuracy	Confidence
3	Education	0.92	0.91	0.91	0.85	0.80
3	Institution	0.65	0.91	0.76	0.62	0.59
3 (high confidence only)	Institution	0.74	0.89	0.81	0.68	0.71
4	Any	0.68	0.90	0.78	0.71	0.64
4 (high confidence only)	Any	0.73	0.87	0.80	0.75	0.76

The confidence value expresses a weighted value for inter-rater agreement, by rater trust scores

Agree = 1, Partly Agree = 0.5, Disagree = 0

Crowdsourcing tasks for Usable predicate generation

Task 5:

Creating a phrase λ that summarises the relation expressed by the content of s , between two given entities mentioned in s , according to the evidence provided by s .

Based on data from C_{general}

At least 3 raters with $t > 0.60$

Answer was open

Evaluation results for Usable Predicate Generation

- Similarity score between human created $\{\lambda_i\}$ and Legalo λ , for a $\varphi_s(e_{\text{subj}}, e_{\text{obj}})_i$
- Jaccard distance measure (string similarity)
- Semantic similarity measure based on the SimLibrary framework*

<i>Task</i>	<i>Relation</i>	<i>Jaccard [0,1]</i>	<i>SimLibrary [0,1]</i>	<i>Confidence</i>
5	Any	0.63	0.80	0.59

* G. Pirró and J. Euzenat. A feature and information theoretic framework for semantic similarity and relatedness. ISWC2010

Conclusion and next steps

- a method to extract the meaning of hyperlinks
- and express it as de-reification of n-ary relations
- ability to identify relevant binary relations
- ability to generate usable labels
- useful for interaction tasks
 - e.g. querying, language generation

Next

- reconciliation strategies
- robot's question answering

Thanks for your attention

Questions?

A hand holding a white marker is pointing to the end of the word 'Questions?' written on a chalkboard. The hand is wearing a blue sleeve.

<http://wit.istc.cnr.it/stlab-tools/legalo>

References

Valentina Presutti, Andrea Giovanni Nuzzolese, Sergio Consoli, Aldo Gangemi, Diego Reforgiato Recupero: *From hyperlinks to Semantic Web properties using Open Knowledge Extraction*. *Semantic Web* 7(4): 351-378 (2016)

Aldo Gangemi, Valentina Presutti, Diego Reforgiato Recupero, Andrea Giovanni Nuzzolese, Francesco Draicchio, Misael Mongiovì: *Semantic Web Machine Reading with FRED*. *Semantic Web* 8(6): 873-893 (2017)

<http://wit.istc.cnr.it/stlab-tools/legalo>

<http://wit.istc.cnr.it/stlab-tools/fred>

Natural language sentence

s

“Joey Foster Ellis has published on The New York Times, and The Wall Street Journal.”

Set of entity pairs

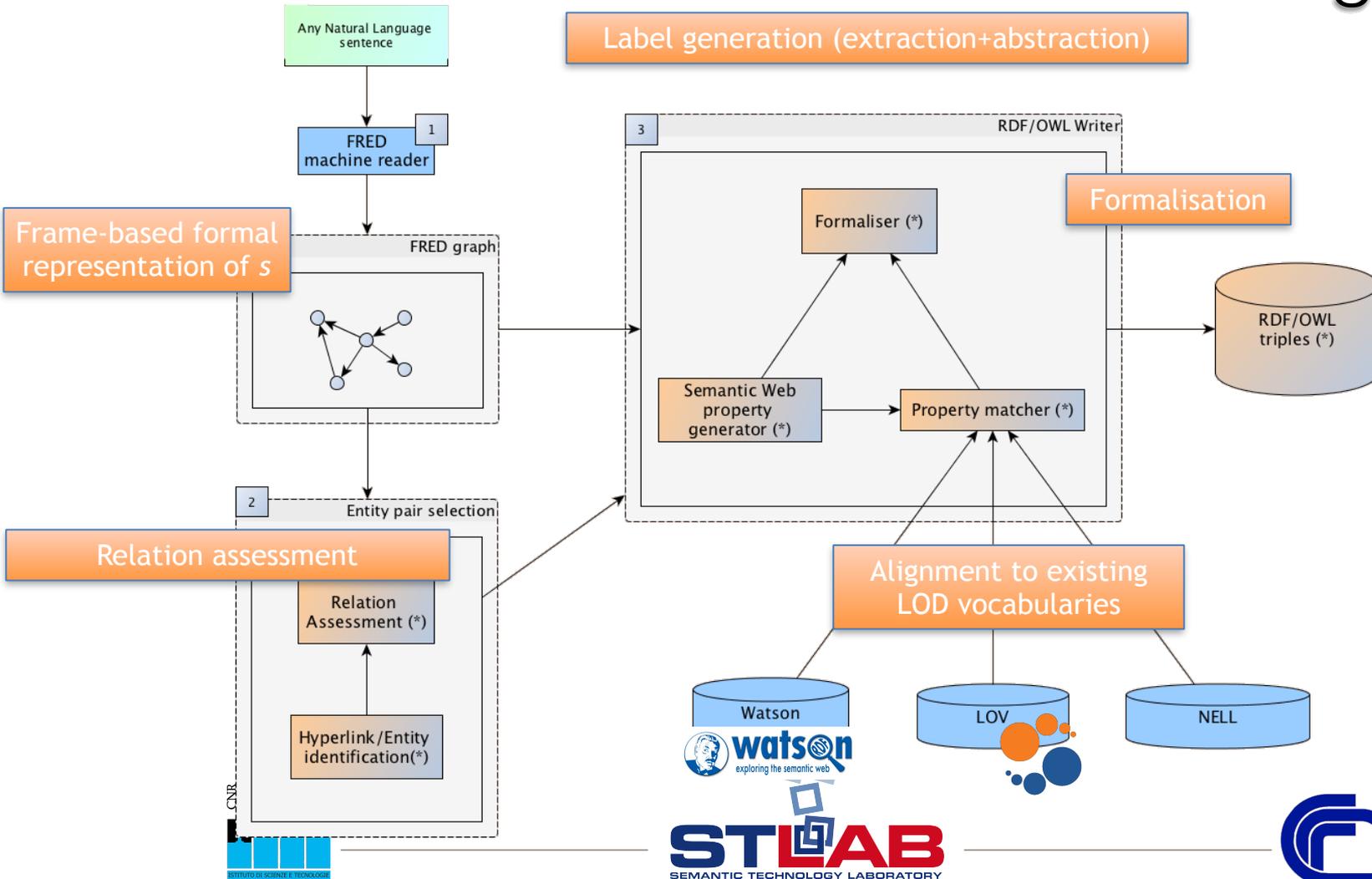
$\{(e_{subj}, e_{obj})_i\}$

Joey Foster Ellis, The New York Times
Joey Foster Ellis, The Wall Street Journal
The New York Times, The Wall Street Journal

Relevant relations evidenced by s

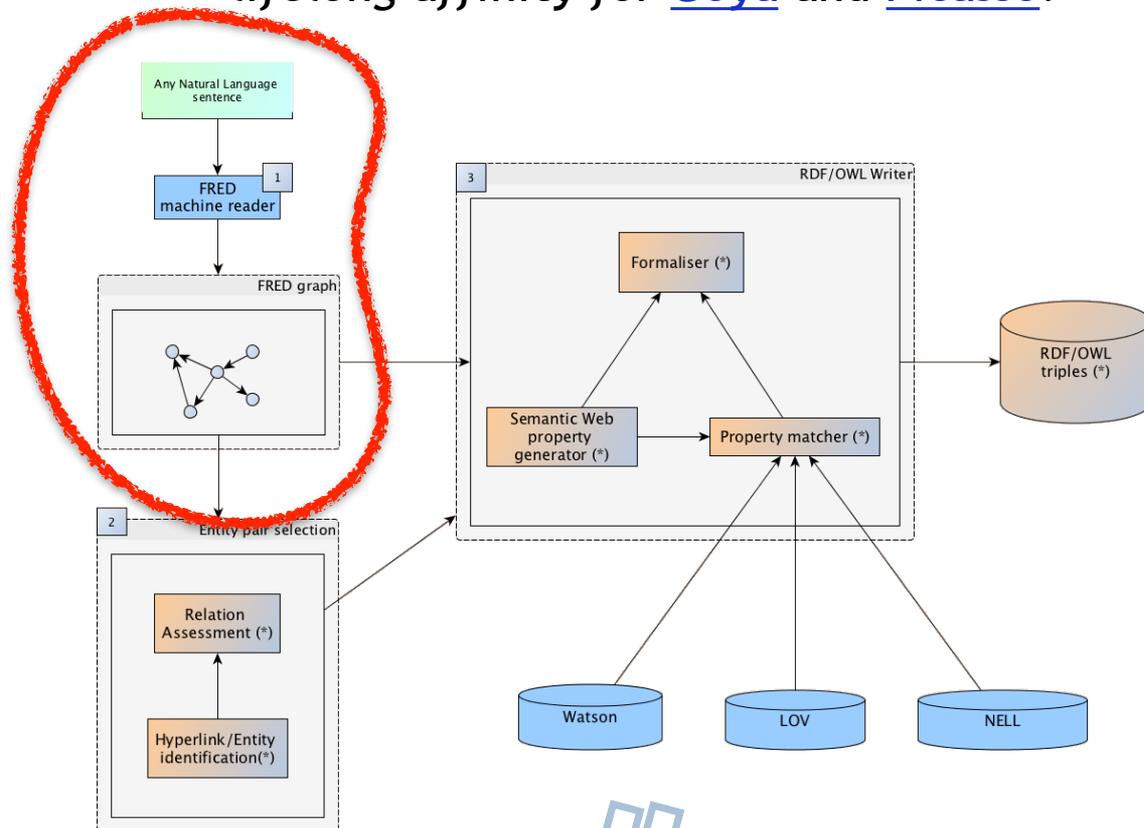
$\varphi_s(e_{subj}, e_{obj})_i$ $\varphi_s(\text{Joey Foster Ellis, The New York Times})$

Method and implementation



Frame-based formal representation

“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



Open Knowledge Extraction

- **unsupervised** -> no need of huge annotated corpora for training
- **open-domain** -> not bound to specific domains
- **abstractive** -> models the text, uses external knowledge resources, uses summarisation and language generation techniques
- **producing formalised output** -> linked data knowledge graph, (de-)reification of expressed relations

OIE vs OKE

“John Stigall received a Bachelor of arts from the State University of New York at Cortland “

<i>Subject</i>	<i>Predicate</i>	<i>Object</i>	<i>Approach</i>
<i>John Stigall</i>	<i>received</i>	<i>a Bachelor of arts</i>	<i>extractive</i>
<i>John Stigall</i>	<i>received</i>	<i>from the State University of New York at Cortland</i>	<i>extractive</i>
<i>dbpedia:John_Stigall</i>	<i>myprop:receives AcademicDegree</i>	<i>dbpedia:Bachelor_of_Arts</i>	<i>abstractive</i>
<i>dbpedia:John_Stigall</i>	<i>myprop:receives AcademicDegree From</i>	<i>dbpedia:State_University_of_New_York</i>	<i>abstractive</i>

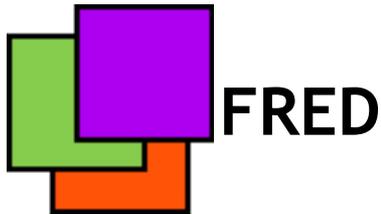
OIE vs OKE

“John Stigall received a Bachelor of arts from the State University of New York at Cortland “

<i>Subject</i>	<i>Predicate</i>	<i>Object</i>	<i>Approach</i>
<i>John Stigall</i>	<i>received</i>	<i>a Bachelor of arts</i>	<i>extractive</i>
<i>John Stigall</i>	<i>received</i>	<i>from the State University of New York at Cortland</i>	<i>extractive</i>
<i>myprop:receive_1</i>	<i>role:Agent</i>	<i>dbpedia:John_Stigall</i>	<i>abstractive</i>
<i>myprop:receive_1</i>	<i>role:Theme</i>	<i>dbpedia:Bachelor_of_Arts</i>	<i>abstractive</i>
<i>myprop:receive_1</i>	<i>role:Source</i>	<i>dbpedia:Univ_NYC_Cortland</i>	<i>abstractive</i>

OKE tools

Performing unsupervised, open domain, and abstractive knowledge extraction from text for producing directly usable machine readable data



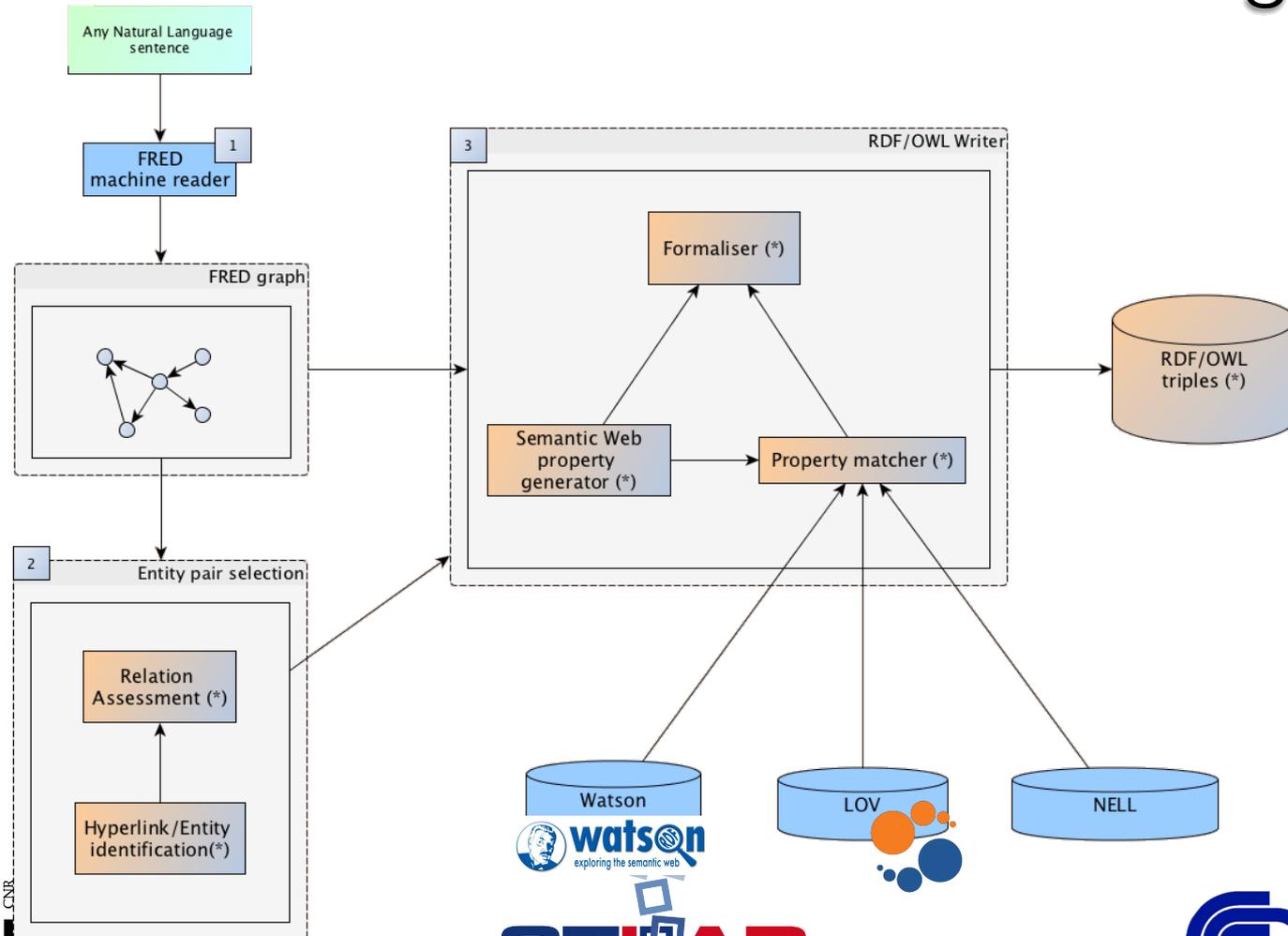
<http://wit.istc.cnr.it/stlab-tools/fred>

<http://wit.istc.cnr.it/stlab-tools/legalo>

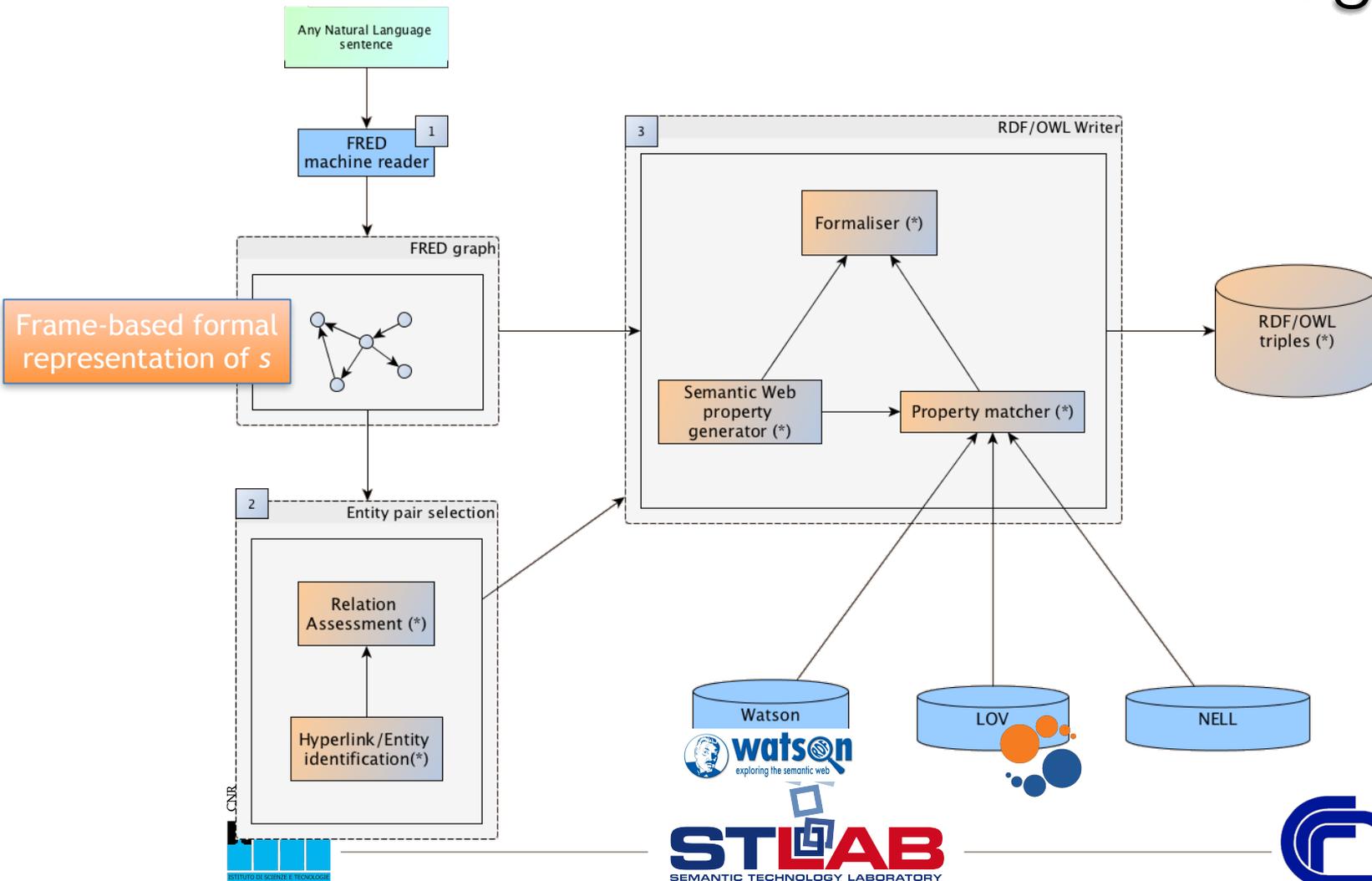


<http://wit.istc.cnr.it/stlab-tools/legalo/wikipedia>

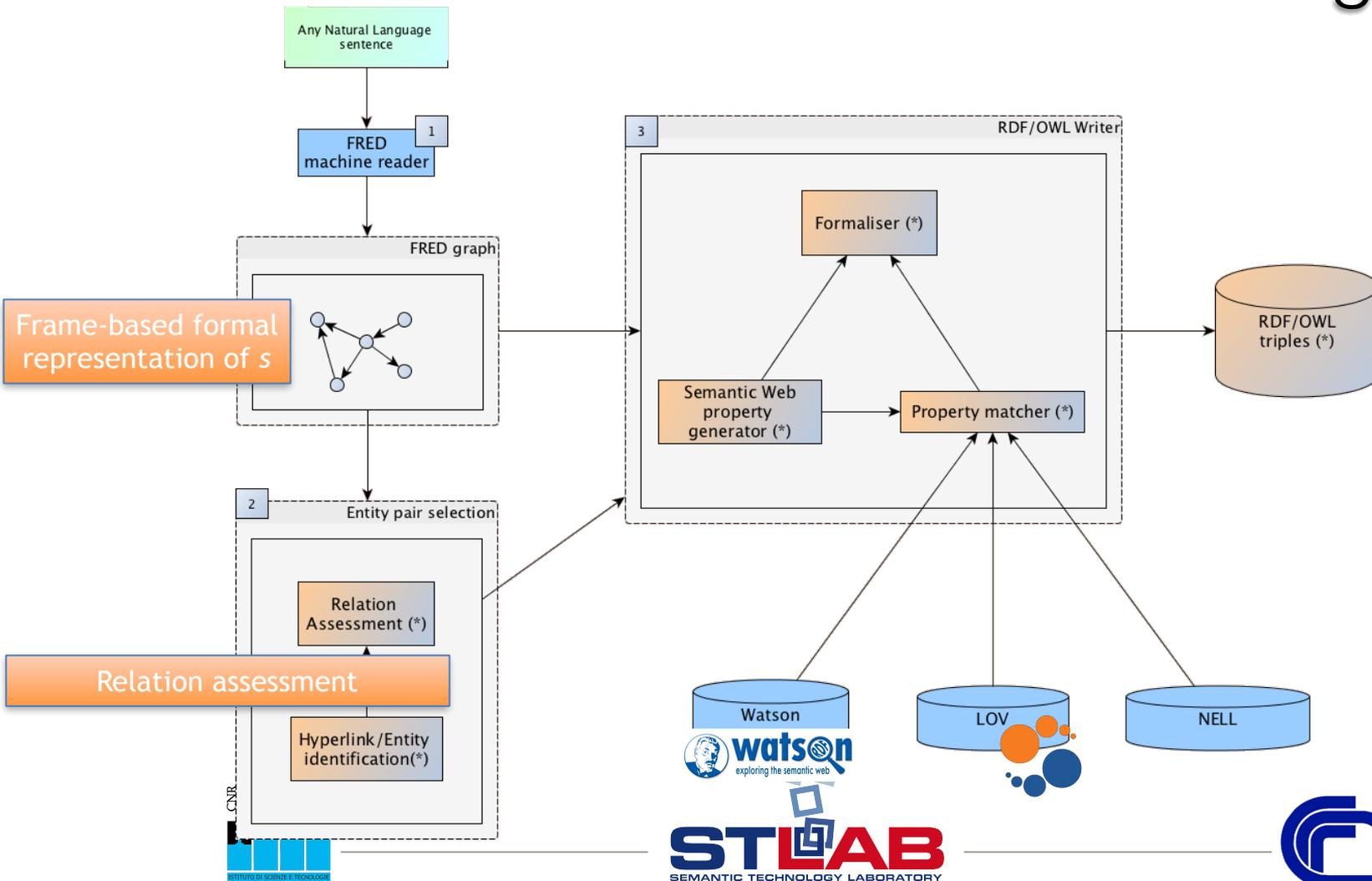
Method and implementation



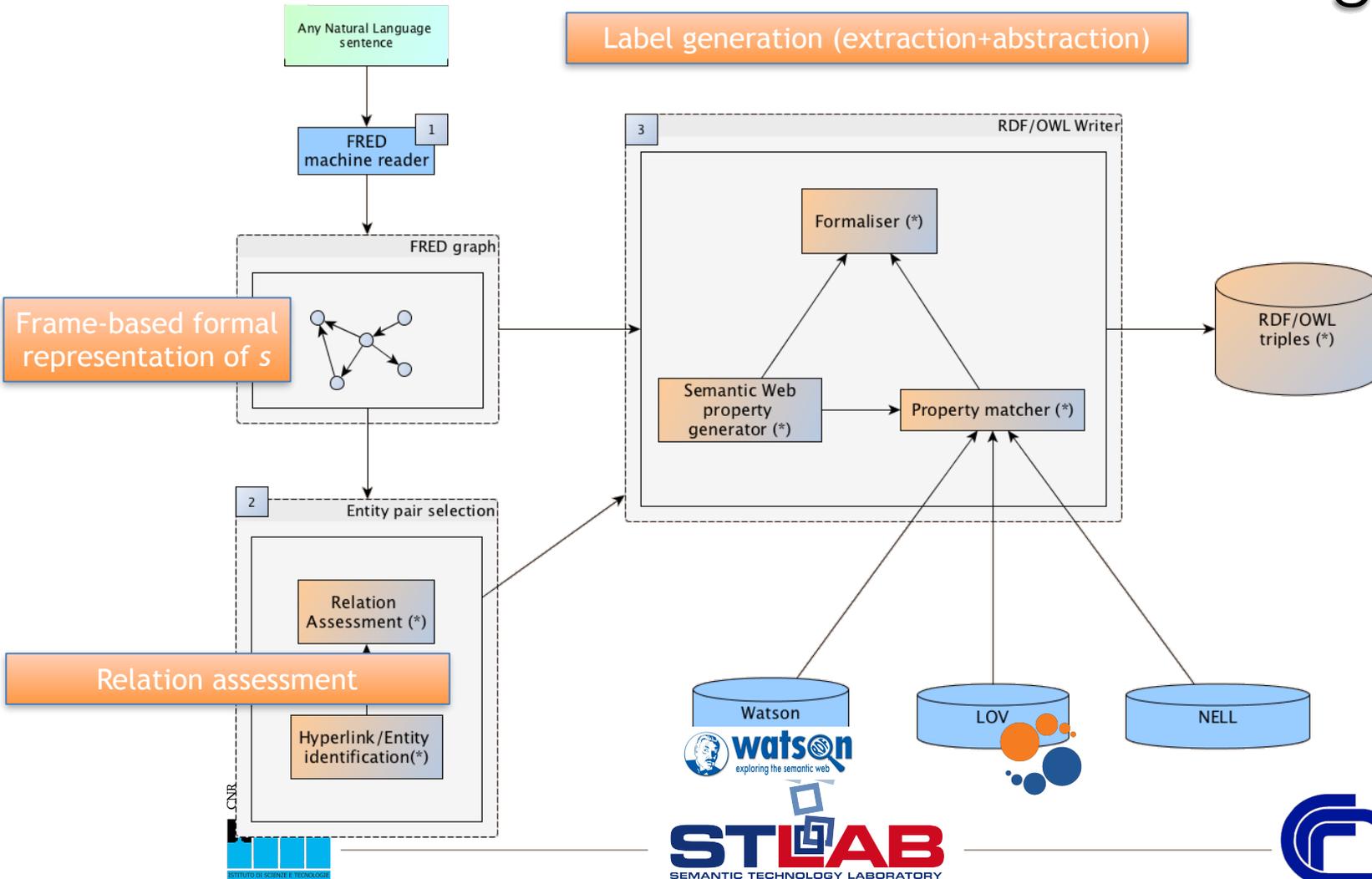
Method and implementation



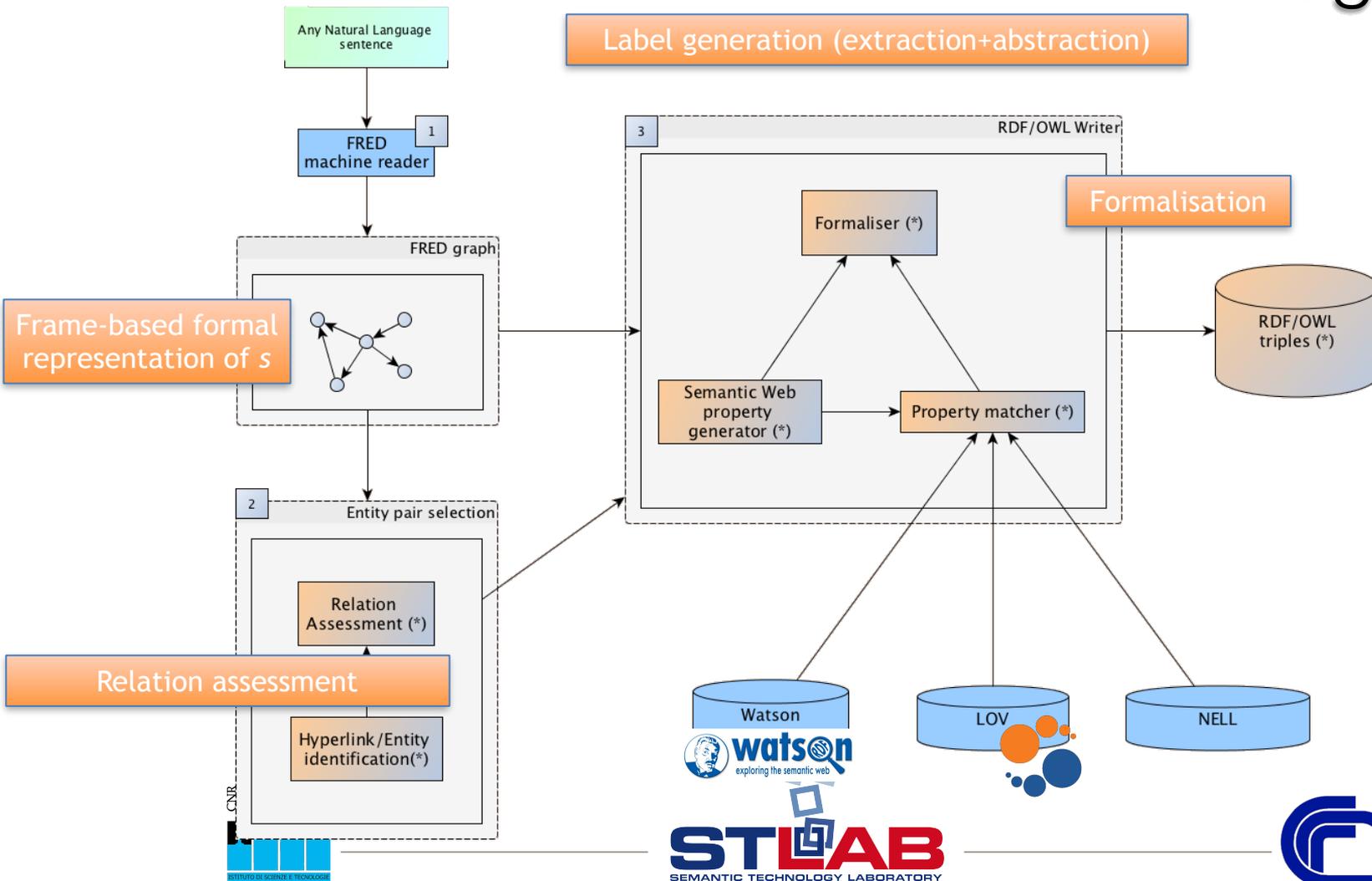
Method and implementation



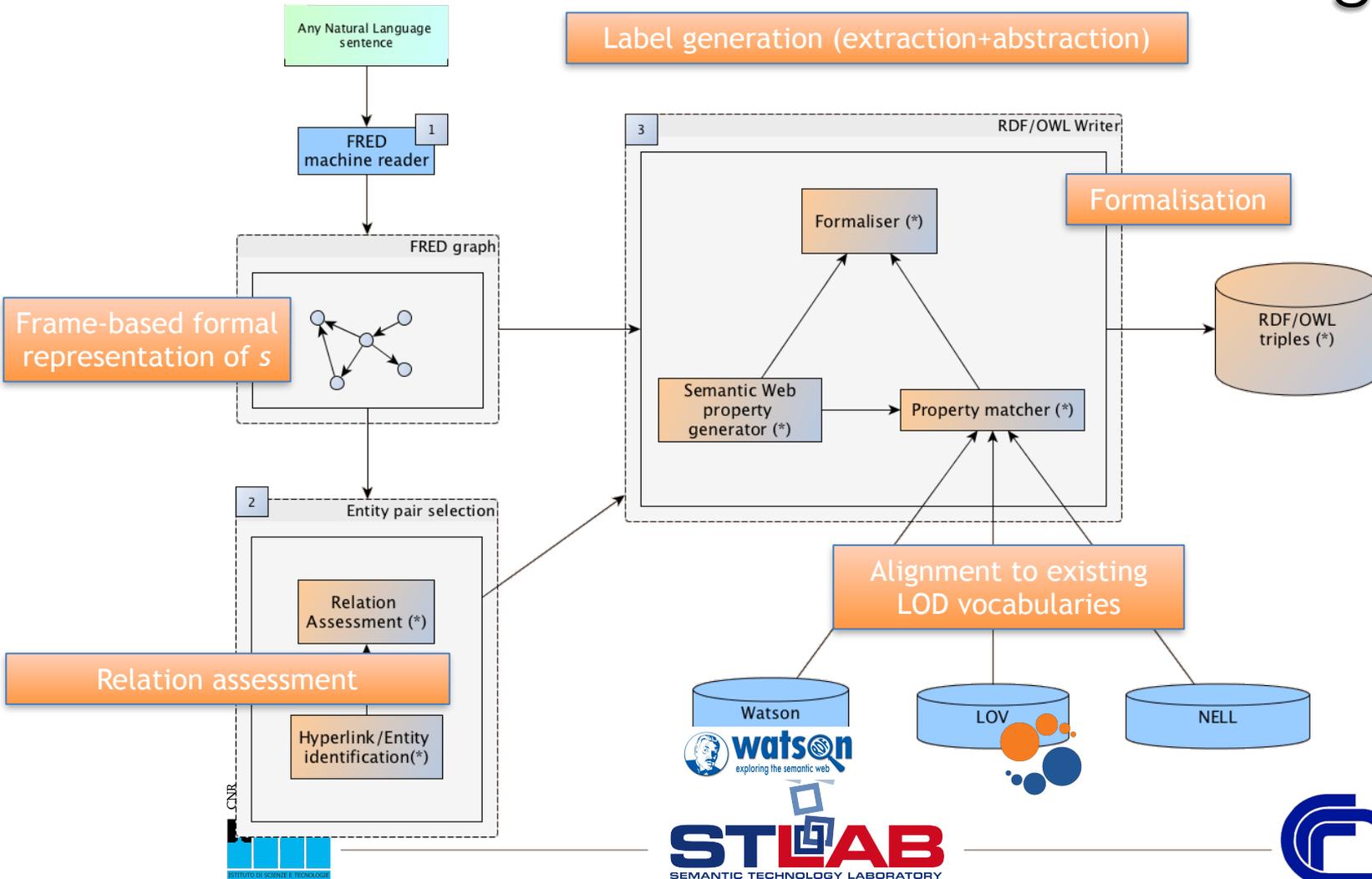
Method and implementation



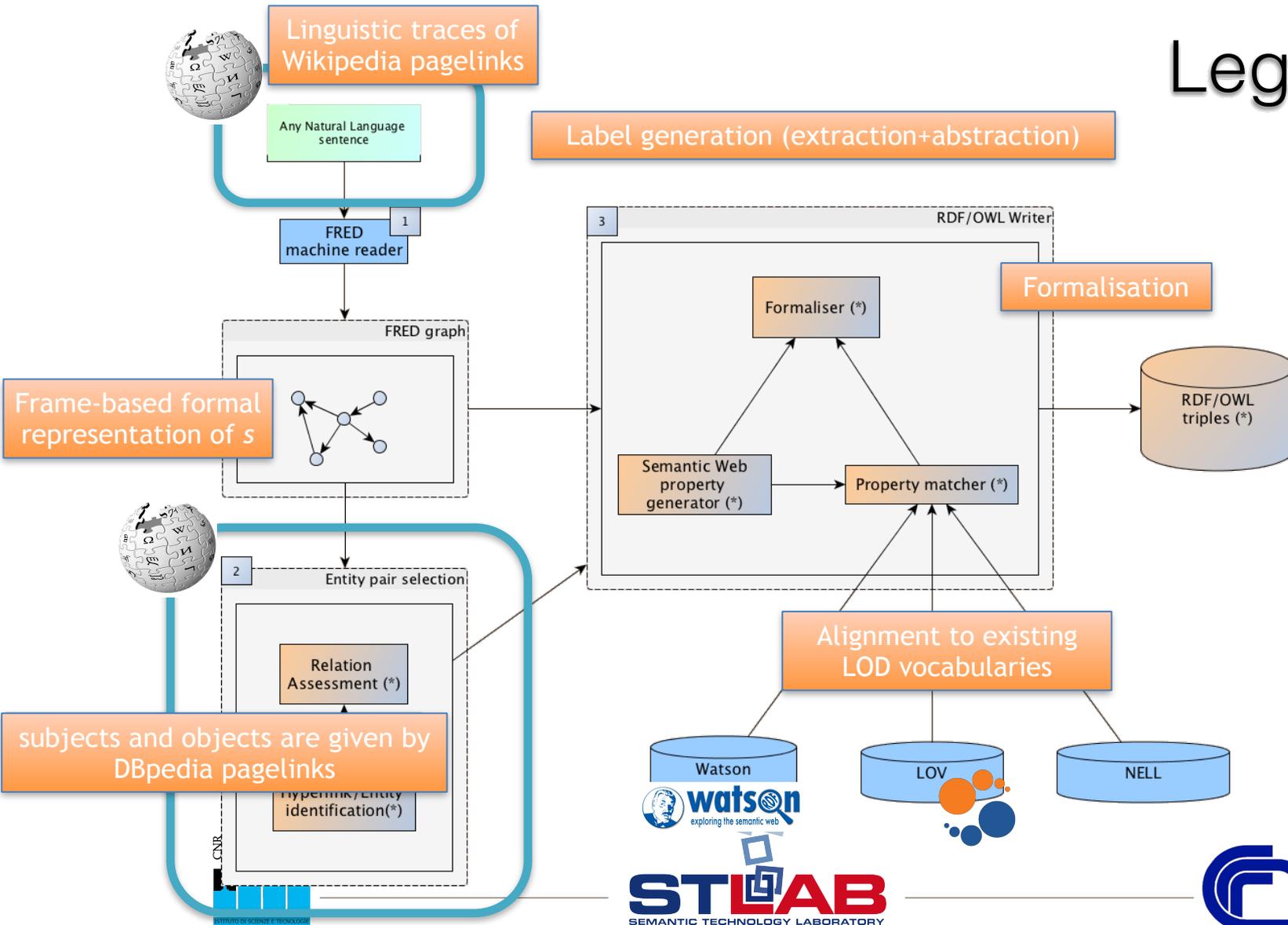
Method and implementation



Method and implementation

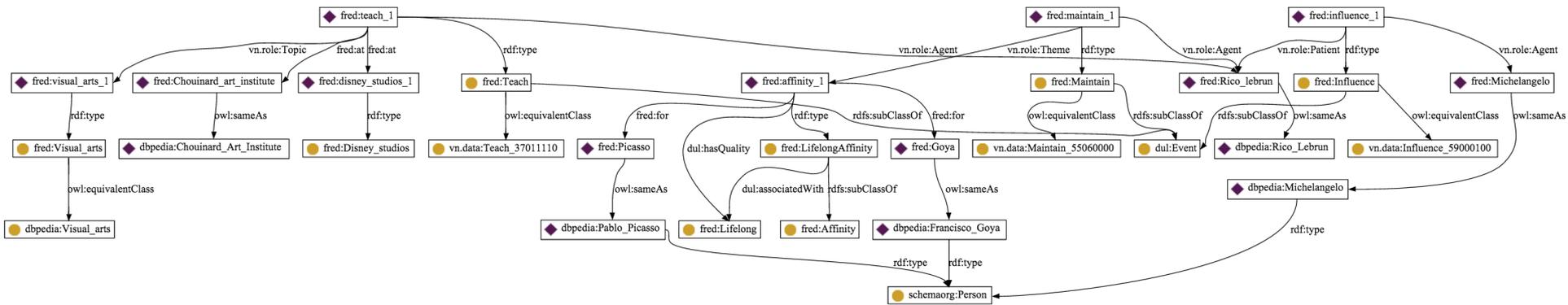


Method and implementation



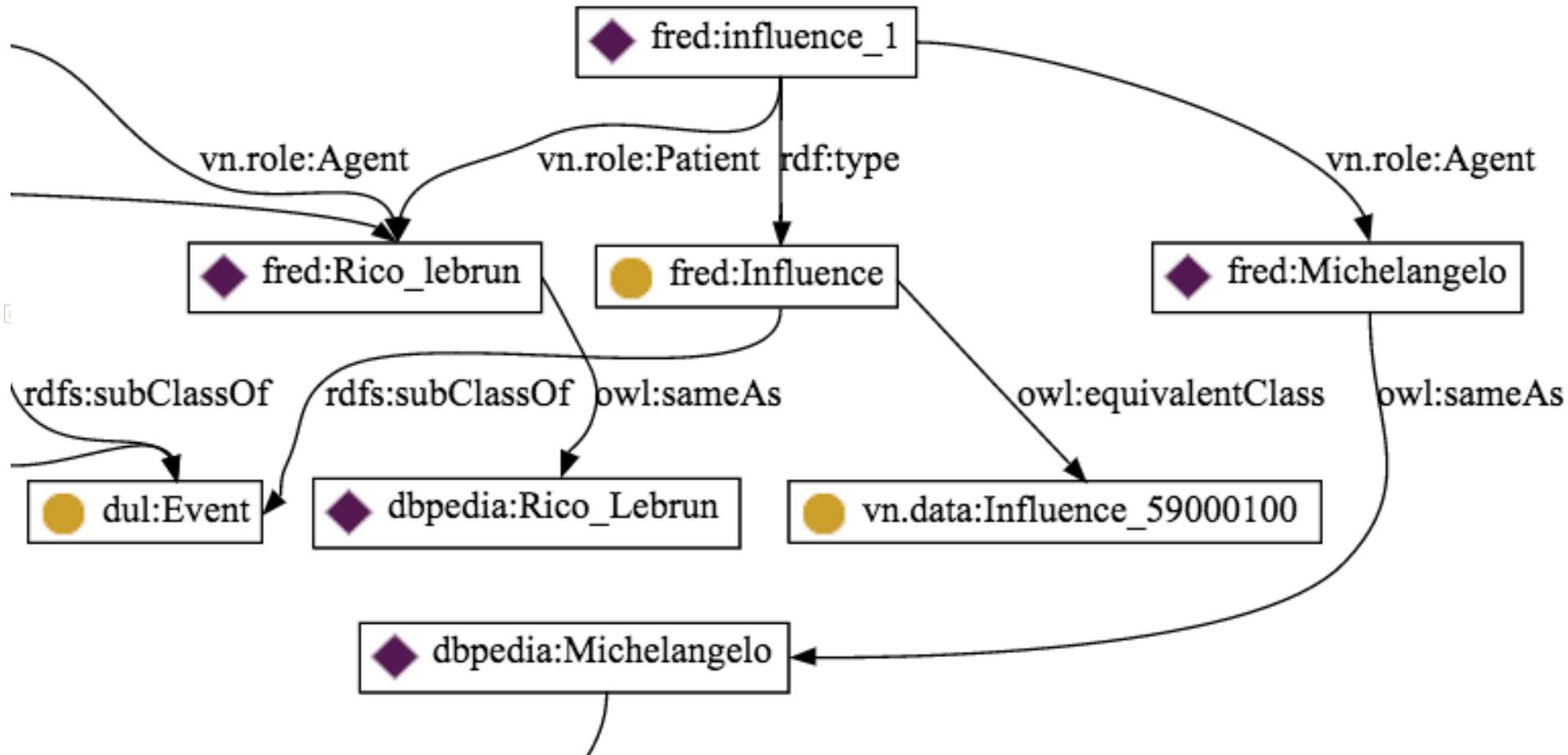
Frame-based formal representation

“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



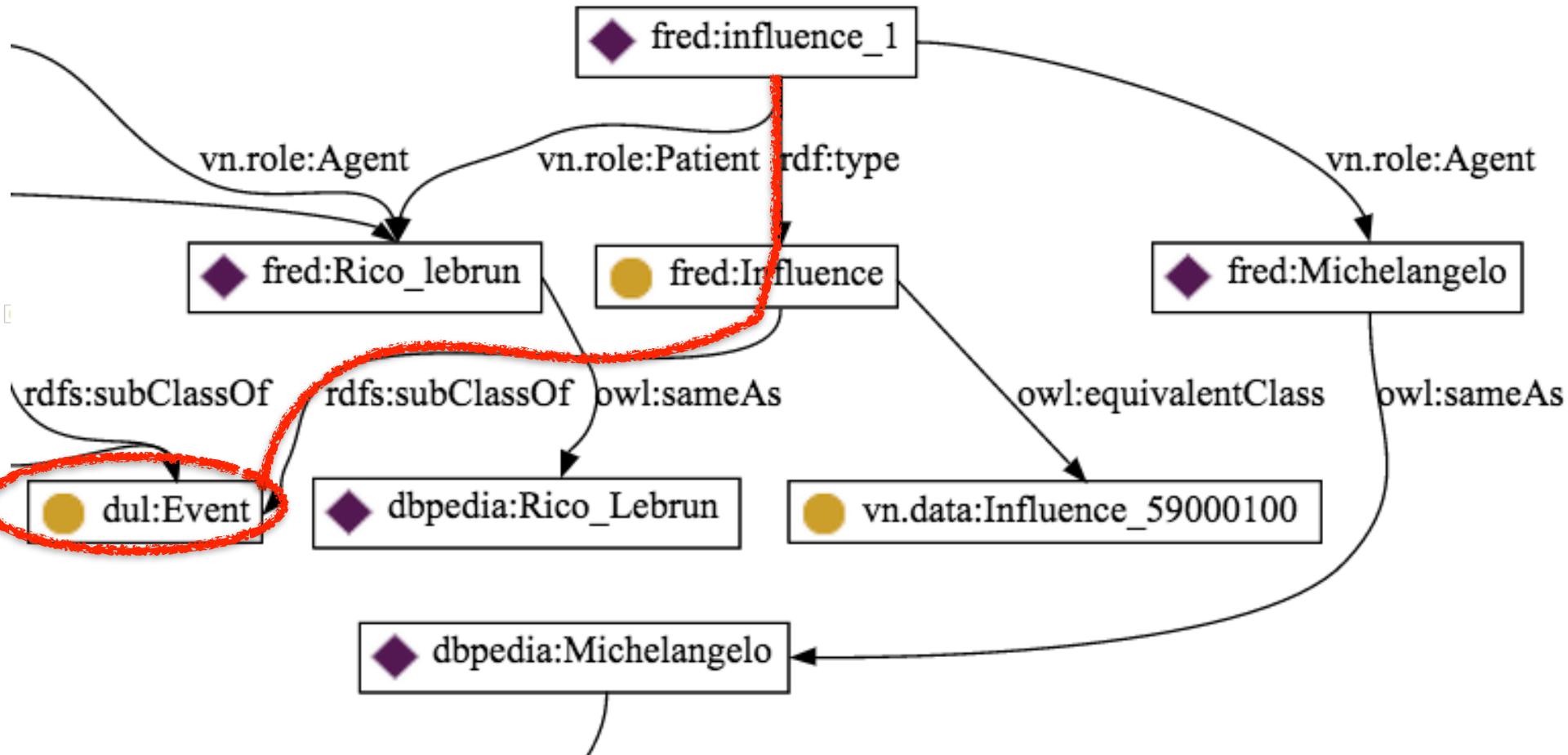
Frame-based formal representation

“[Rico Lebrun](#) taught [visual arts](#) at the [Chouinard Art Institute](#) and at the [Disney Studios](#). He was influenced by [Michelangelo](#) and maintained a lifelong affinity for [Goya](#) and [Picasso](#).”



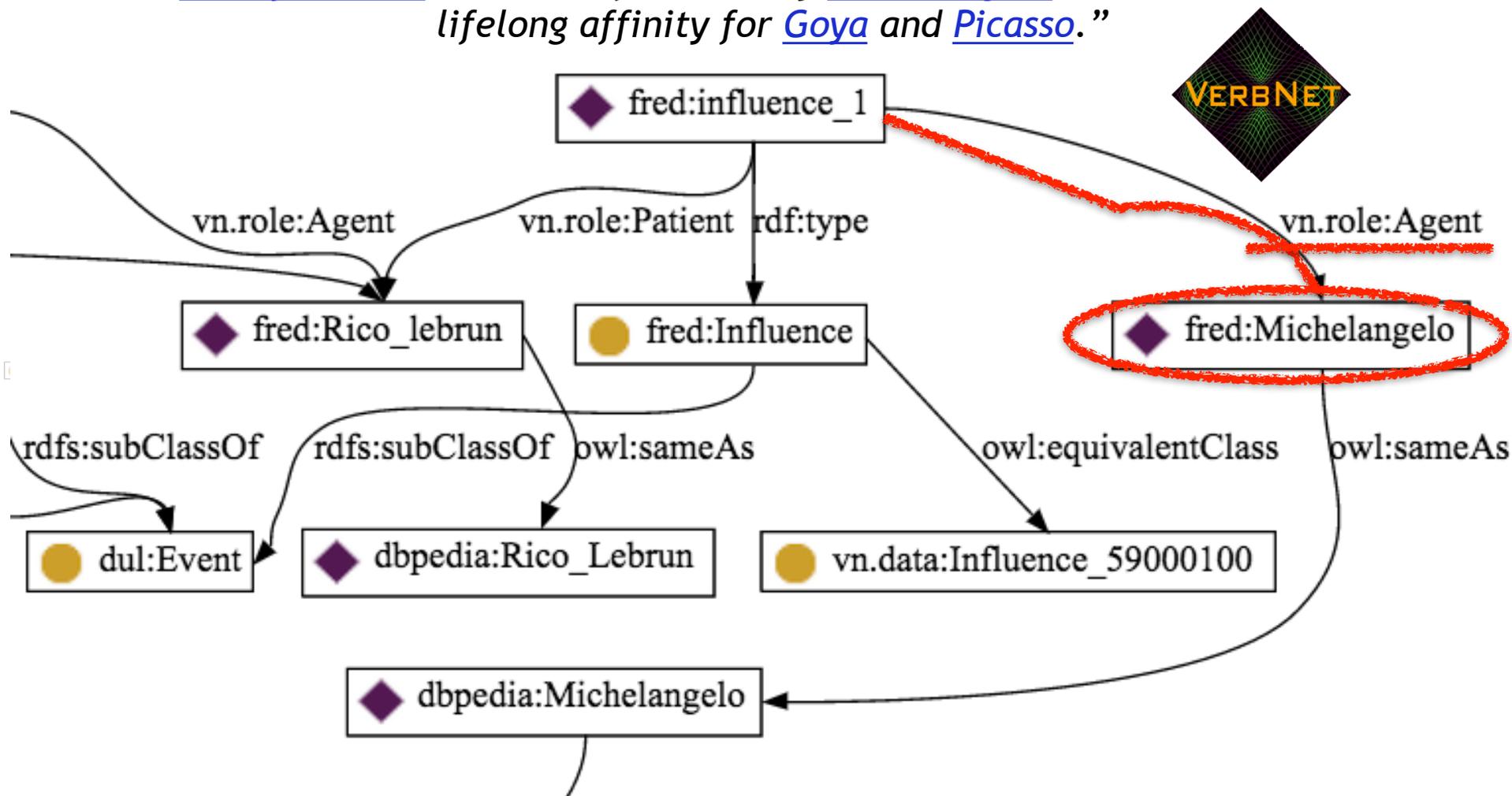
Frame-based formal representation

“[Rico Lebrun](#) taught [visual arts](#) at the [Chouinard Art Institute](#) and at the [Disney Studios](#). He was influenced by [Michelangelo](#) and maintained a lifelong affinity for [Goya](#) and [Picasso](#).”



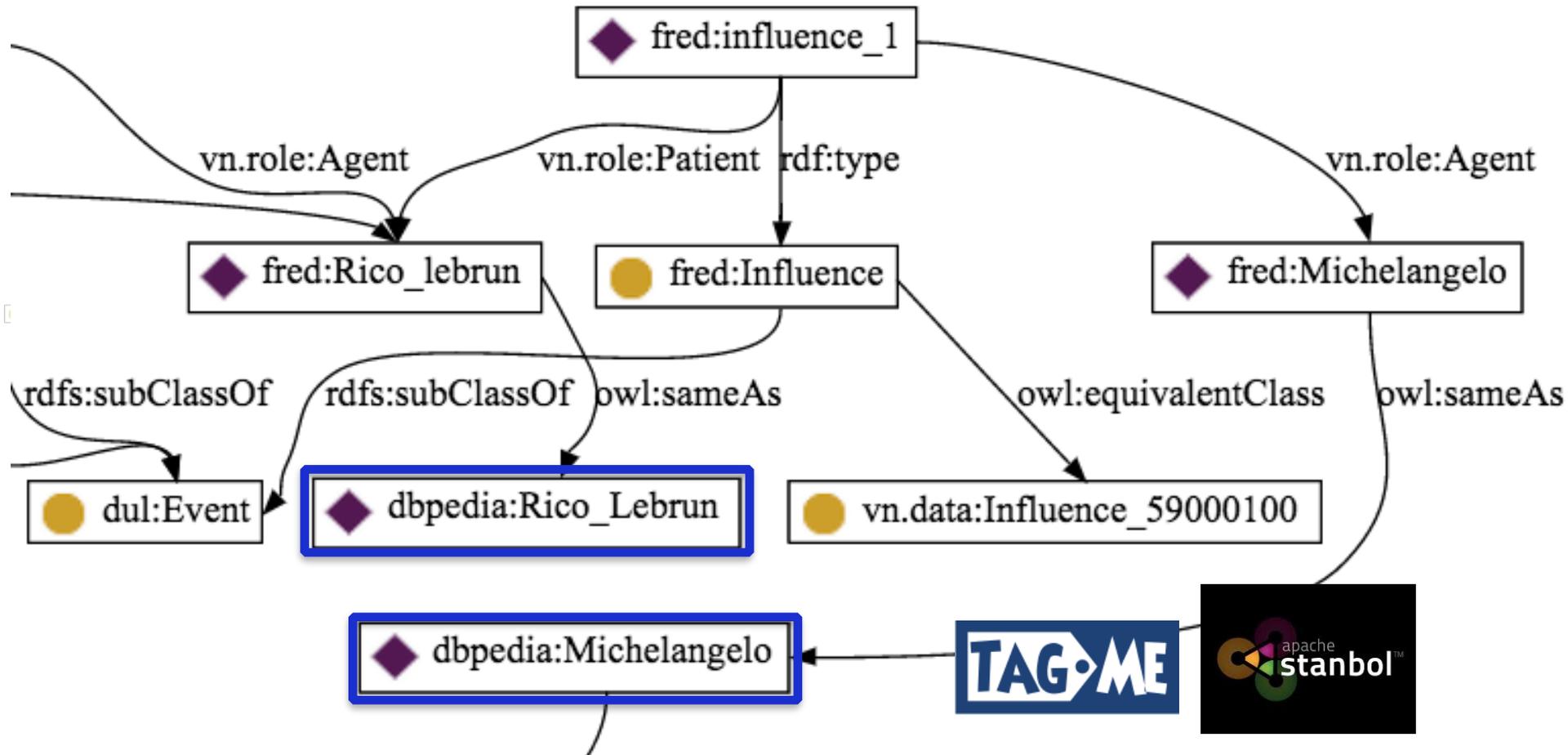
Frame-based formal representation

“[Rico Lebrun](#) taught [visual arts](#) at the [Chouinard Art Institute](#) and at the [Disney Studios](#). He was influenced by [Michelangelo](#) and maintained a lifelong affinity for [Goya](#) and [Picasso](#).”



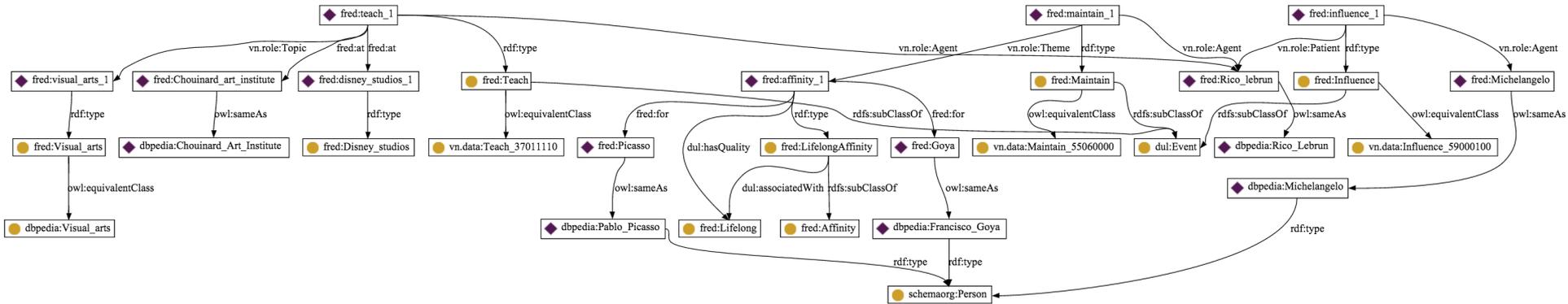
Frame-based formal representation

“[Rico Lebrun](#) taught [visual arts](#) at the [Chouinard Art Institute](#) and at the [Disney Studios](#). He was influenced by [Michelangelo](#) and maintained a lifelong affinity for [Goya](#) and [Picasso](#).”



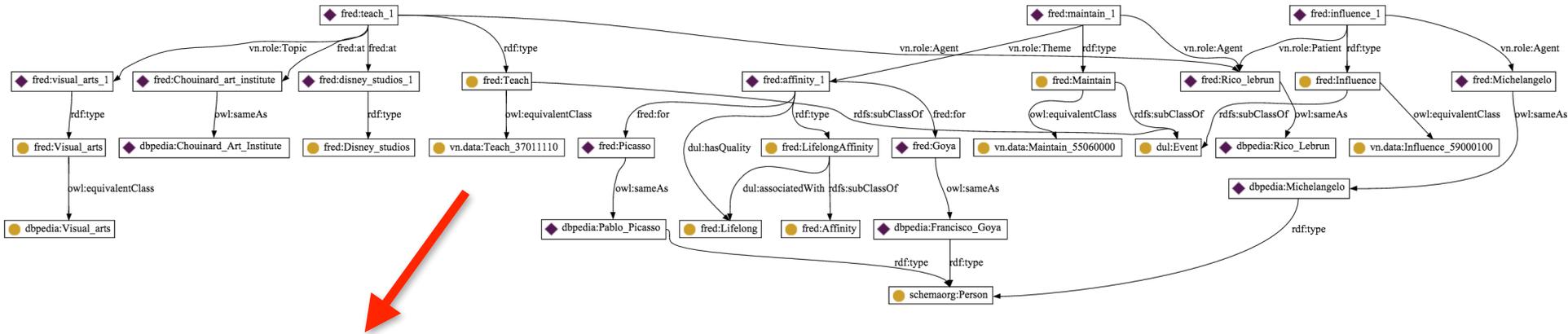
Binary relation assessment

“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



Binary relation assessment

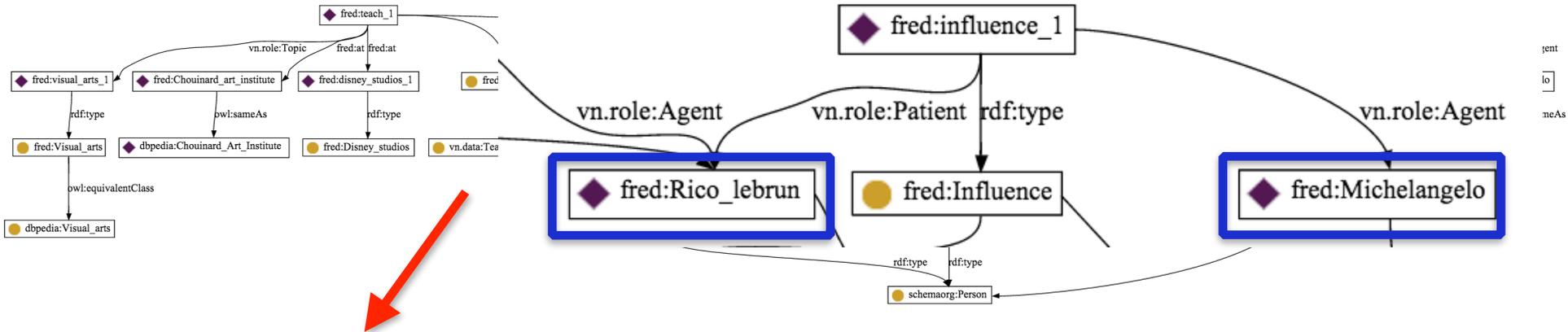
“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



$$G=(V,E), V \equiv \{v_0, \dots, v_n\}, E \equiv \{edge_1, \dots, edge_n\}$$

Binary relation assessment

“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



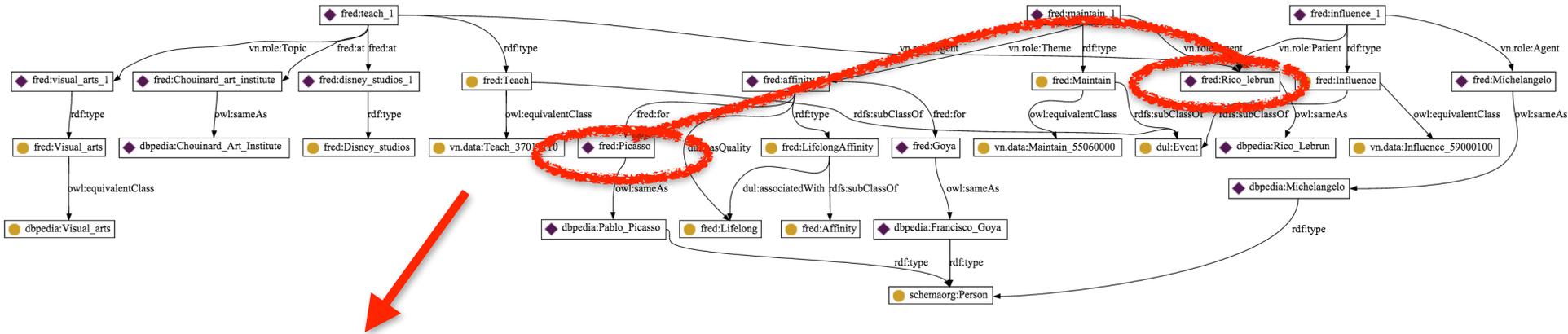
$G=(V,E)$, $V \equiv \{v_0, \dots, v_n\}$, $E \equiv \{edge_1, \dots, edge_n\}$

$G'=(V,E')$ is the undirected version of $G=(V,E)$

$v_i \in V$ is the node in G representing the entity e_i mentioned in s .

Binary relation assessment

“Rico Lebrun taught visual arts at the Chouinard Art Institute and at the Disney Studios. He was influenced by Michelangelo and maintained a lifelong affinity for Goya and Picasso.”



$G=(V,E)$, $V \equiv \{v_0, \dots, v_n\}$, $E \equiv \{edge_1, \dots, edge_n\}$

$G'=(V,E')$ is the undirected version of $G=(V,E)$

$v_i \in V$ is the node in G representing the entity e_i mentioned in s .

$P(v_{subj}, v_{obj})=[v_0, edge_1, \dots, edge_n, v_n]$, $Pset_{subj,obj} \equiv \{edge_1, v_1, \dots, v_{n-1}, edge_n\}$



Evaluation

Legalo-Wikipedia (EKAW 2014)

Number of p_{new}	Precision	Recall	F-measure	Kendall's W
629	0.72	0.97	0.83	0.73

3 raters, computer science, non-familiar with Legalo, familiar with Linked Data