

SPATIO-TEMPORAL VERACITY ASSESSMENT

JOANA GONZALES MALAVERRI (LAHDAK)

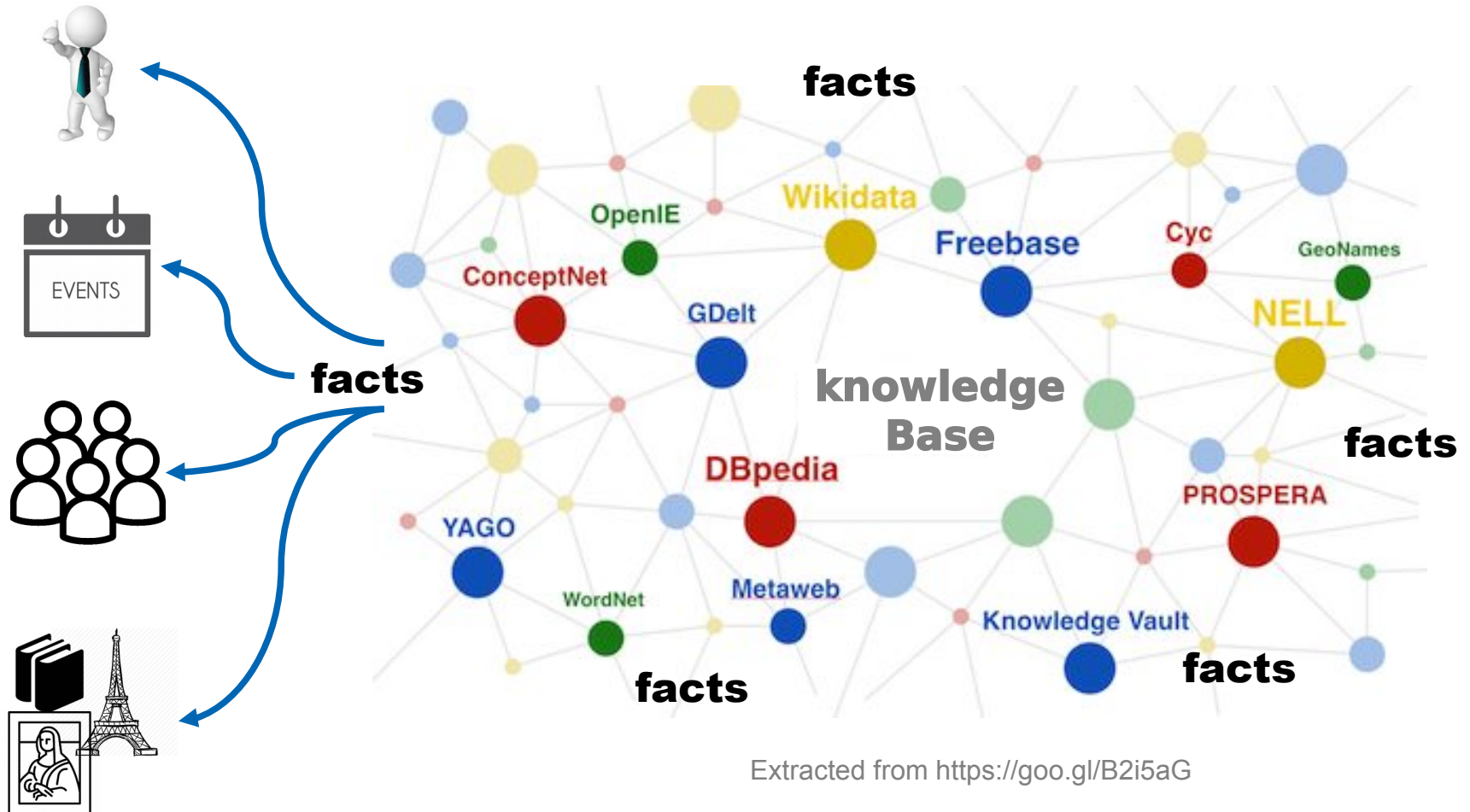
FATIHA SAÏS (LAHDAK)

GIANLUCA QUERCINI (MODHEL)

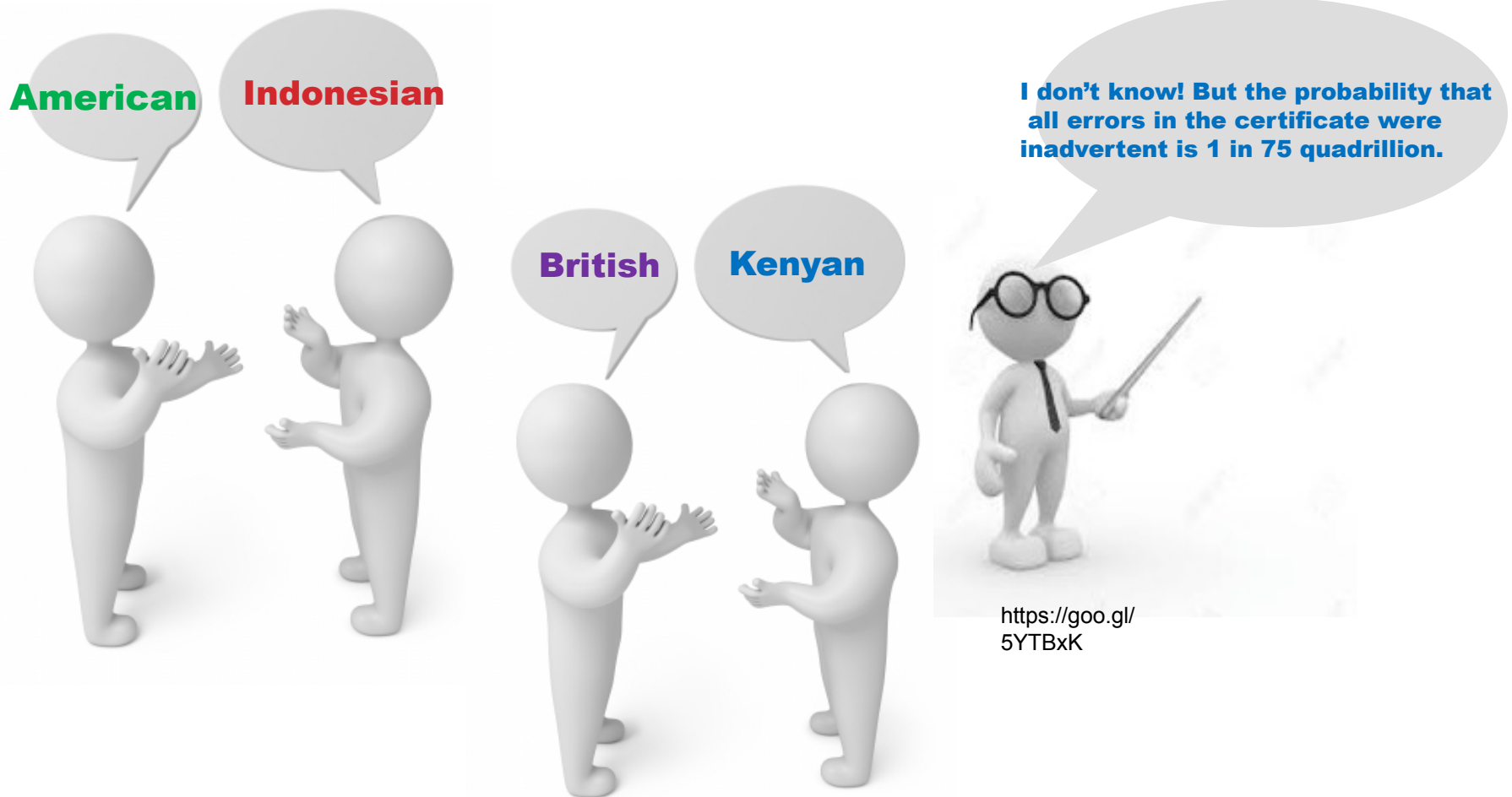
LABORATOIRE DE RECHERCHE EN INFORMATIQUE (LRI)
{MALAVERRI, FATIHA.SAIS,GIANLUCA.QUERCINI}@LRI.FR

JOURNEE ROD

MOTIVATION



REAL-WORLD SCENARIO: WHAT IS THE BARACK OBAMA'S CITIZENSHIP





WIKIPEDIA
The Free Encyclopedia



tumblr



TypePad



SQUARESPACE



facts

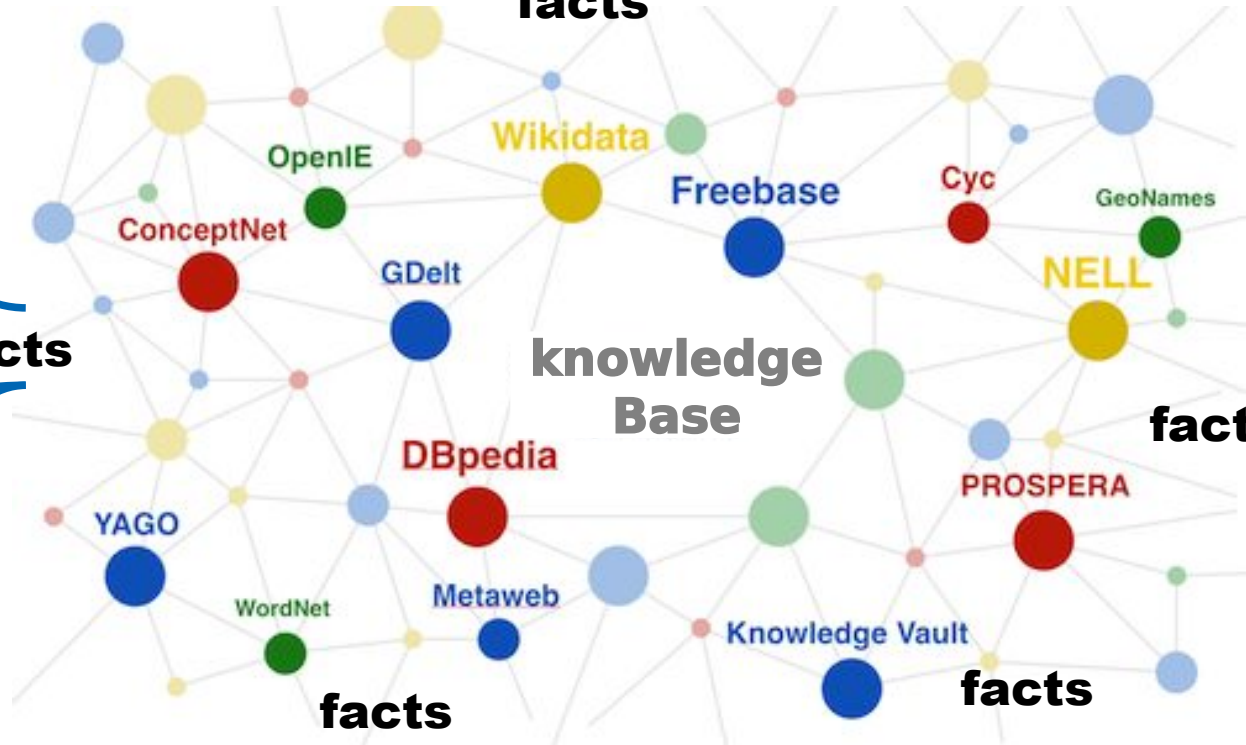
facts

**knowledge
Base**

facts

facts

facts

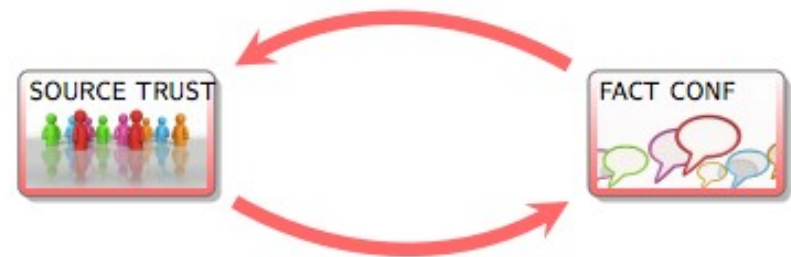


Extracted from <https://goo.gl/B2i5aG>

VERACITY ASSESSMENT APPROACHES

[Beretta et al. 16]

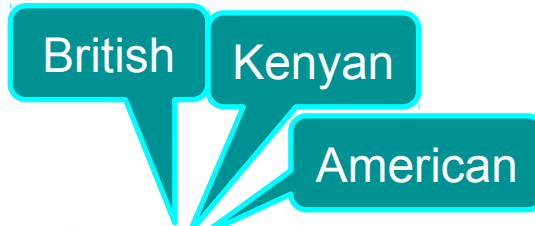
- Majority voting
- Basic approaches
- Extended approaches
 - Extra knowledge (ontologies)
 - Source dependency detection



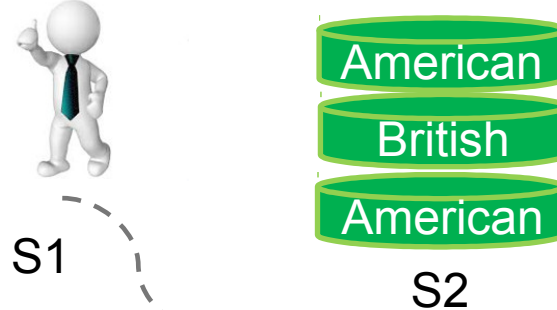
VERACITY ASSESSMENT

What is the nationality of those US presidents?

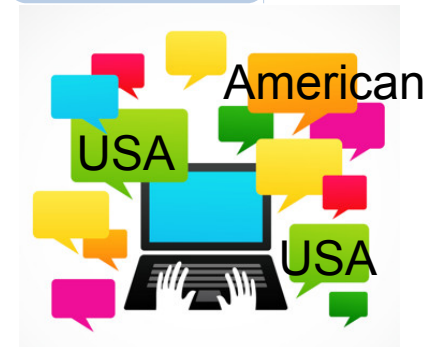
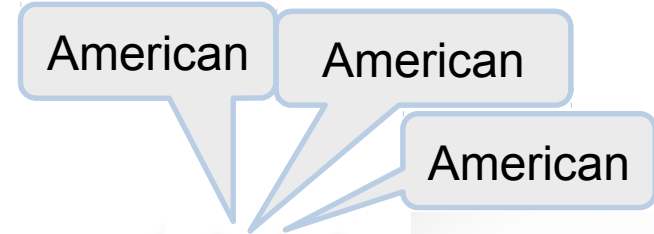
d1. Edward
Dickinson
Baker



d2.
Barack
Obama



d3. Bill
Clinton

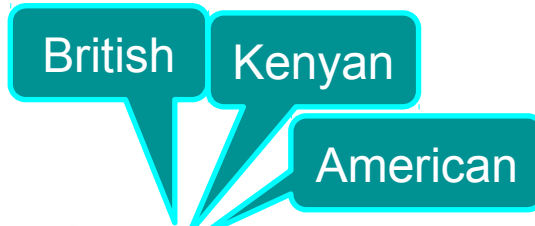


<u>d1:nat</u>	<u>d2:nat</u>	<u>d3:nat</u>
S1 <British, Kenyan, American>		
S2 <American, British, American>		
S3 <American, American, American>		
S4 <USA, USA, American>		

VERACITY ASSESSMENT

What is the nationality of those US presidents?

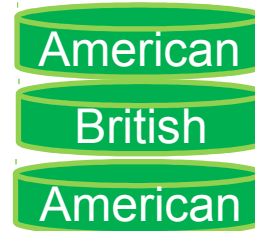
d1. Edward
Dickinson
Baker



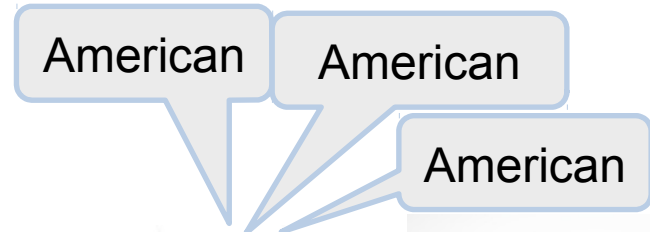
d2. Barack
Obama



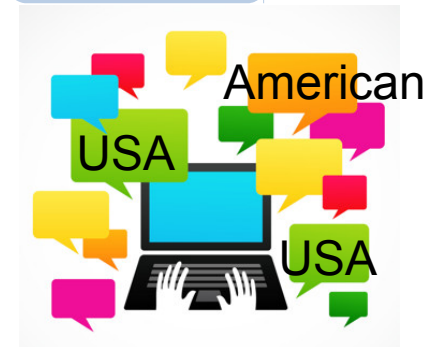
S1



S2



S3

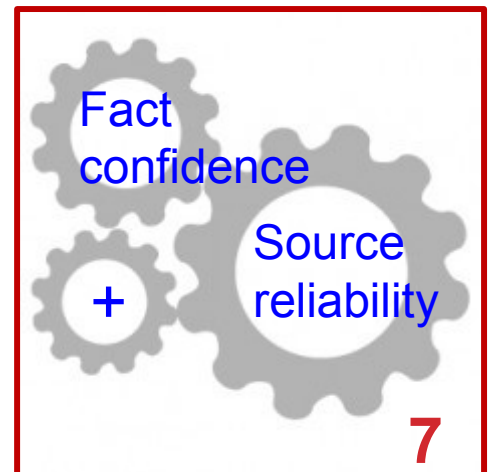


S4

d3. Bill
Clinton



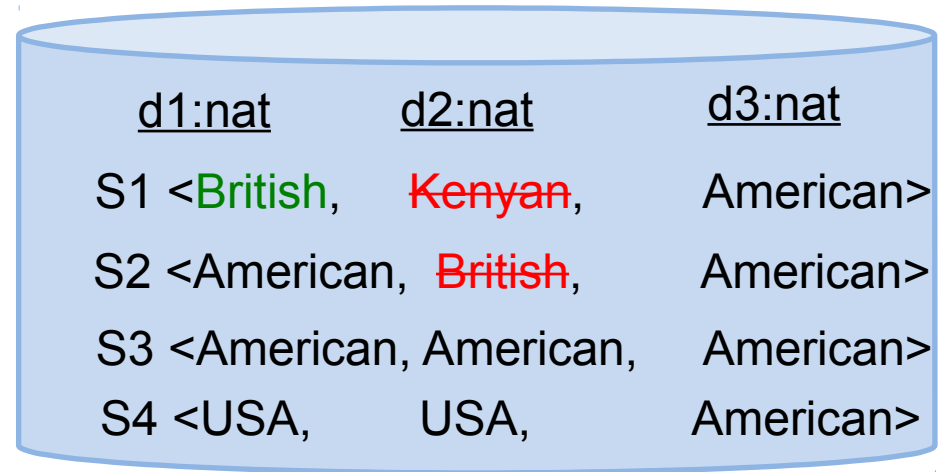
<u>d1:nat</u>	<u>d2:nat</u>	<u>d3:nat</u>
S1 <British, Kenyan,		American>
S2 <American, British,		American>
S3 <American, American,		American>
S4 <USA, USA,		American>



VERACITY ASSESSMENT APPROACHES: LIMITATIONS

- **Single-Truth**

S1: <d1:nat, British> is also true.



<u>d1:nat</u>	<u>d2:nat</u>	<u>d3:nat</u>
S1 <British,	Kenyan,	American>
S2 <American,	British,	American>
S3 <American, American,		American>
S4 <USA,	USA,	American>

VERACITY ASSESSMENT APPROACHES: LIMITATIONS

- **Single-Truth**

S1: <d1:nat, British> is also true.

- **No contextual information**

S1: <d1:nat, British> is true in the temporal context [1811-1816]

S1: <d1:birthdate, 1811>

S1: <d1:birthPlace, London>

S1: <d1:immigrationDate, 1816>

<u>d1:nat</u>	<u>d2:nat</u>	<u>d3:nat</u>
S1 <British,	Kenyan,	American>
S2 <American,	British,	American>
S3 <American, American,		American>
S4 <USA,	USA,	American>

VERACITY ASSESSMENT

APPROACHES: LIMITATIONS

- **Single-Truth**

S1: <d1:nat, British> is also true.

- **No contextual information**

S1: <d1:nat, British> is true in the temporal context [1811-1816]

S1: <d1:birthdate, 1811>

S1: <d1:birthPlace, London>

S1: <d1:immigrationDate, 1816>

- **Some sources mainly created from data extracted from Wikipedia**

How reliable S1: <d1:nat, British> is?

S1: <d1:birthPlace, London>

S1: <d1:immigrationDate, 1816>

<u>d1:nat</u>	<u>d2:nat</u>	<u>d3:nat</u>
S1 <British, Kenyan,		American>
S2 <American, British,		American>
S3 <American, American,		American>
S4 <USA, USA,		American>

VERACITY ASSESSMENT APPROACHES: LIMITATIONS

- **No explanations**

Why S1: <d1:nat, British> is true?

S1: <d1:birthPlace, London>

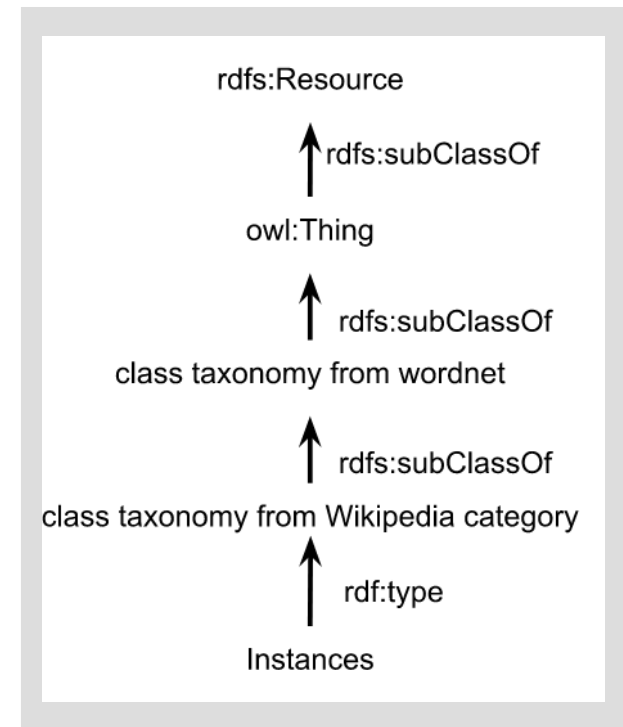
<u>d1:nat</u>	<u>d2:nat</u>	<u>d3:nat</u>
S1 <British,	Kenyan,	American>
S2 <American,	British,	American>
S3 <American, American,		American>
S4 <USA,	USA,	American>

GOAL

Build an approach to assess the veracity of facts taken from knowledge bases based on spatio-temporal information.

YAGO KNOWLEDGE BASE

- General purpose semantic knowledge base (KB)
 - Integrates information extracted from Wikipedia infoboxes, WordNet, and GeoNames
 - > 10 million entities (persons, cities, organizations),
 - > 120 million facts about these entities
- Attaches **temporal** and **spatial** dimensions to many of its facts and entities – meta facts.



Yago structure

APPROACH: RULE BASED TEMPORAL META FACTS GENERATION

RULE BASED TEMPORAL META FACTS GENERATION

Focus on facts that may change over time:

- Brad Pitt *acted* in
 - the Fight Club in 1999
 - the Curious Case of Benjamin Button in 2008
- Paul McCartney was/is *married* with
 - Heather Mills from 2002 to 2008
 - Nancy Shevell since 2011

BASIC ALGORITHM TO INFER META FACTS

Input:

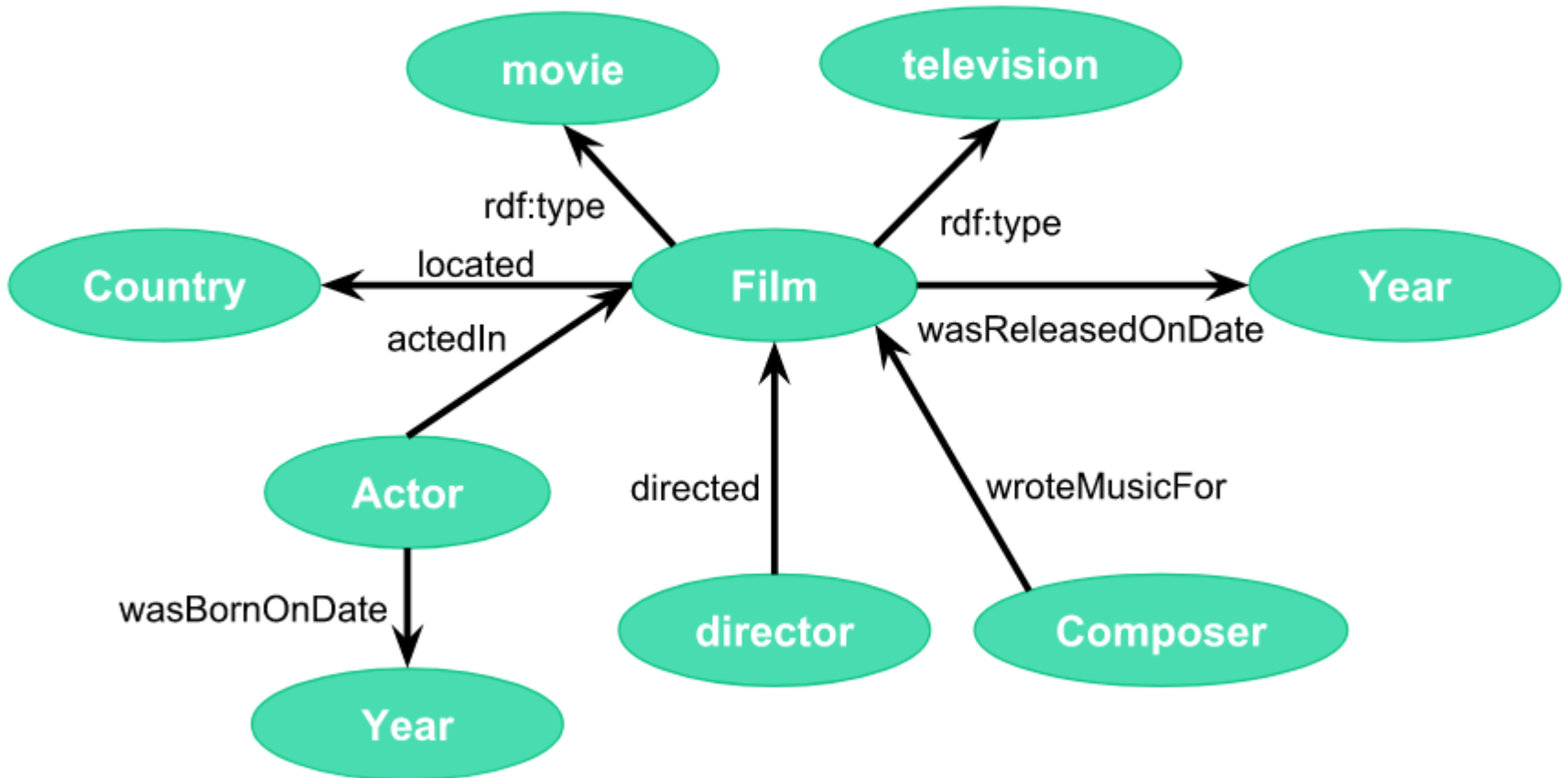
- F: a set of facts $\langle s, p, o \rangle$
- P: a set of predicates time sensitive, e.g.: actedIn
- R: a set of rules which generates the time meta facts for P.

Output: VD: a set of valid_on_date meta facts

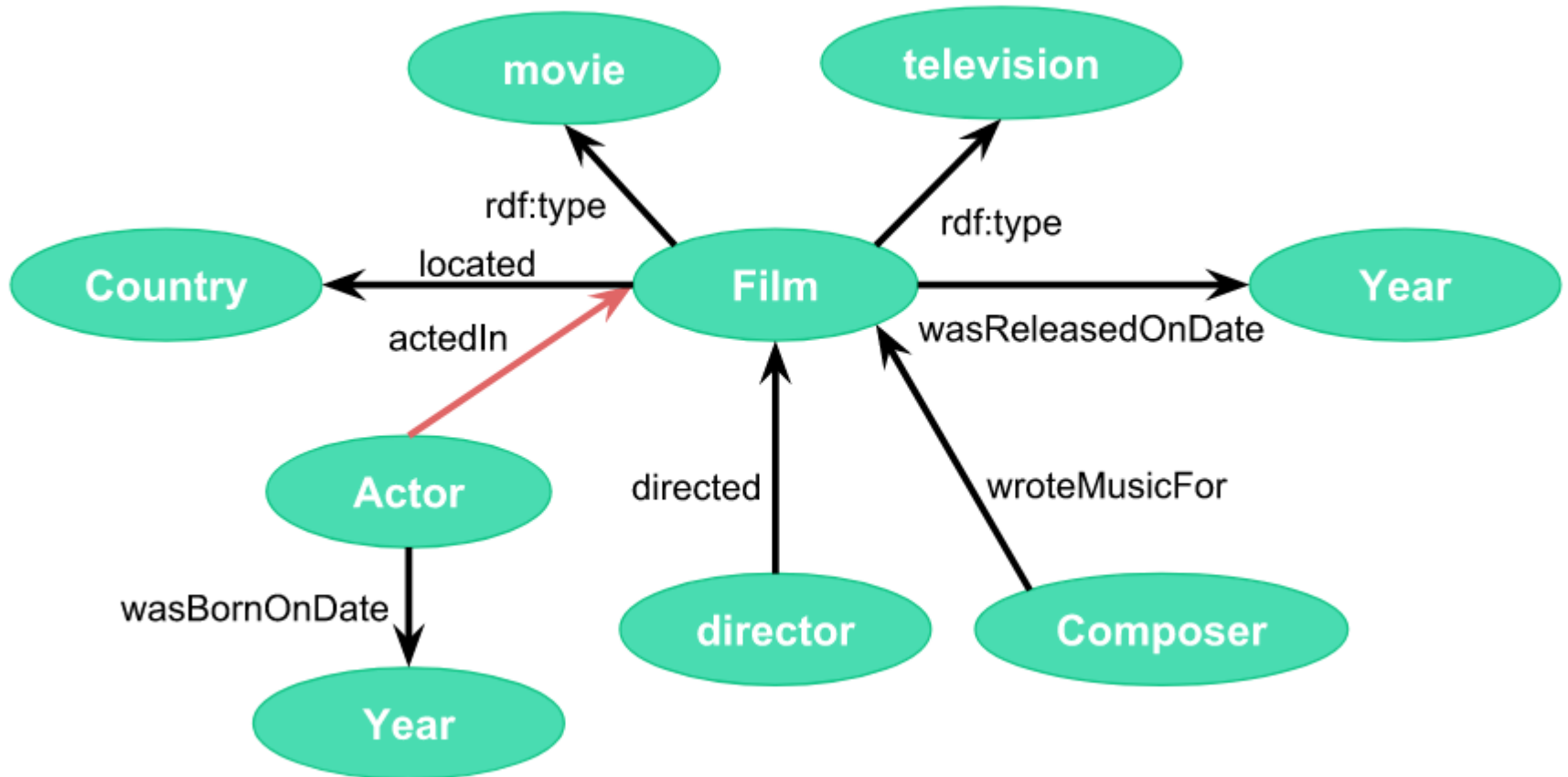
Algorithm

```
1 foreach  $p \in P$  do  
2    $r \leftarrow \text{getRules}(R, p)$   
3    $VD \leftarrow VD \cup \text{applyRules}(r, F)$   
4 end  
5 return VD
```

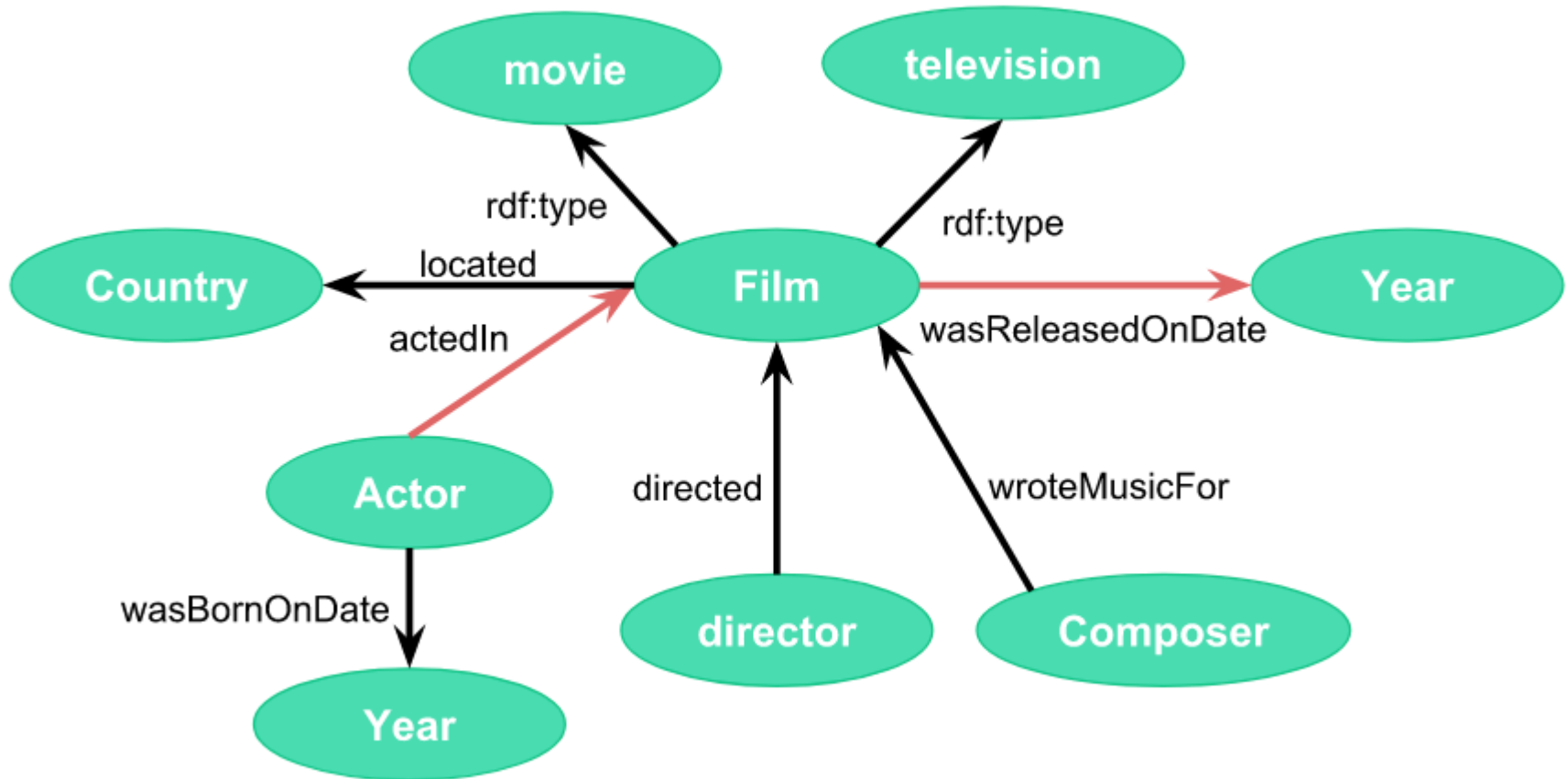
TIMESTAMP GENERATION RULE



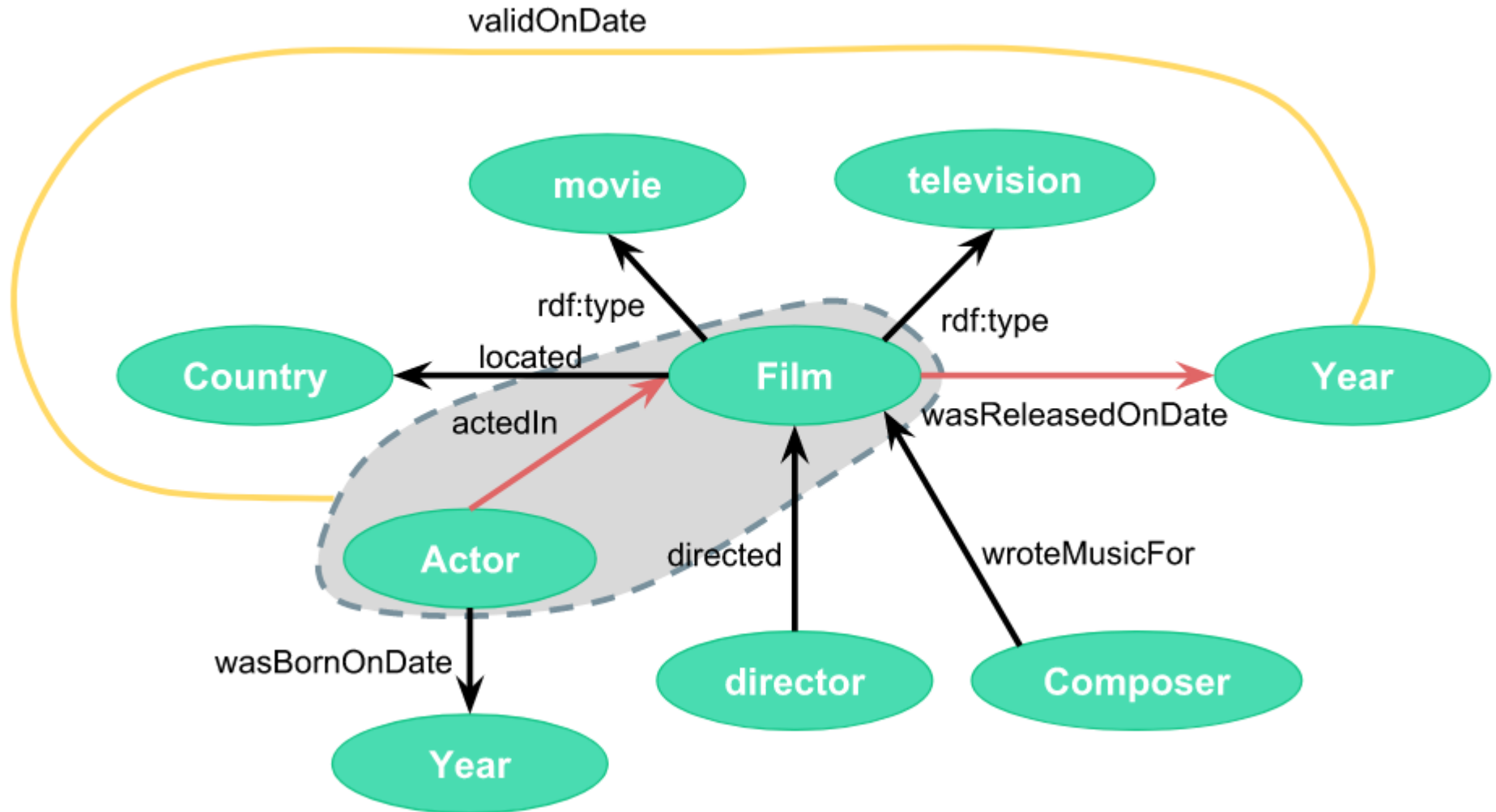
TIMESTAMP GENERATION RULE



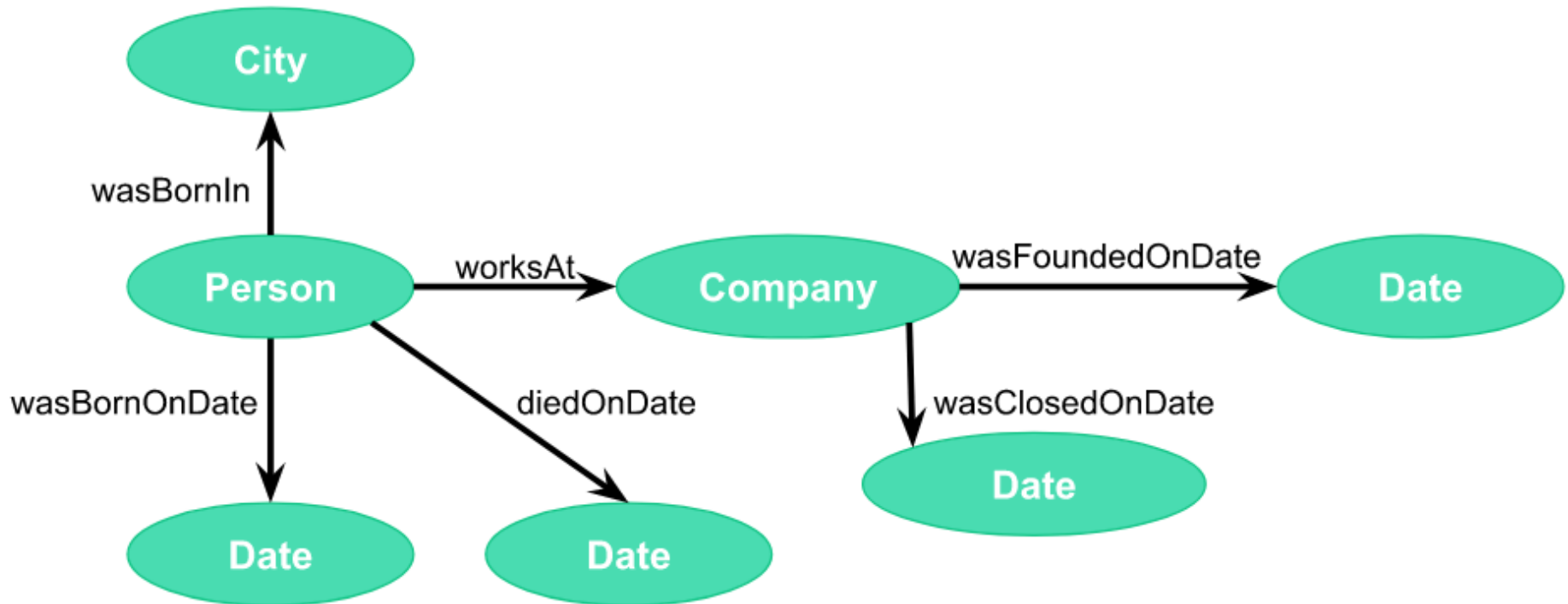
TIMESTAMP GENERATION RULE



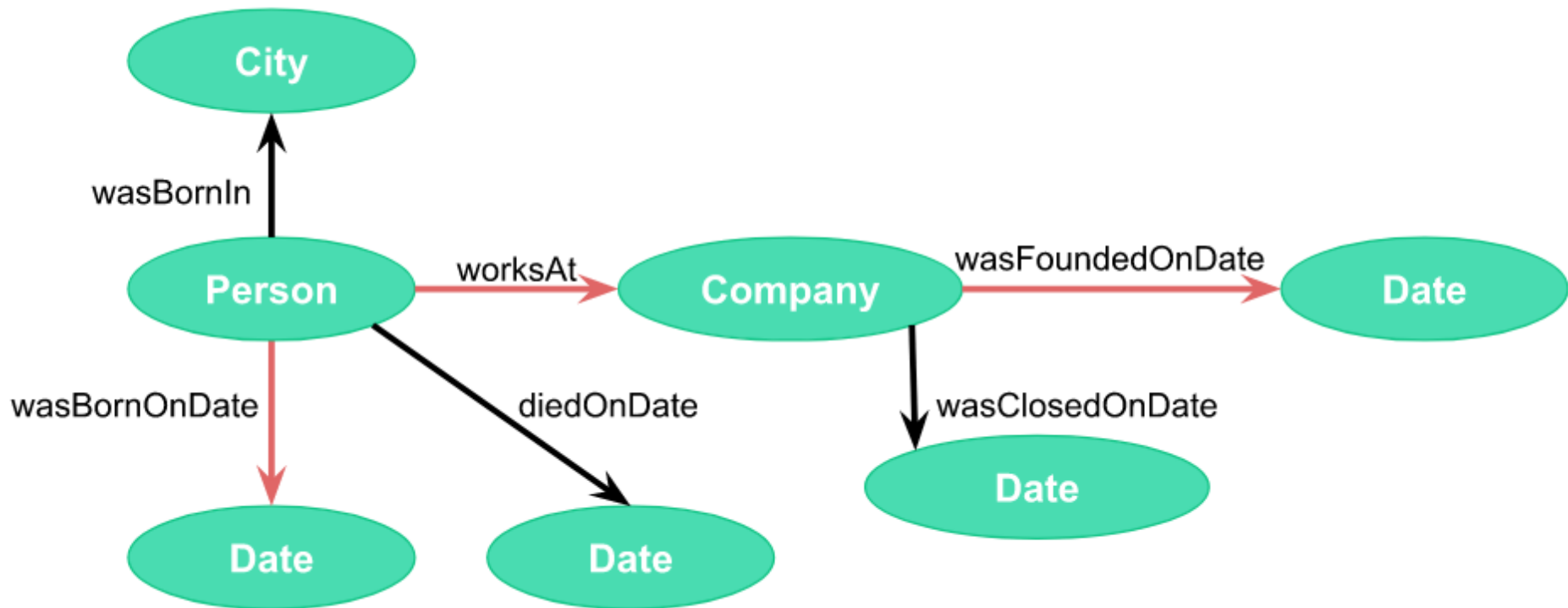
TIMESTAMP GENERATION RULE



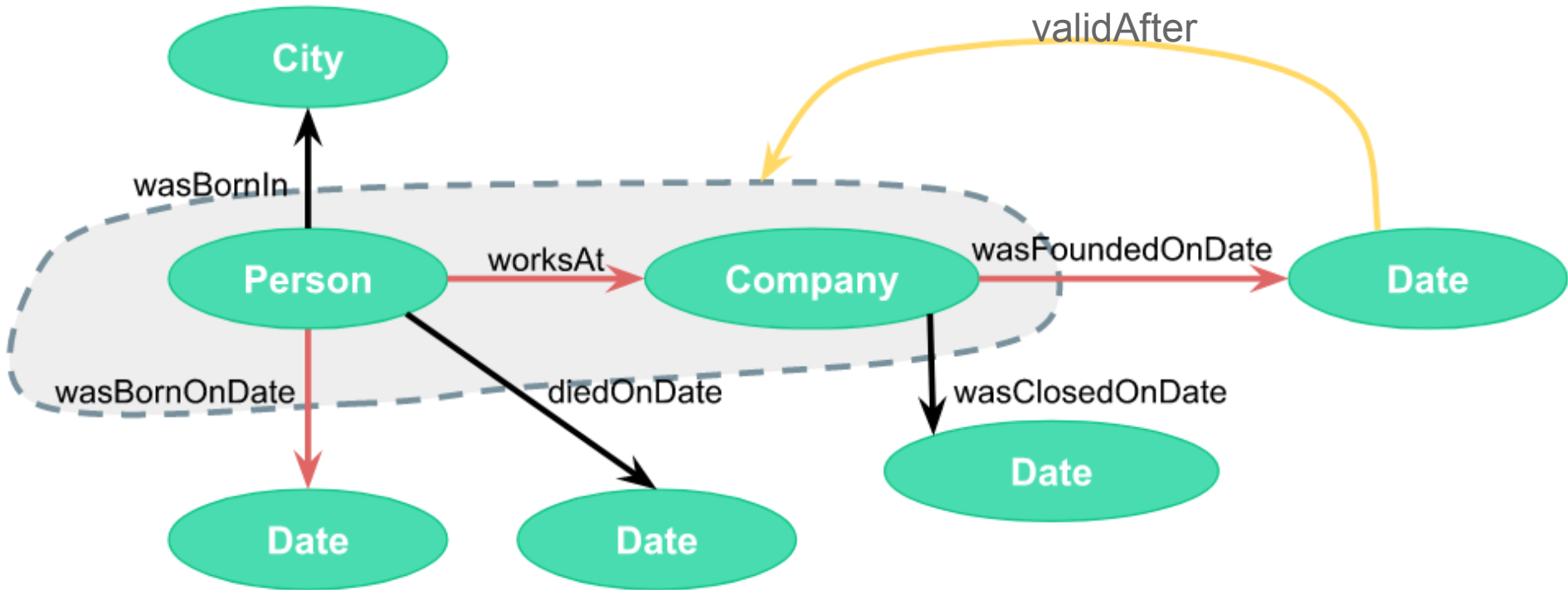
TEMPORAL INTERVAL RULE



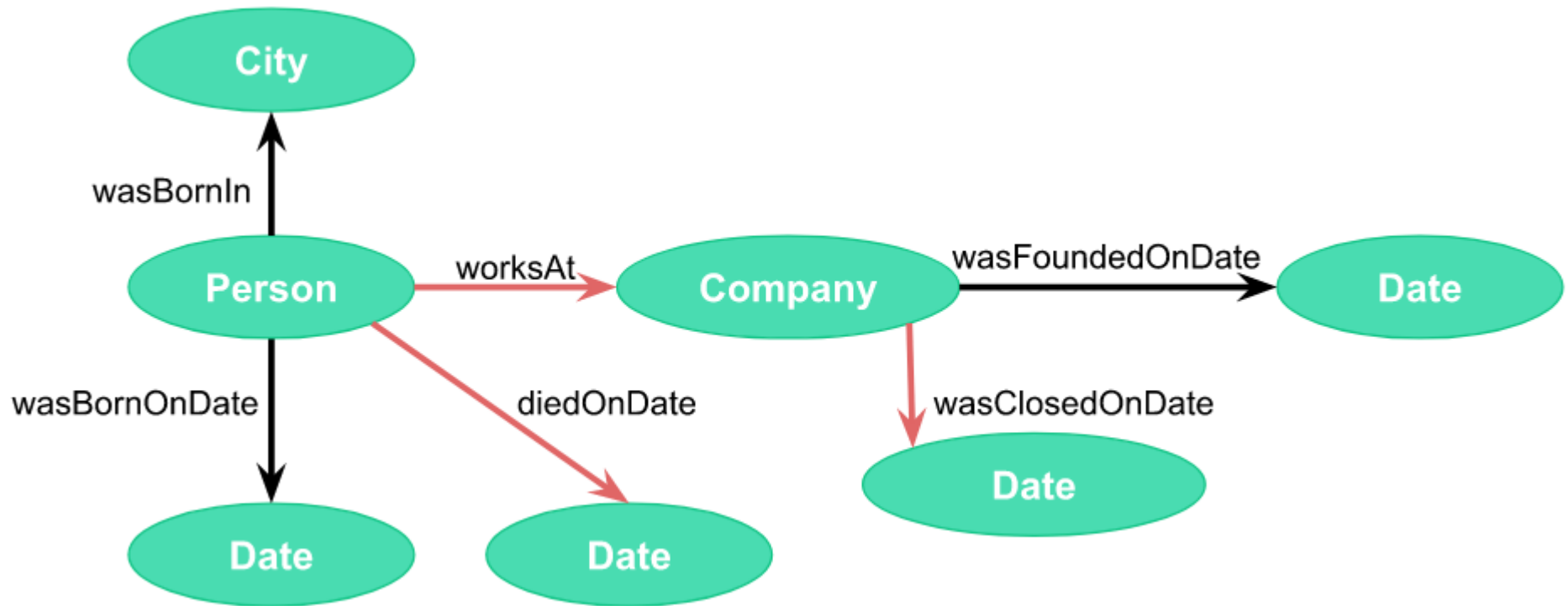
TEMPORAL INTERVAL RULE



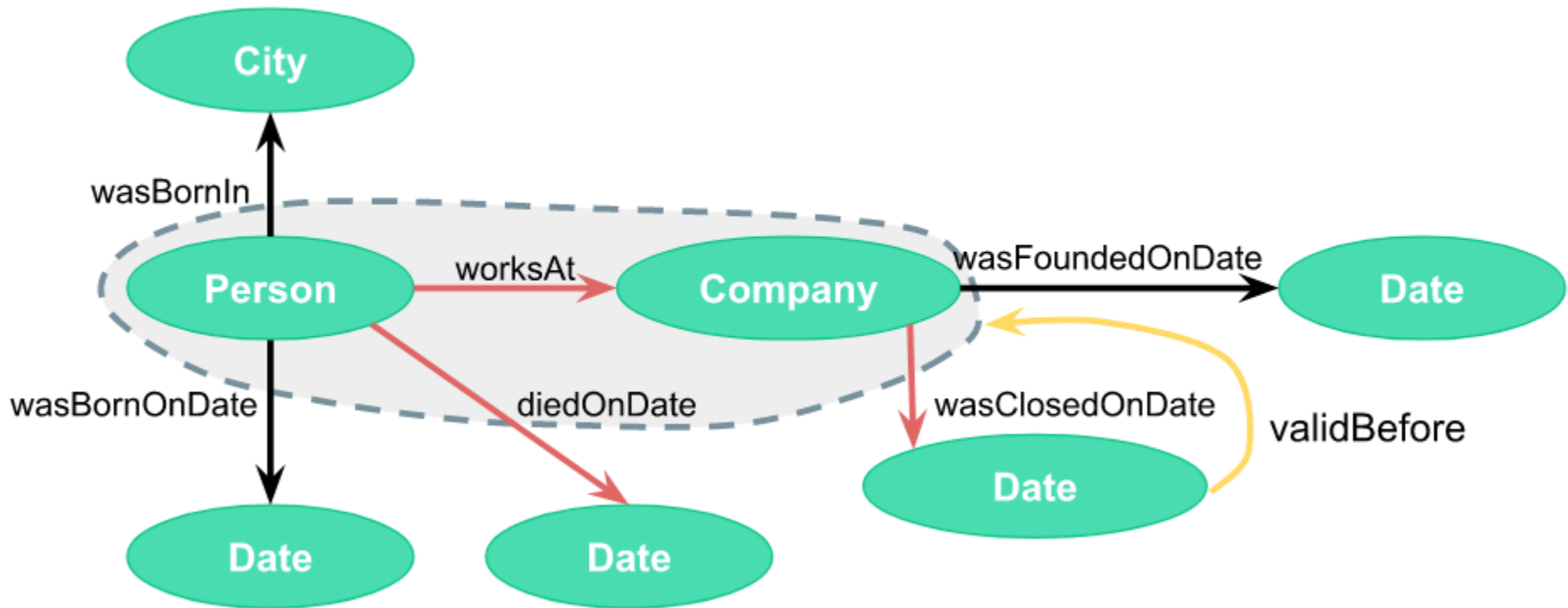
TEMPORAL INTERVAL RULE



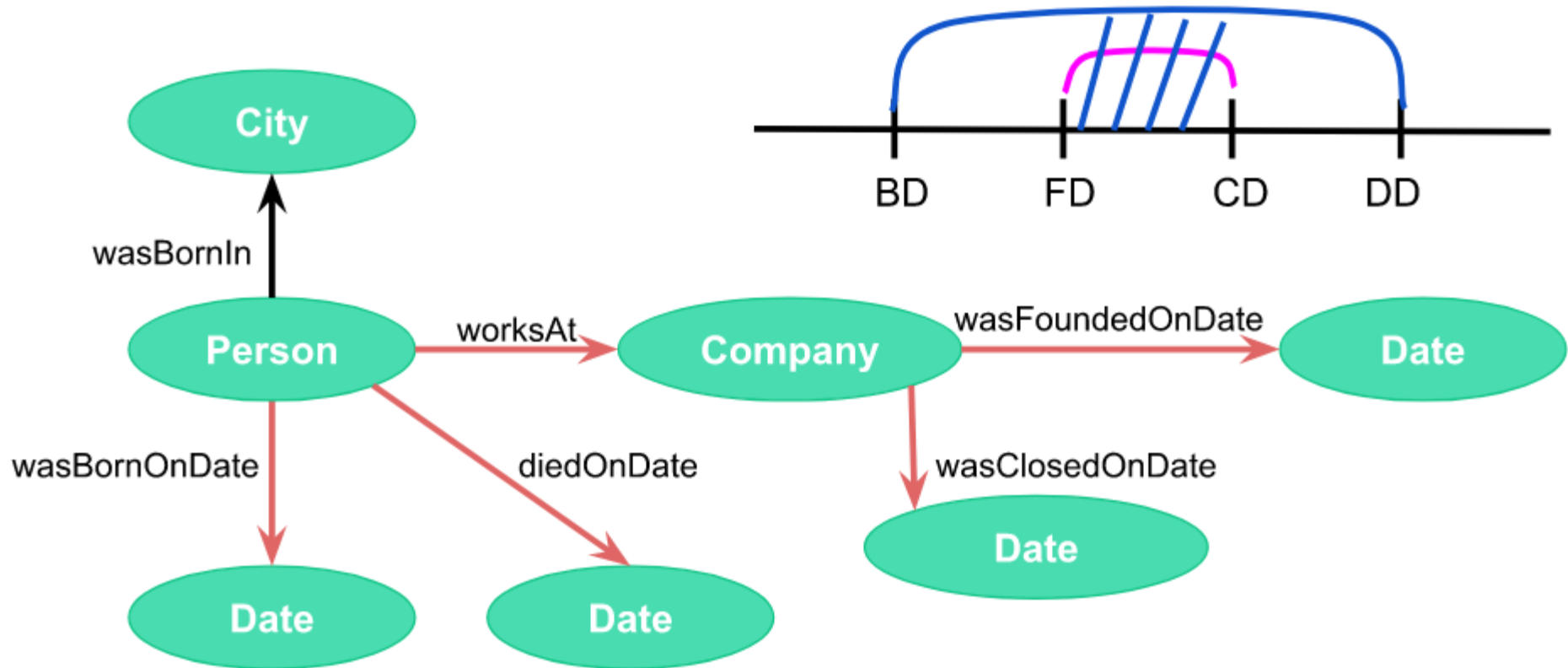
TEMPORAL INTERVAL RULE



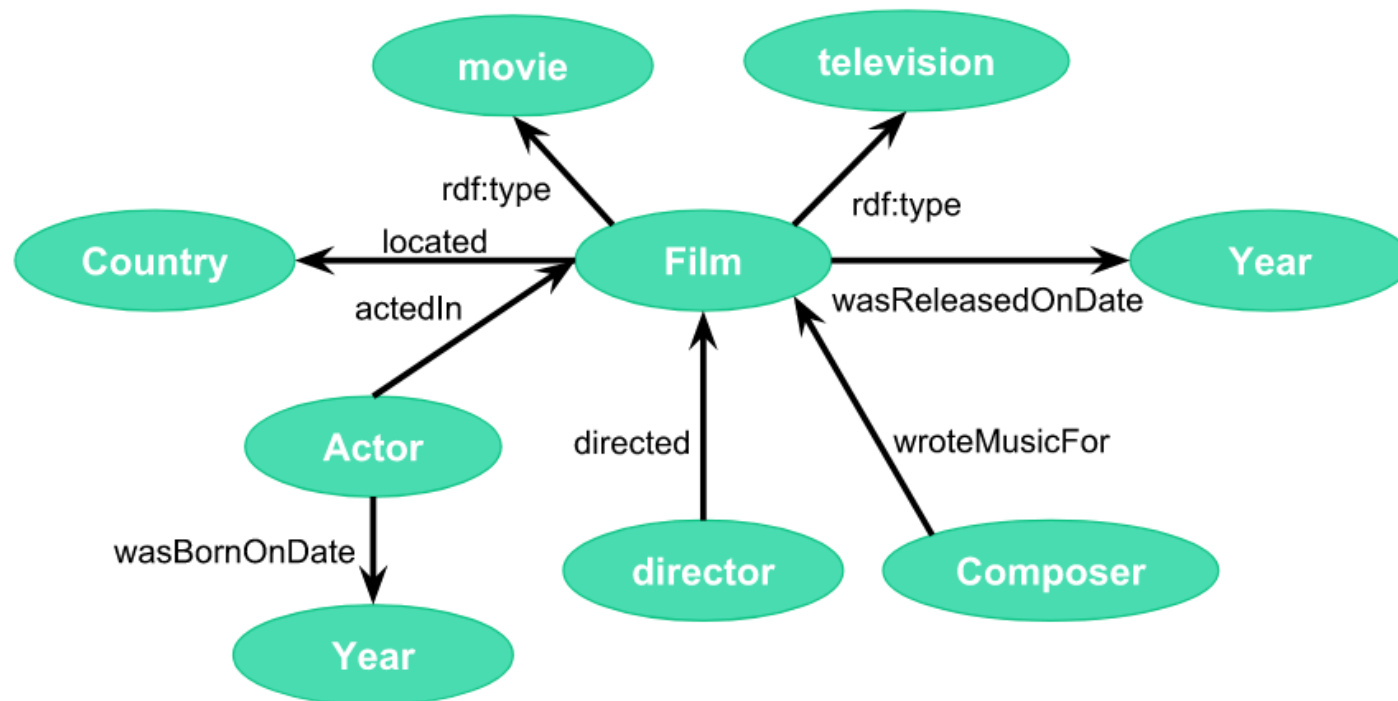
TEMPORAL INTERVAL RULE



TEMPORAL INTERVAL RULE



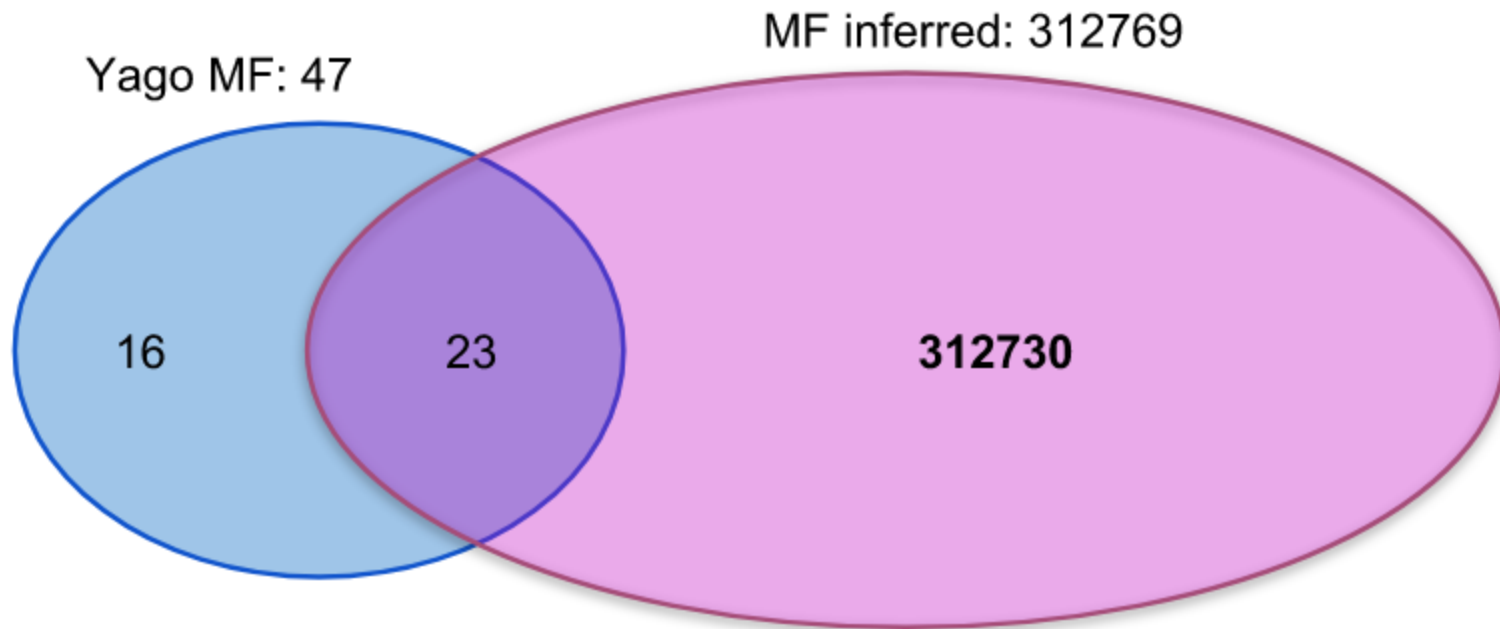
CASE STUDY: MOVIE DOMAIN (YAGO)



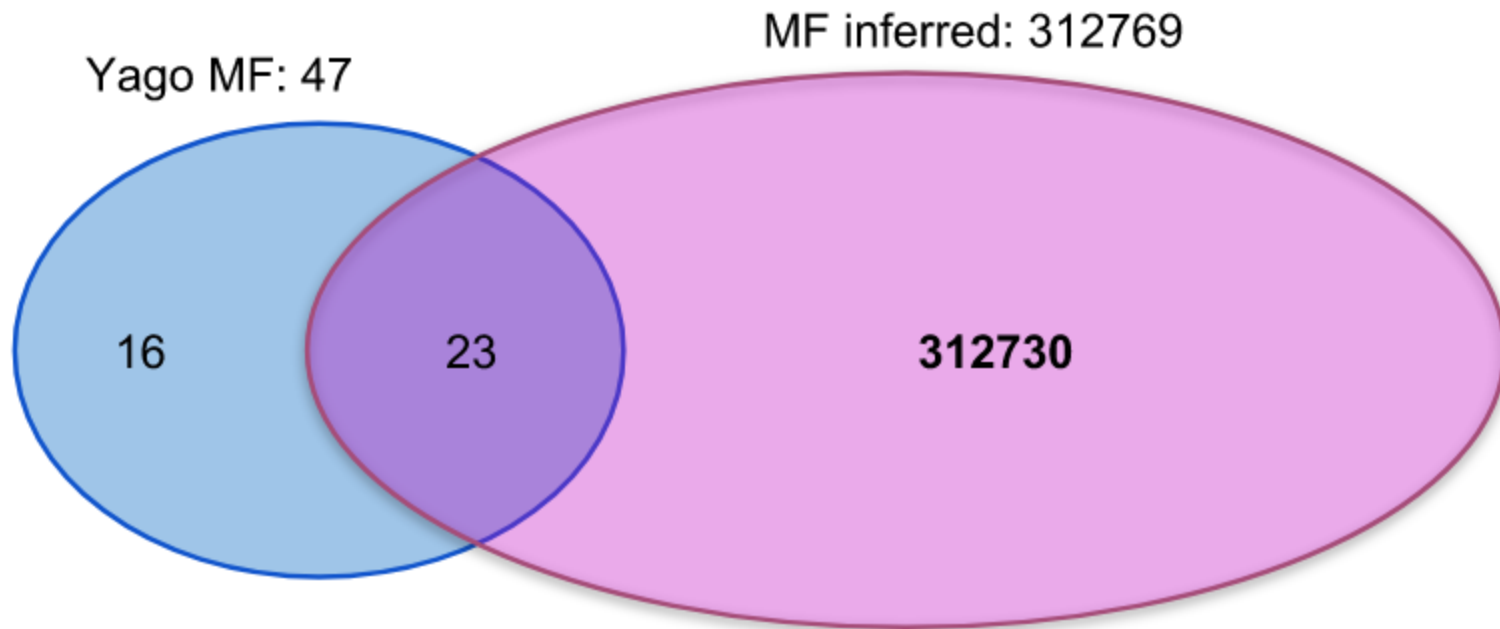
MOVIE DOMAIN (YAGO)

- Yago:
 - # of films: 151427
 - # of actors in Yago: 47800
 - # of release dates available: 136234

RESULTS: actedIn



RESULTS: actedIn



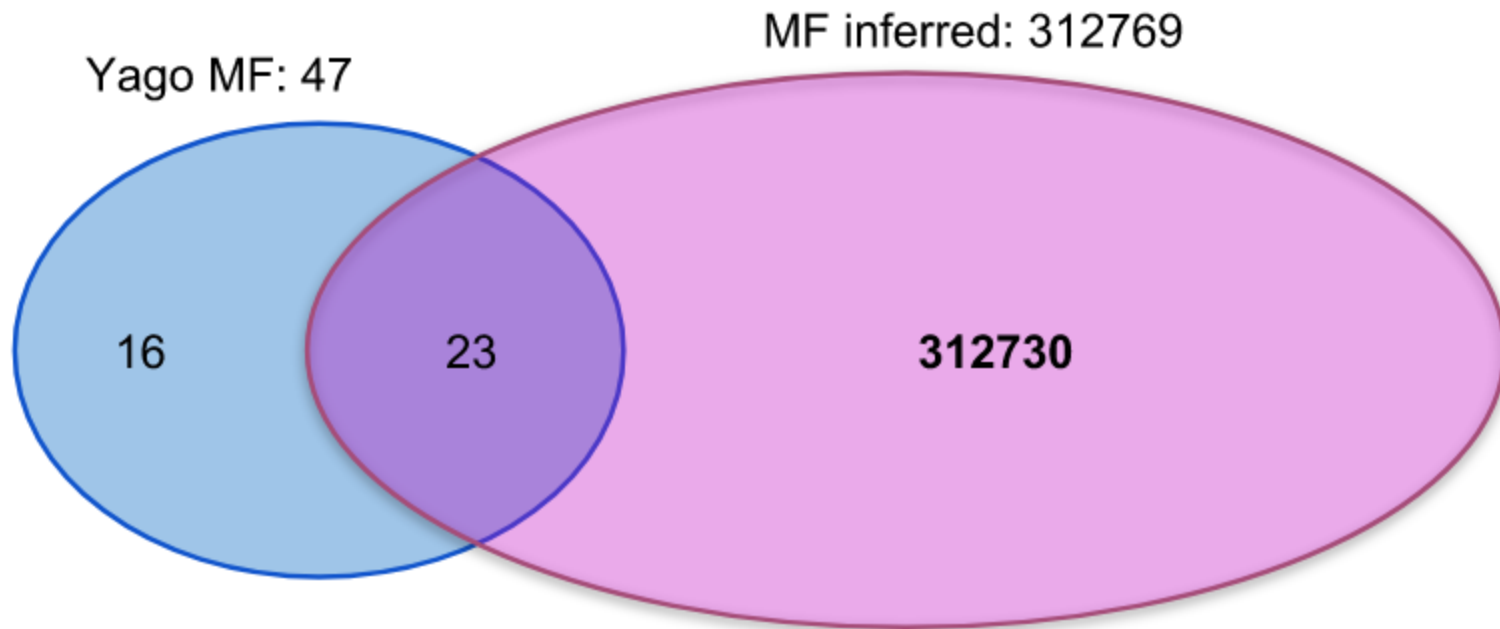
- Results vs Yago MFs & IMDb:

- 23 share the same information (Yago MFs & IMDb)
- 16 have the same IMDb information

Notice: Yago MFs are more accurate.

- 8 MFs not inferred because some facts don't have release date associated.

RESULTS: actedIn



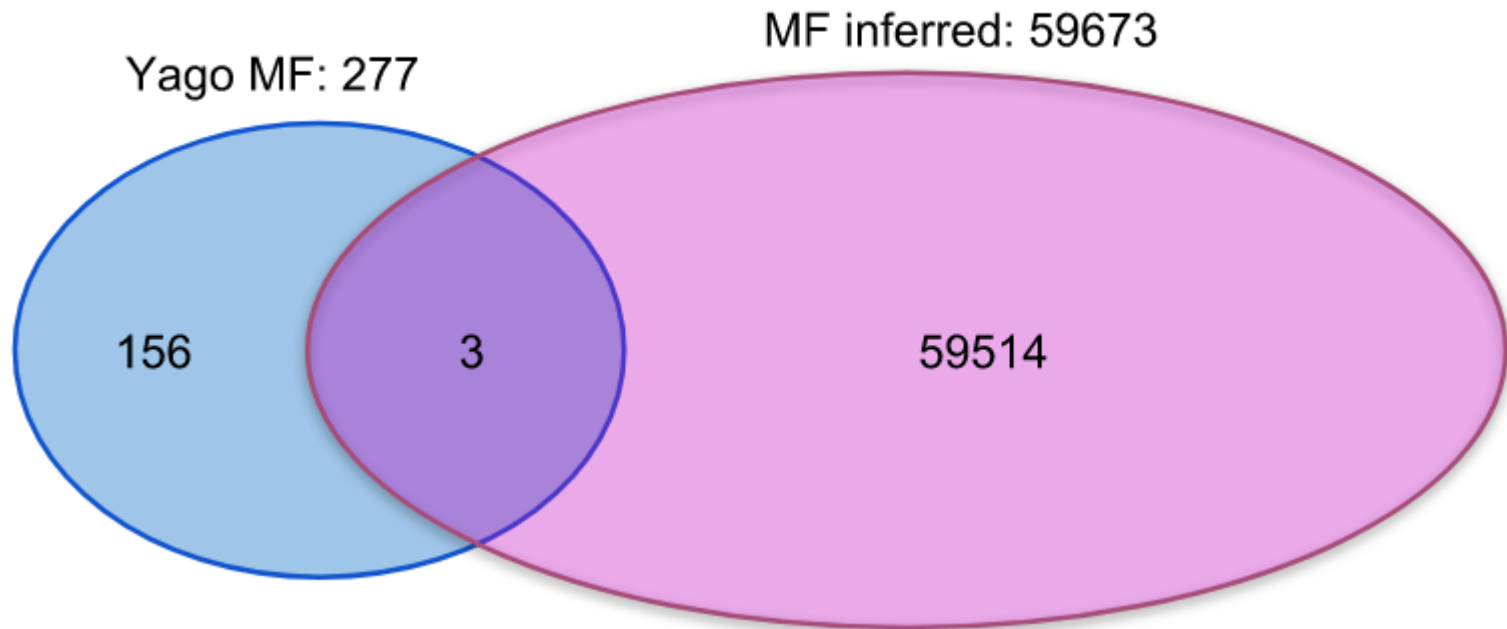
- Results vs Yago MFs & IMDb:

- 23 share the same information (Yago MFs & IMDb)
- **16** have the same IMDb information

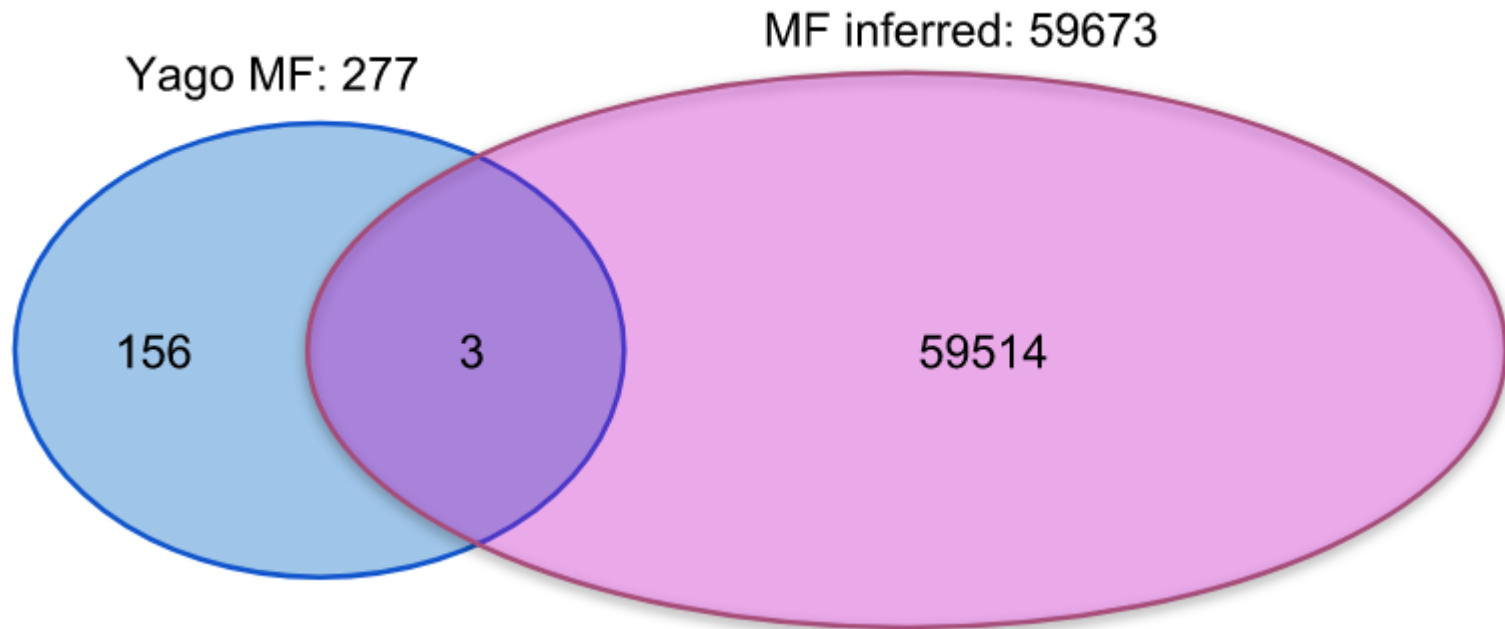
Notice: Yago MFs are more accurate.

- 8 MFs not inferred because some facts don't have release date associated.

RESULTS: wroteMusicFor



RESULTS: wroteMusicFor



- Results vs Yago MFs & IMDb:
 - 3 share the same information (Yago MFs & IMDb)
- 118 MFs not inferred because some facts don't have release date associated.

ONGOING WORK:

- Qualitative evaluation of the meta facts inferred.
- Creating new set of rules.
- Extend the approach to reason more globally on the whole graph while inferring meta facts.

FUTURE WORK:

- Spatial reasoning.
 - E.g.: Film release dates are associated to specific locations (country, city).
- (Semi-)automatic approach for rule generation.

MERCI !