

# Introduzione al web semantico

Carlo Meghini

CNR ISTI

Angelica Lo Duca

CNR IIT

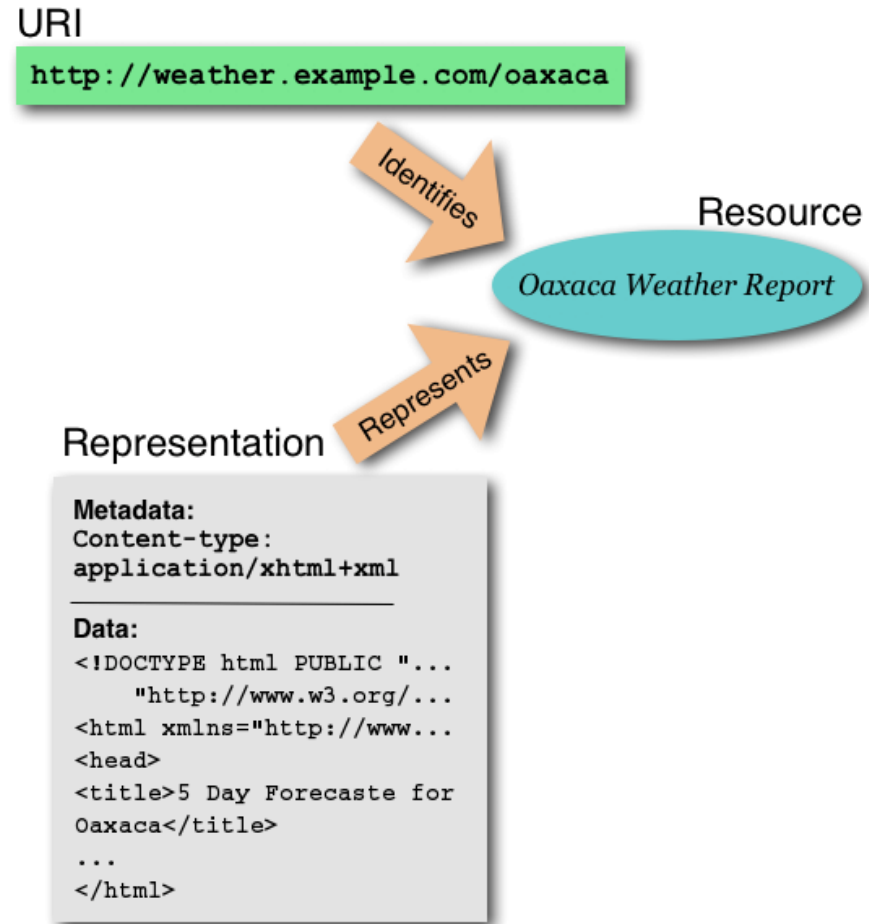
# Il Web

- Il World Wide Web ha cambiato il modo in cui comunichiamo e in cui conduciamo le nostre attività.
- È al centro di una rivoluzione che sta trasformando il mondo sviluppato verso una economia della conoscenza e, più in generale, verso una società della conoscenza.

# Il web come base di conoscenza

- Una base di conoscenza informalmente espressa (testo, immagini, video, grafica) contenuta in ipertesti HTML.
- Un meccanismo per accedere alla conoscenza basato sul protocollo HTTP:
  - HTTP Client: browser
  - HTTP Server: una parte del Web server
- Concettualmente, il web si base su tre nozioni:
  - risorsa: tutto ciò che ha una identità chiara e distinta
  - IRI: identificatori di risorsa
  - rappresentazione: dati che codificano informazione riguardo allo stato di una risorsa
- Ogni IRI identifica univocamente una risorsa, e una risorsa può avere zero, una o più rappresentazioni.

# Il web come base di conoscenza



# Il Web è per gli umani

- La maggior parte dei contenuti del web sono per gli umani. Essi incorporano conoscenza che richiede l'apparato cognitivo e mentale degli umani per essere estratta, compresa e utilizzata.
- Anche i contenuti che sono generati automaticamente prelevando le informazioni da database, sono presentati senza l'originaria struttura, per essere più appetibili per l'utente umano.

# Uso del web

- L'uso tipico del web comprende la ricerca e l'accesso alle conoscenze, direttamente dalle pagine del web o in documenti che sono raggiungibili da pagine web.
- Una volta ottenute le conoscenze cercate, l'utente le usa per svolgere una qualche attività, tornando a usare il web quando si rendono necessarie altre informazioni, e così via.

# I motori di ricerca sul web

- L'uso del web non sarebbe possibile senza l'ausilio dei motori di ricerca, che ci propongono le pagine di nostro interesse tra miliardi di pagine disponibili in centinaia di lingue diverse.

# Si può fare di più o di meglio?

- Una domanda che è stata presente nella mente degli inventori del web, e di Tim Berners-Lee in particolare, è se non sia possibile utilizzare il web in modo più esteso, supportando in maniera migliore o maggiore le attività che vengono quotidianamente svolte dalle persone.



# Approccio evolutivo

- L'approccio evolutivo prevede il graduale miglioramento delle prestazioni del web a partire dal web iniziale:
  - motori di ricerca più efficaci
  - supporto a un ventaglio più ampio di tipi di dato
  - supporto a un ventaglio più ampio di app
- Questo approccio ha funzionato molto bene, trasformando il web iniziale da una piattaforma di scambio di *documenti* a una piattaforma di scambio di *servizi*

# Approccio rivoluzionario

- C'è però anche un approccio rivoluzionario, che parte dalla considerazione che per migliorare in modo significativo l'uso del web sia necessario rendere accessibile alle macchine il *significato* della conoscenza espressa nel web.
- Se un agente artificiale potesse interpretare correttamente le informazioni che trova nelle pagine web, l'agente potrebbe svolgere molti compiti che gli sono attualmente preclusi.

# L'accesso al significato

- Le tecniche di estrazione del senso dal testo, dalle immagini, dai video sono molto migliorate negli ultimi anni, grazie alla disponibilità di risorse computazionali sempre più potenti a costi sempre più accessibili.
  - Deep learning realizzato con reti neurali sempre più sofisticate, grazie alle maggiori capacità di calcolo delle CPU attuali

# L'accesso al significato

- Nonostante questi progressi, siamo ancora lontani però dalla possibilità di carpire algoritmicamente il senso del testo o delle immagini.
- Soprattutto, risulta difficile costruire in modo automatico la rete dei significati che connette fra di loro i termini del linguaggio naturale
  - Il significato è olistico

# Il web semantico

- L'unica alternativa al momento è di rappresentare il contenuto del web in una forma che renda il senso di tale contenuto direttamente accessibile alle macchine, usando tecniche intelligenti per fare uso di tale contenuto.
- Questo approccio rivoluzionario viene indicato con il termine di *web semantico*.
- Il web semantico è dunque innanzitutto una visione, che si affianca a quella del web attuale, e che mira a rendere il significato della conoscenza espressa sul web accessibile alle macchine.

# Il World Wide Web Committee

- L'organismo impegnato nella realizzazione della visione del web semantico è il World Wide Web Consortium (W3C), un organismo di standardizzazione internazionale per il Web.
- Gli standard necessari alla realizzazione del web semantico **non** sono definiti dal W3C, ma dagli *esperti* che possiedono le competenze necessarie.
- Tali esperti provengono dal mondo della ricerca e della tecnologia.

# Il ruolo del W3C

- Il ruolo del W3C è di
  - *selezionare* gli esperti
  - *organizzare e supportare* il processo di definizione
  - *assicurare* che gli standard definiti siano
    - conformi ai principi ispiratori e di sviluppo del web
    - adeguati alle aspettative della comunità che dovrà poi usare tali standard
    - aggiornati o eventualmente sostituiti da altri standard quando diventino obsoleti
    - preservati, rendendo i relativi documenti di specifica accessibili per un periodo di tempo il più lungo possibile

# Scenario: B2C e-commerce

- Il commercio elettronico Business-to-Consumer (B2C) è una delle esperienze più comuni degli utenti del web.
- Nello scenario tipico, un utente sceglie un negozio on-line, ne consulta il catalogo, valutando le offerte, e infine scegliendo i prodotti, ordinandoli e pagandoli.



# Scenario: B2C e-commerce

- Per essere sicuro di ottenere il miglior rapporto qualità-prezzo, l'utente dovrà ripetere queste operazioni su diversi negozi on-line.
- Dovrà farlo *manualmente*, ogni volta identificando l'indirizzo del negozio e quindi visitando il negozio.
- Per facilitare l'utente, esistono agenti software che accedono ai siti dei diversi negozi, estraggono informazioni sui prodotti e relativi prezzi, e offrono agli utenti una panoramica del mercato.
- Questi agenti operano usando i cosiddetti “wrappers”, programmi che accedono ai negozi on-line e catturano l'informazione cercata interagendo con il negozio.
- Ogni negozio necessita di un wrapper diverso, perché ogni negozio offre le informazioni in modo diverso dagli altri.
  - Approccio costoso e soggetto ad errori, perché spesso l'informazione è estratta analizzando il testo in linguaggio naturale delle pagine web dei negozi.

# Scenario: B2C e-commerce

- La soluzione offerta dal semantic web risolverebbe il problema:
  - Ogni negozio on-line esporrebbe le proprie informazioni usando un linguaggio formale, lo stesso per tutti i negozi
  - Gli agenti software troverebbero le informazioni cercate rappresentate sempre allo stesso modo, codificate secondo un linguaggio dal significato chiaro e conosciuto.

# Quale linguaggio?

- Ma quale potrebbe essere questo linguaggio?
- La logica matematica si propone come strumento ideale per questo ruolo:
  - Offre un linguaggio formale che cattura gli aspetti essenziali del linguaggio naturale
  - Il linguaggio è dotato di una specifica del significato, o semantica, ben formalizzata, su cui si definiscono operazioni molto utili al caso:
    - rispondere a domanda (deduzione)

# Quale linguaggio?

- I primi studi sull'uso della logica per la rappresentazione dell'informazione in forma digitale risalgono agli anni 50 del secolo scorso, nell'ambito dell'Intelligenza Artificiale.
  - L'obiettivo era ricreare le funzioni della mente in un software.
- Questo obiettivo non è ancora stato raggiunto, ma si sono raggiunti obiettivi parziali molto utili al caso nostro:
  - scoperta di logiche matematiche pensate per la descrizione di oggetti del mondo reale
  - dotate di proprietà computazionali positive, prime fra tutte la decidibilità dei problemi principali, e tra questi quello della domanda-risposta
- Queste sono le logiche descrittive (Description Logics), studiate a partire dalla metà degli anni 80 e oggi ben comprese e implementate

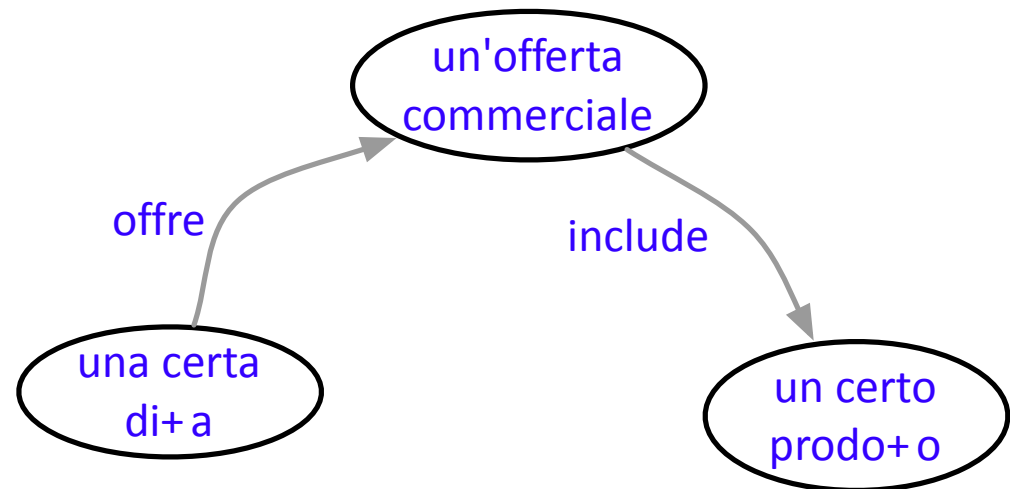
# Le logiche descrittive

- Uno degli standard più importanti del web semantico, l'Ontology Web Language (OWL), è una logica descrittiva.
- I sotto-linguaggi di OWL pensati per specifiche applicazioni (OWL Profiles) sono anch'essi logiche descrittive.
- Tutte le logiche descrittive più comuni sono implementabili, anche se non tutte lo sono in maniera efficiente.
  - Per le logiche più espressive, il problema della domanda-risposta può richiedere in qualche caso tempi lunghi di elaborazione, ma questa è una limitazione naturale, che si riscontra anche nelle basi di dati.

# B2C e-commerce

- L'informazione che ci interessa rappresentare può essere schematizzata in due semplici frasi:

1. Una certa ditta (*soggetto*)  
offre (*predicato*)  
una certa offerta (*oggetto*)
2. una certa offerta  
include  
un certo prodotto



Ogni frase ha tre parti, e per questo è detta *tripla*

# Mettiamo qualche nome

Tutte le entità che compaiono nelle nostre triple sono dette risorse. Per il web, il mondo è popolato di risorse.

Le risorse sono identificate da nomi speciali: IRI (International Resource Identifier).

- <http://dhtools.labcd.unipi.it/programma/>

Gli IRI si possono abbreviare *pref:nome*

Usando il prefisso aiucd per la prima parte dell'IRI:

- aiucd = http://dhtools.labcd.unipi.it/

l'IRI si può scrivere in forma più leggibile

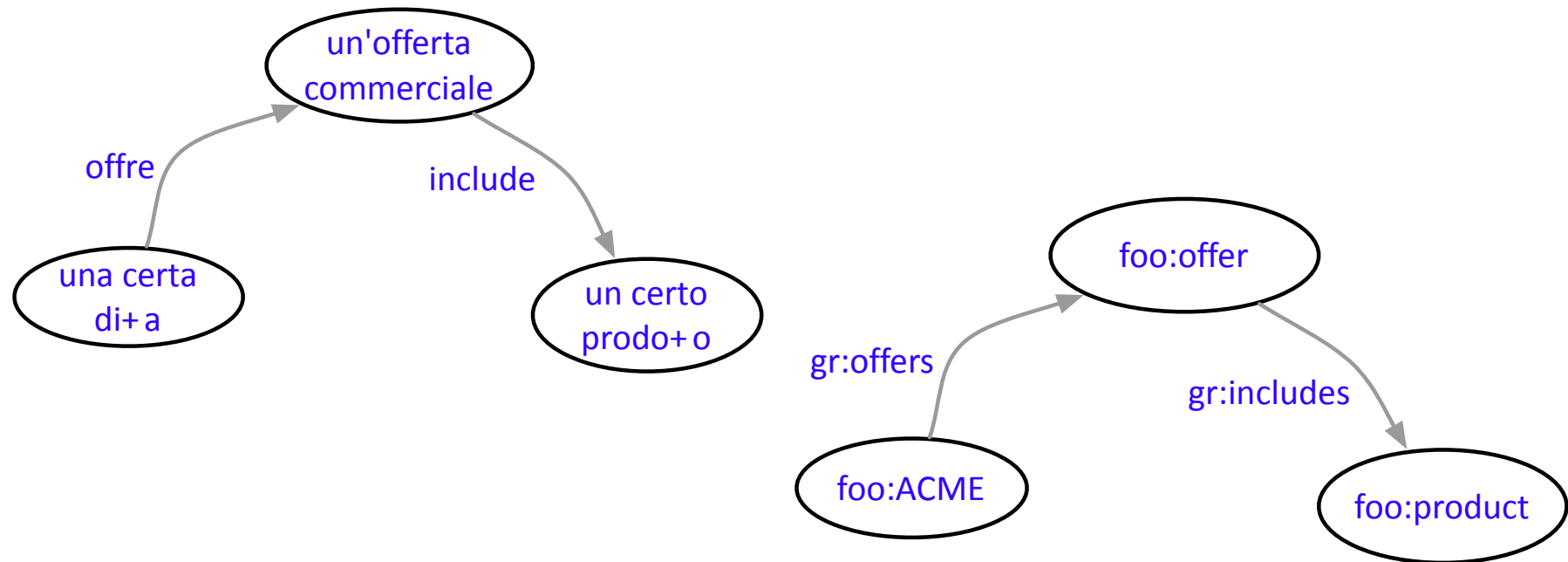
- aiucd:programma/

Il prefisso indica il vocabolario da cui provengono i nomi.

# Mettiamo qualche nome

foo: il nostro vocabolario giocattolo

[gr: Good Relations ontology](#)





# In forma machine-readable

```
@prefix foo: <http://example.com/> .
```

```
@prefix gr: <http://purl.org/goodrelations/v1#> .
```

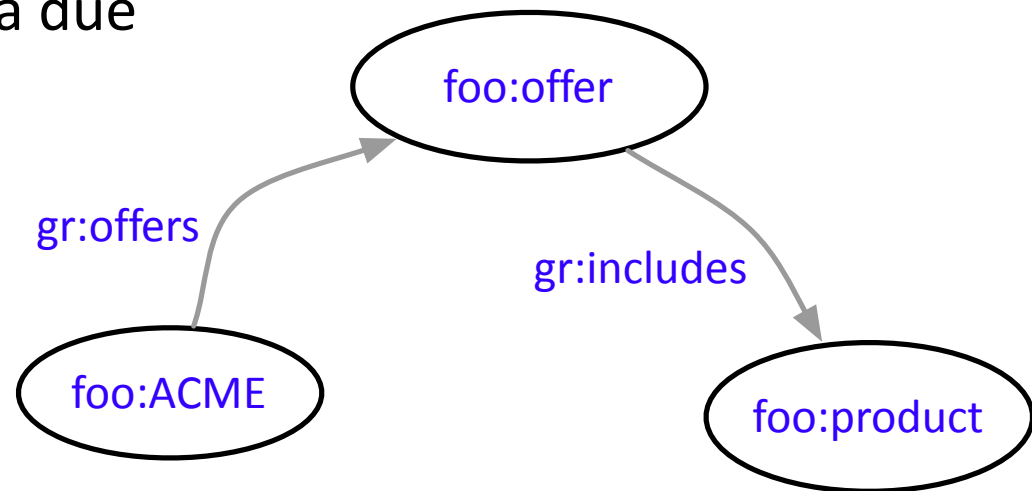
```
foo:ACMECorp gr:offers foo:Offer .
```

```
foo:Offer gr:includes foo:product .
```

Un grafo RDF, composto da due dichiarazioni di prefisso e da due triple.

La notazione è Turtle.

Altra notazione: XML



# Completiamo l'informazione

```
@prefix foo: <http://example.com/> .
@prefix gr: <http://purl.org/goodrelations/v1#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix pto: <http://www.productontology.org/id/> .
```

```
foo:ACMECorp a gr:BusinessEntity .
foo:ACMECorp gr:legalName "ACME Corp" .
foo:ACMECorp gr:offers foo:Offer .
```

```
foo:Offer a gr:Offering ;
    gr:includes foo:product;
    gr:validFrom "2017-01-24T00:00:00+01:00"^^xsd:dateTime ;
    gr:validThrough "2017-12-24T00:00:00+01:00"^^xsd:dateTime ;
    gr:hasPriceSpecification
    [ a gr:UnitPriceSpecification ;
      gr:hasCurrency "USD"^^xsd:string ;
      gr:hasCurrencyValue "19.99"^^xsd:float ;
      gr:validThrough "2011-12-24T00:00:00+01:00"^^xsd:dateTime ] .
```

```
foo:product a <http://www.productontology.org/id/Hammer> ;
    a gr:ClawHammer ;
    gr:name "martello da carpentiere"@it ;
    gr:description "specifico per l'inserimento dei chiodi (con profilo presente sulla bocca), con un'estremità
    biforcuta per l'estrazione"@it .
```

# Vantaggi

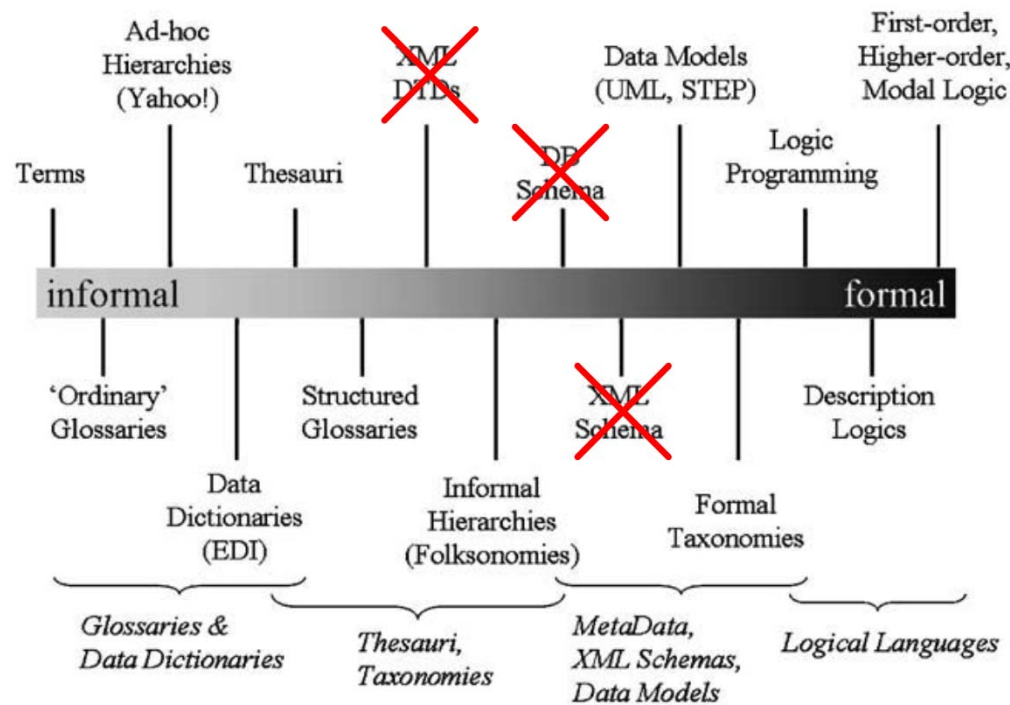
- L'informazione sul prezzo e sul prodotto viene estratta correttamente,
- senza ricorrere a tecniche di basso livello come i wrapper
- Si può completare l'informazione cercando su altre fonti, per esempio il rating di ACME
- Si possono progettare agenti sofisticati in grado di condurre negoziati con i venditori per conto degli acquirenti.

# Principio cardine

- Gli agenti “capiscono” le descrizioni perché le descrizioni sono espresse in un'ontologia accessibile e quindi condivisa.
- Un'ontologia è un vocabolario.
- Fornisce esplicite definizioni dei termini, rimuovendo quanto più possibile ogni ambiguità.

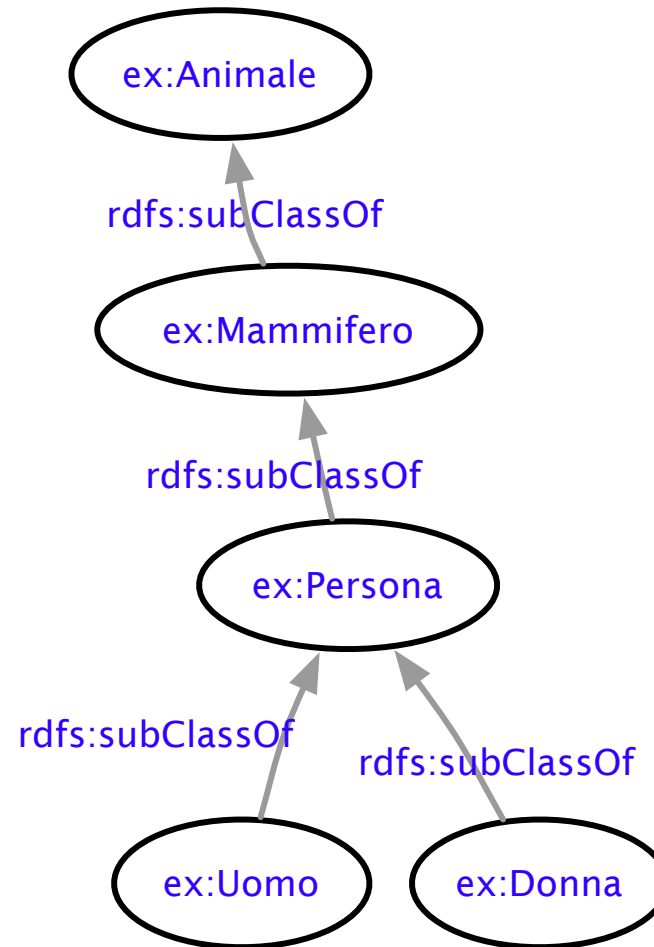
# Ontologia

- Ci sono vari tipi di ontologie, dai più semplici ai più complessi:



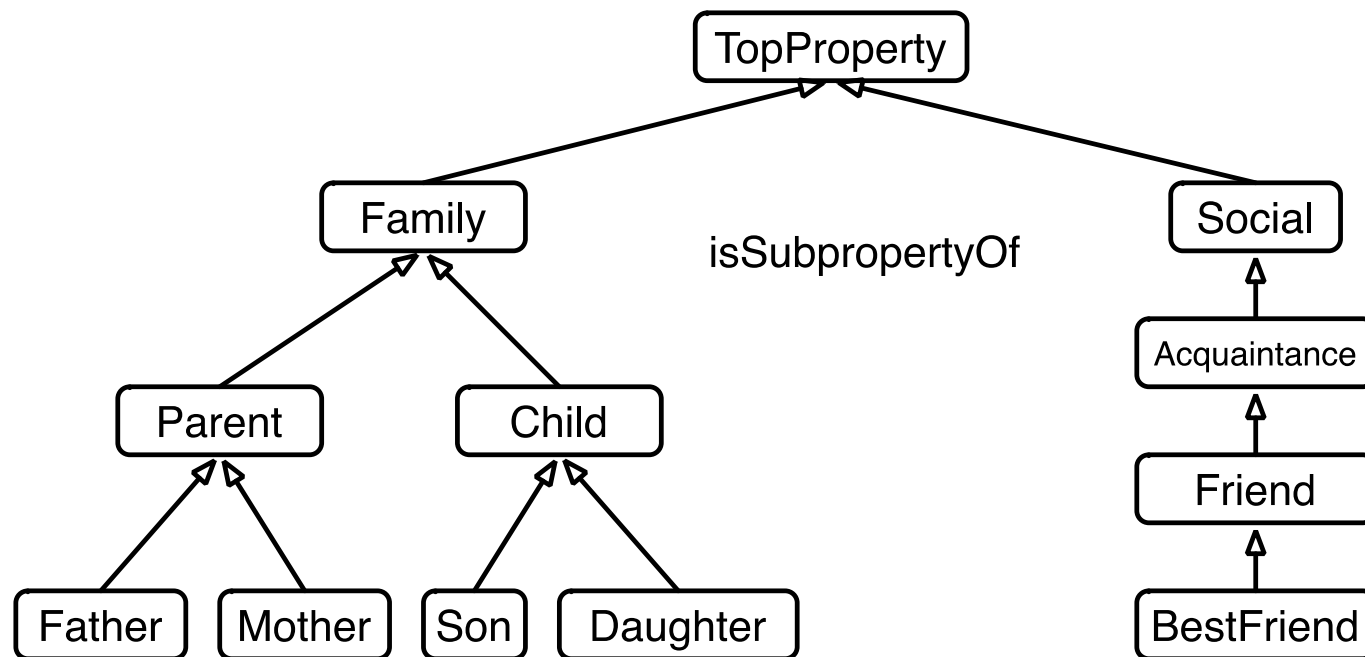
# Ontologie

Classi e loro  
relazioni  
tassonomiche



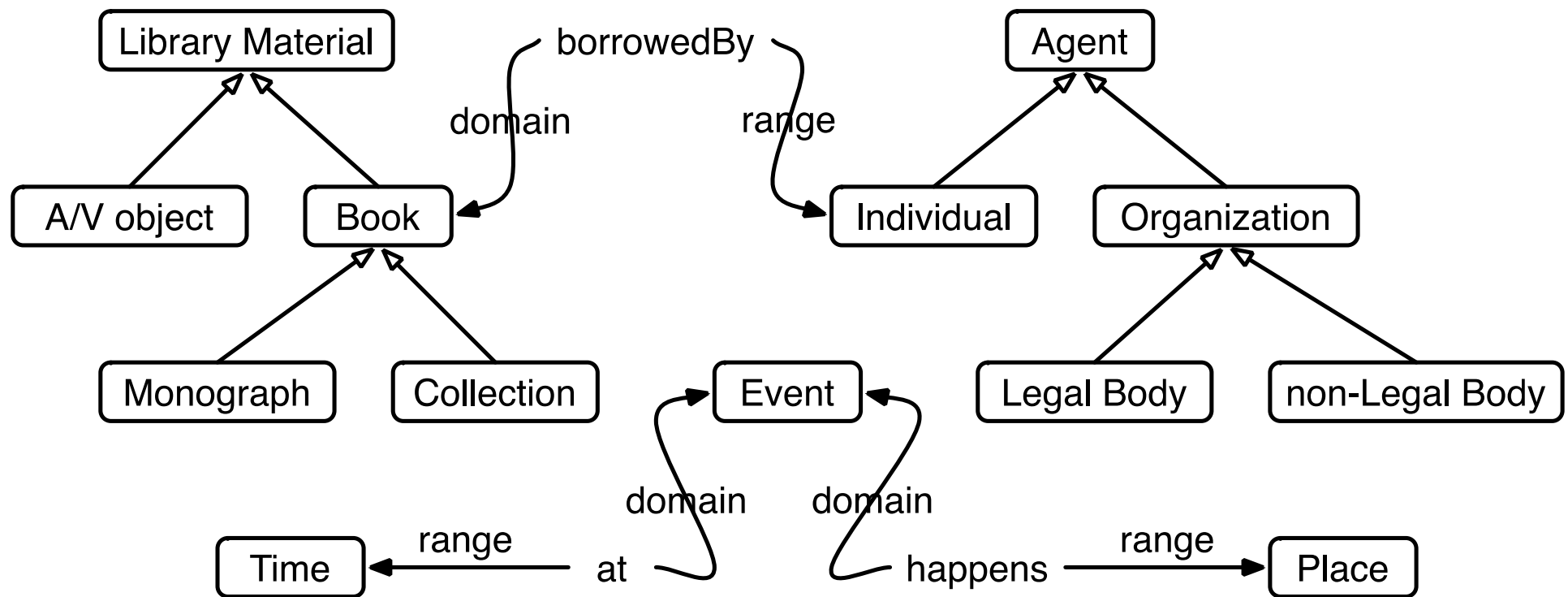
# Ontologie

Proprietà e loro relazioni tassonomiche



# Ontologie

## Proprietà e loro definizione

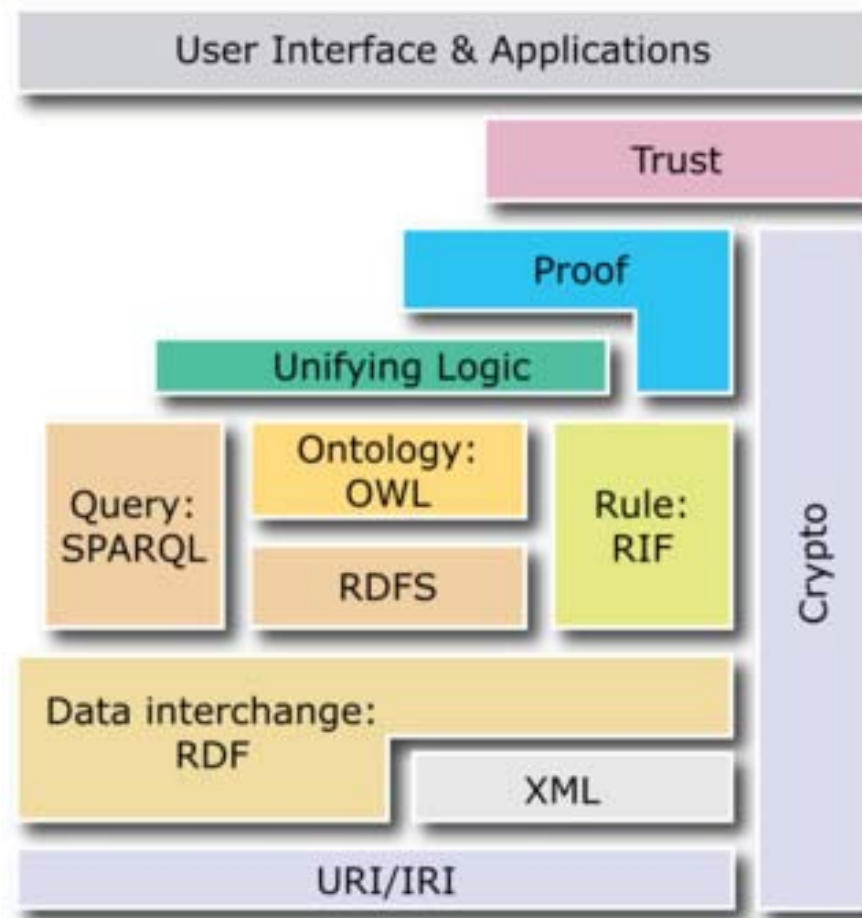




# Ontologia delle buone relazioni commerciali

- Vedi esempio.

# Il quadro completo



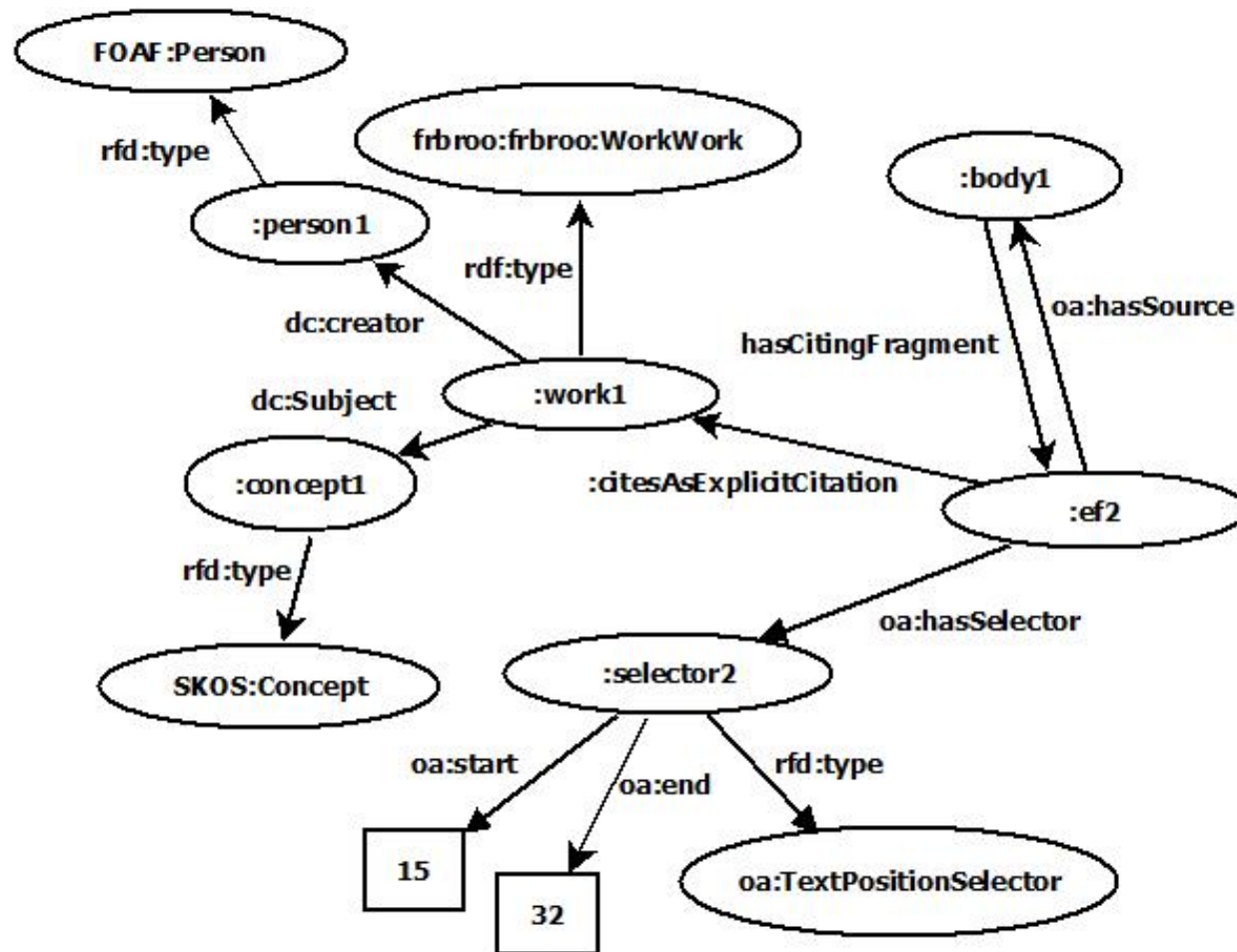
# Applicazione alle DH

- Possibile?
  - *“Ciò che può essere detto può essere detto chiaramente”*
- Come?
  - Definendo ontologie che diano conto dei termini che appartengono al discorso delle DH.
    - Interoperabilità = riuso di ontologie esistenti
  - Usando tali ontologie per costruire rappresentazioni degli oggetti, degli eventi e dei processi rilevanti
  - Linkando tali rappresentazioni ad altre, analoghe, e così formando la rete di conoscenze che “mima” il web a livello formale

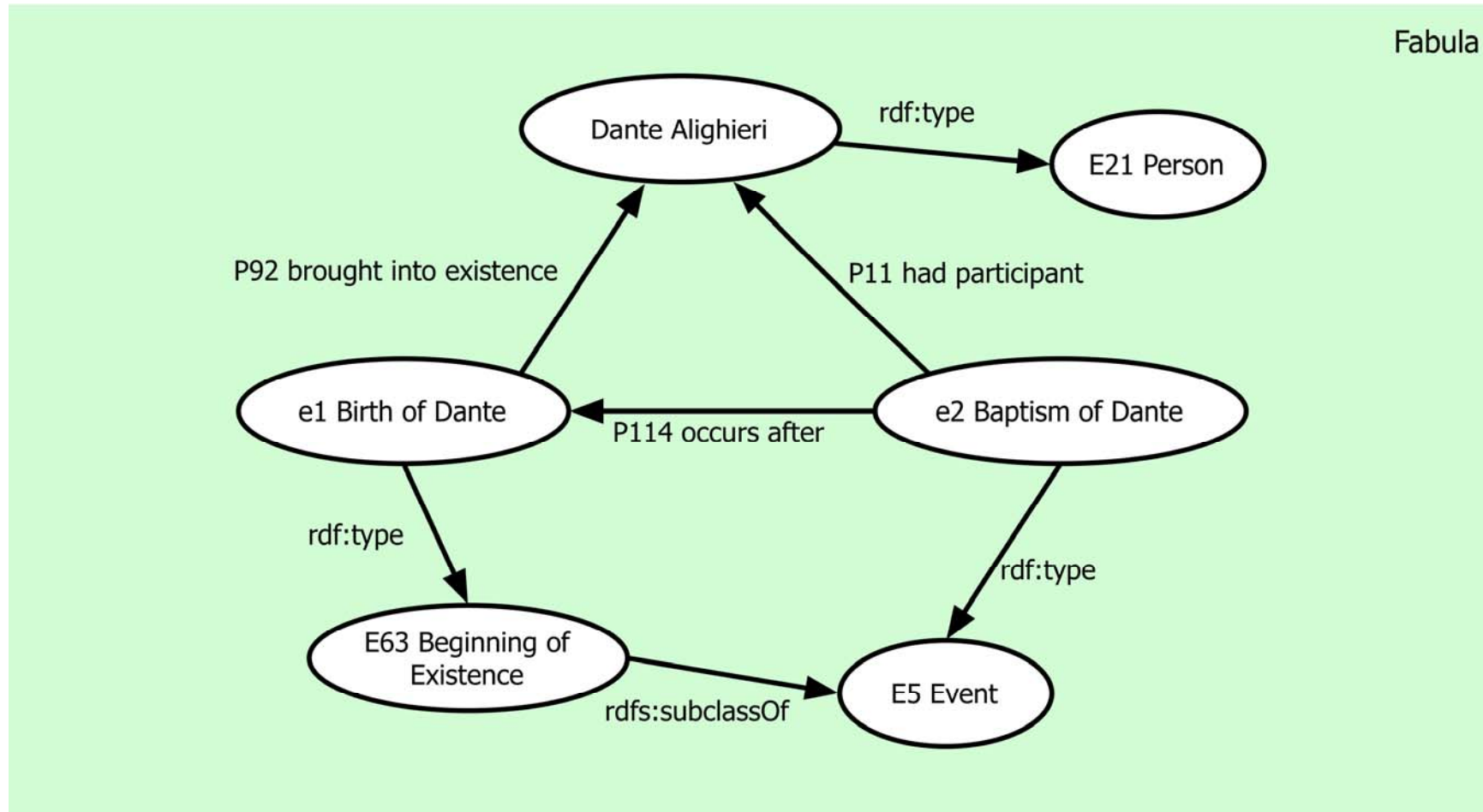
# Due esempi

- [Dante Sources](#)
- [Narratives](#)

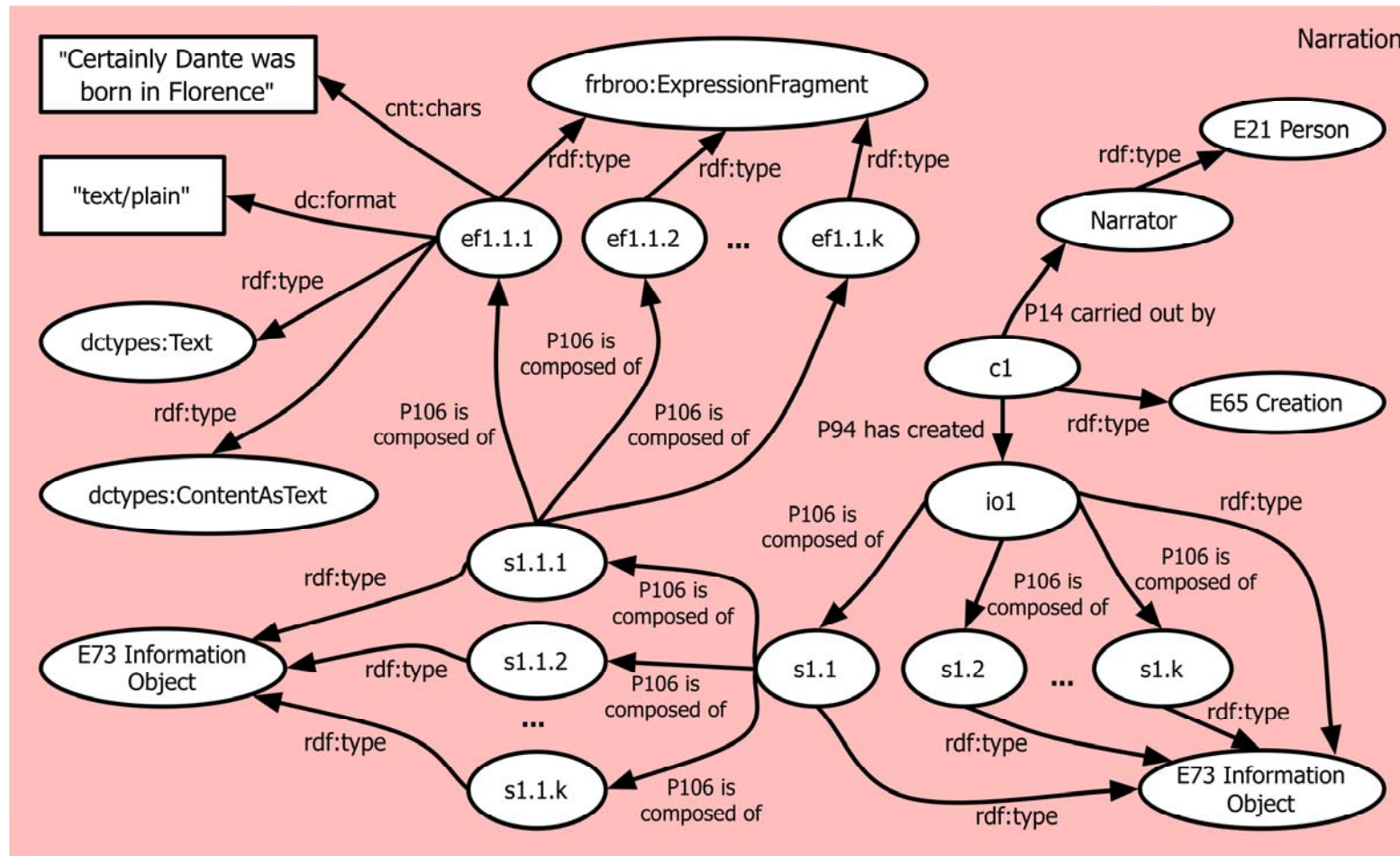
# Semplice rappresentazione delle citazioni alle fonti primarie



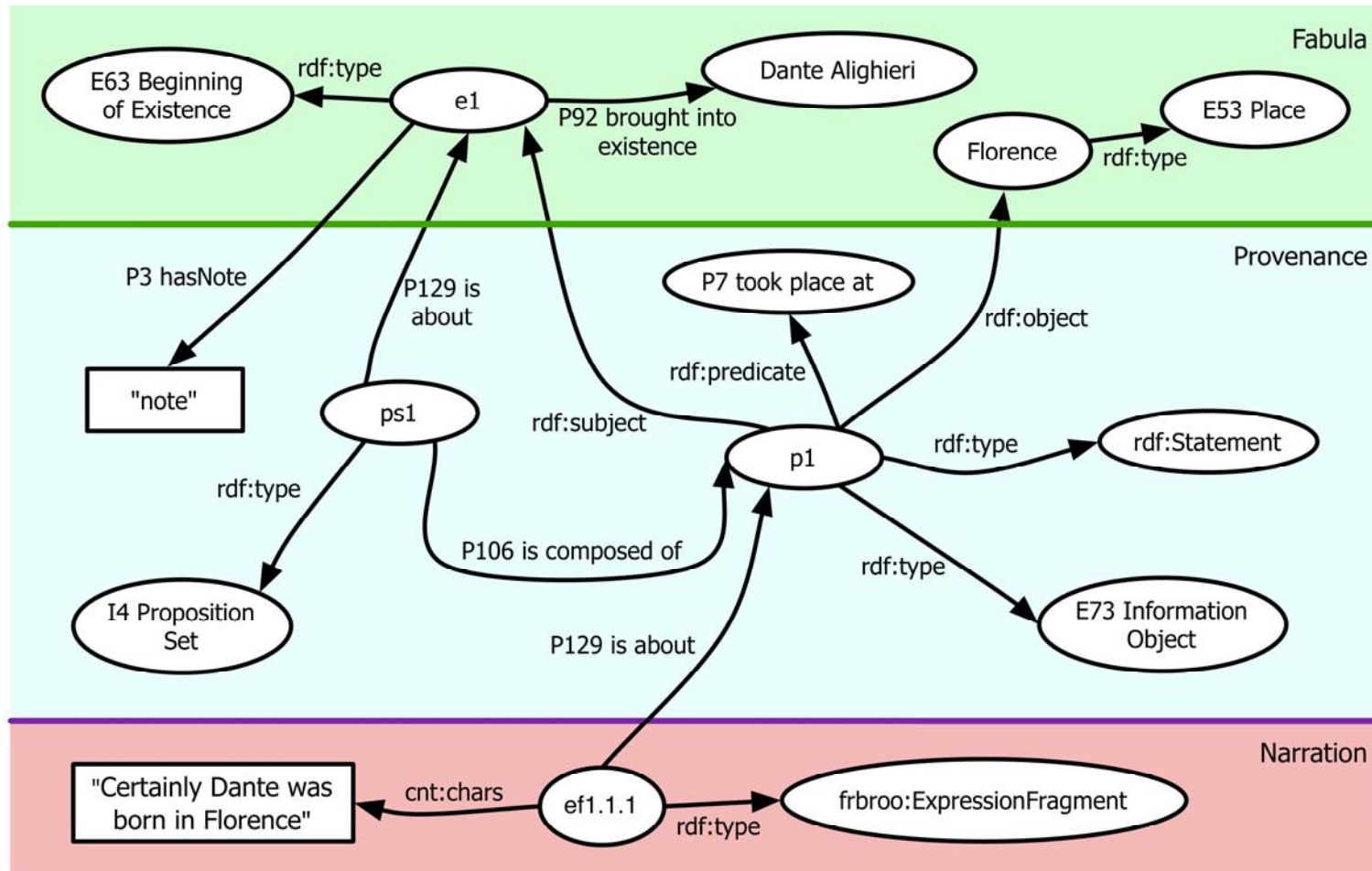
# Semplice rappresentazione di una narrazione: Fabula



# Semplice rappresentazione di una narrazione: Narrazione



# Semplice rappresentazione di una narrazione: Reference function





# Conclusioni

- Il web semantico è una visione per condividere conoscenza formalmente espressa così come il web permette la condivisione di conoscenza espressa nei linguaggi naturali
- Grazie agli standard sviluppati sotto il coordinamento del W3C, la visione si sta realizzando.
- Contributo fondamentale: Linked Data