Narratives in Digital Libraries

Carlo Meghini, Valentina Bartalesi, Daniele Metilli, Filippo Benedetti

ISTI-CNR, Via Moruzzi,1 - 56124 Pisa

Outline of the Presentation

1. Research question
2. The proposal
3. Study of the Narratology and Artificial Intelligence literature
4. The methodology
5. The ontology for narratives
6. The tool for bulding and visualising narratives
7. The software architecture
8. The Web interface for visualising the knowledge
9. A qualitative evaluation of the developed work
10. Conclusions and future work

The Problem We Started From

- Digital Libraries (DLs) are **information systems that offer information services over large sets of digital objects**
- The basic information service of a contemporary DL is essentially the same as that of a traditional library: to **support users in discovering the digital objects** that satisfy an information need, typically expressed as a query consisting of a short list of terms
- **Limitation** of the **informative services** offered to the user by the current Digital Libraries → result given to the query is a **list of the information objects without semantic relations among them**

The Proposal

…the introduction of a **new search functionality** for DLs that does not return just a list of objects but a **narrative**

Narrative

a semantic network of **events,** with a possibly associated **text**, that are linked to the **digital objects** of the existing digital libraries and are endowed with a set of **semantic relations** connecting them to each other

Narratology Background

- Aristotle
- Russian formalists
- **Computational narratology** studies narratives from a computational perspective and develops narratological models → In the Artificial Intelligence field storytelling systems use narratives as interactive method. **Ontologies** are developed to give a **formal computable representation of a narrative**

Artificial Intelligence Background

Study of **Situation and Event Calculus theories** in order to:

- identify the **logical components of narratives** (e.g. events, actions, fluents, physical objects, agents)
- identify **types of events** (e.g. generalized events, processes, mental events)
- **Logical definitions:**
  - Event→ *Initiates*(e, f, t) means that the occurrence of event *e* at time *t* causes fluent *f* to become true, while *Terminates*(w, f, t) means that *f* ceases to be true
  - Mental event→ *Believes*(Lois, x)

    x represents *Flies*(Superman)

    Flies(Superman) is a proposition that specifies a mental object: reification of the mental object that *Superman flies*

Evaluation of Ontology Learning Systems

Ontology Learning (OL) is a method to create ontologies in an automatic or semi-automatic way from text analysis

Limits:

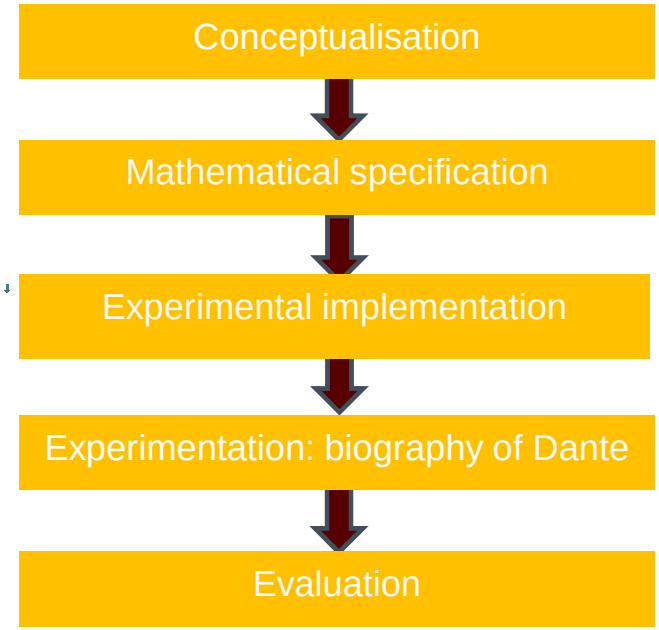
- **language dependent**
- **style dependent**
- the **quality** of the **ontologies** developed using semi-automatic methods is still **worse** than the quality of **ontologies developed by experts**

Our main goals:

- **Reliability** of collected data
- Ontology **independent from population techniques** (automatic, semi-automatic, manual)

Methodology

The creation of a **formal ontology** in order to **COLLECT** and **VISUALIZE** knowledge about **narratives**



Conceptualisation

Abstract model of the **components of narratives,** as derived from the narratology and AI literature

A **narrative** consists of three main elements:

1. the **fabula,** the sequence of the events that compose the story in chronological order
2. the **narratives,** one or more texts that narrate the fabula and its/their author/s
3. the **plot,** a **reference function** that connects the **events** of a narration with a fragment of a **text** that describes them

Fabula

The fabula is built on top of **events**

In a fabula, events are connected to each other by three kinds of relations:

1. A **mereological** relation, relating events to other events that include them as parts, *e.g.*, the birth of Dante Alighieri is part of the life of Dante
2. A **temporal occurrence** relation, associating each event with a time interval during which the event occurs. We formalize these relations between events using the Allen's temporal logic
3. A **causal dependency** relation, relating events that in normal discourse are predicated to have a *cause-effect* relation in the narrator's opinion, *e.g.*, the eruption of the Vesuvius caused the destruction of Pompeii

Narration

Each narration of a fabula consists of **one or more narrators and a text**, which is *authored by* the narrator(s) and constitutes the narration

We focus on the only aspect that is functional to our model of narrative, namely **textual content** that is the language expression that constitutes the content of a piece of text

The Plot - Reference Function

- The reference function connects **each portion of text that narrates an event to the narrated event**

- In order to model a reference we need to identify textual **narrative fragments**, each of which narrates a single event

- The reference function allows **deriving the plot** of the narrative. By visiting the text of the narration in its natural order, it is possible to access the *narrative fragments* and, via these, the events in the fabula, **in the order established by the narrator**

A Mathematical Specification of the Conceptualisation

- A specification of conceptualisation in **mathematical** (set-theoretic) terms is provided
- The choice of mathematics allows concentrating on the proper capturing of the notions of the conceptualisation **postponing any language issue**

A fabula f is a 5-tuple $f=(Ef, pf, bf, df, cf)$ consisting of:

- A finite set of events, $Ef \subset E$
- The event composition function pf: $Ef \rightarrow Ef$ associating some event $e1$ in Ef with a different event $e2$ in Ef, such that $e1$ is a part of $e2$. In this case, $e1$ is a sub-event of $e2$ and $e2$ is a super-event of $e1$.
- The event beginning function bf: $Ef \rightarrow T$, associating each event e in Ef with a time-point $t=bf(e)$ in T, such that event e starts at time $bf(e)$.
- The event ending function df: $Ef \rightarrow T$, associating each event e in Ef with a time-point $t=df(e)$ in T, such that event e ends at time $df(e)$.
- The causal dependence relation $cf \subseteq Ef \times Ef$, such that $e1, e2 \in cf$ if and only if event $e2$ causally depends on event $e1$

Validation of the Conceptualisation

The experiment consists in using the conceptualisation for expressing a narrative based on a short biography of the Italian poet Dante Alighieri

In order to carry out the experiment, two instruments were used:

- An **ontology** to express the conceptualisation
- A **computer system** to construct, store and access narratives

Selection of an Existing Ontology 1/2

The conceptualisation was expressed using standard ontologies, for **interoperability** reasons

We selected two standard ontologies:

1. the **CRM**, a high-level ontology that allows to integrate the information contained in data of the cultural heritage domain along with their correlation with knowledge stored in libraries and archives
2. **DOLCE-Lite-Plus** (an extension of DOLCE), aiming at representing the ontological categories underlying natural language and human common-sense

Both DOLCE+ and the CRM are adequate to express the conceptualisation of narratives

Selection of an Existing Ontology 2/2

Our choice fell on the **CRM** for the following reasons:

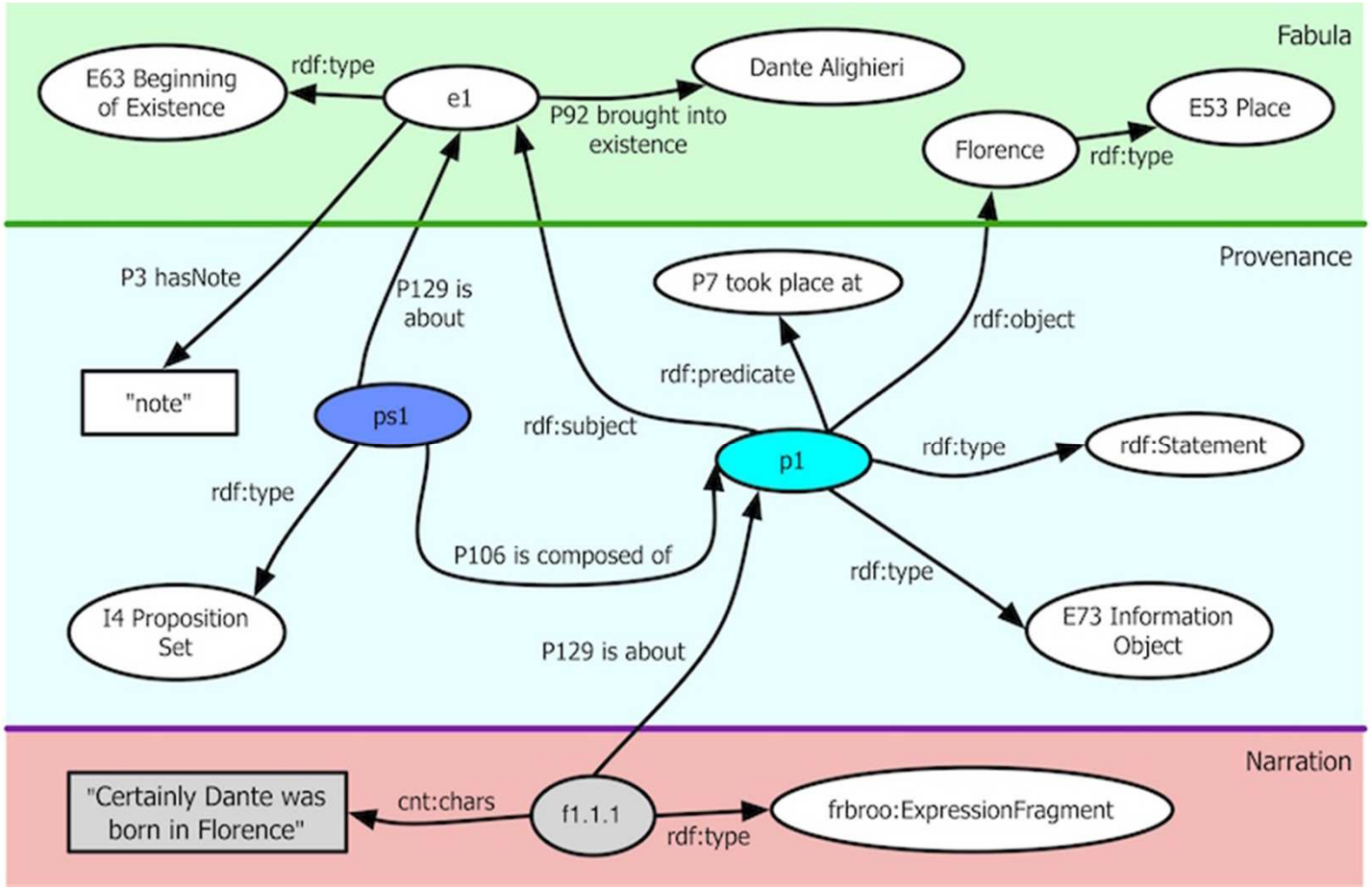
- The CRM is an **ISO standard** since 2006 (ISO21127:2006) and renewed 2014 (ISO21127:2014). It offers a stronger guarantee under many aspects: it is widely known, it is regularly revised, it is universally accessible
- The CRM is specifically thought for the **cultural heritage domain**, and as such it is **closer** to the domain of **narratives** than DOLCE+, which is built with software engineering in mind
- The **Special Interest Group of the CRM** continuously works for expanding the domain of applicability of the ontology

An Ontology for Narratives 1/2

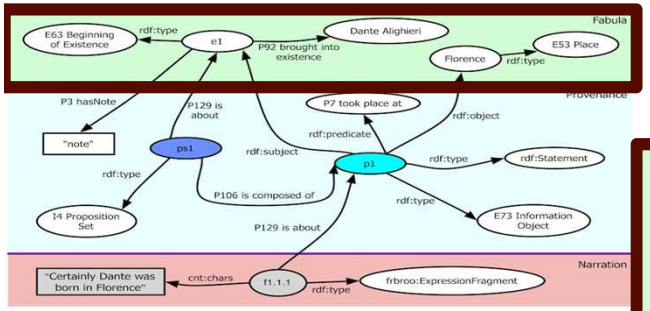
- ***CRM as reference vocabulary***

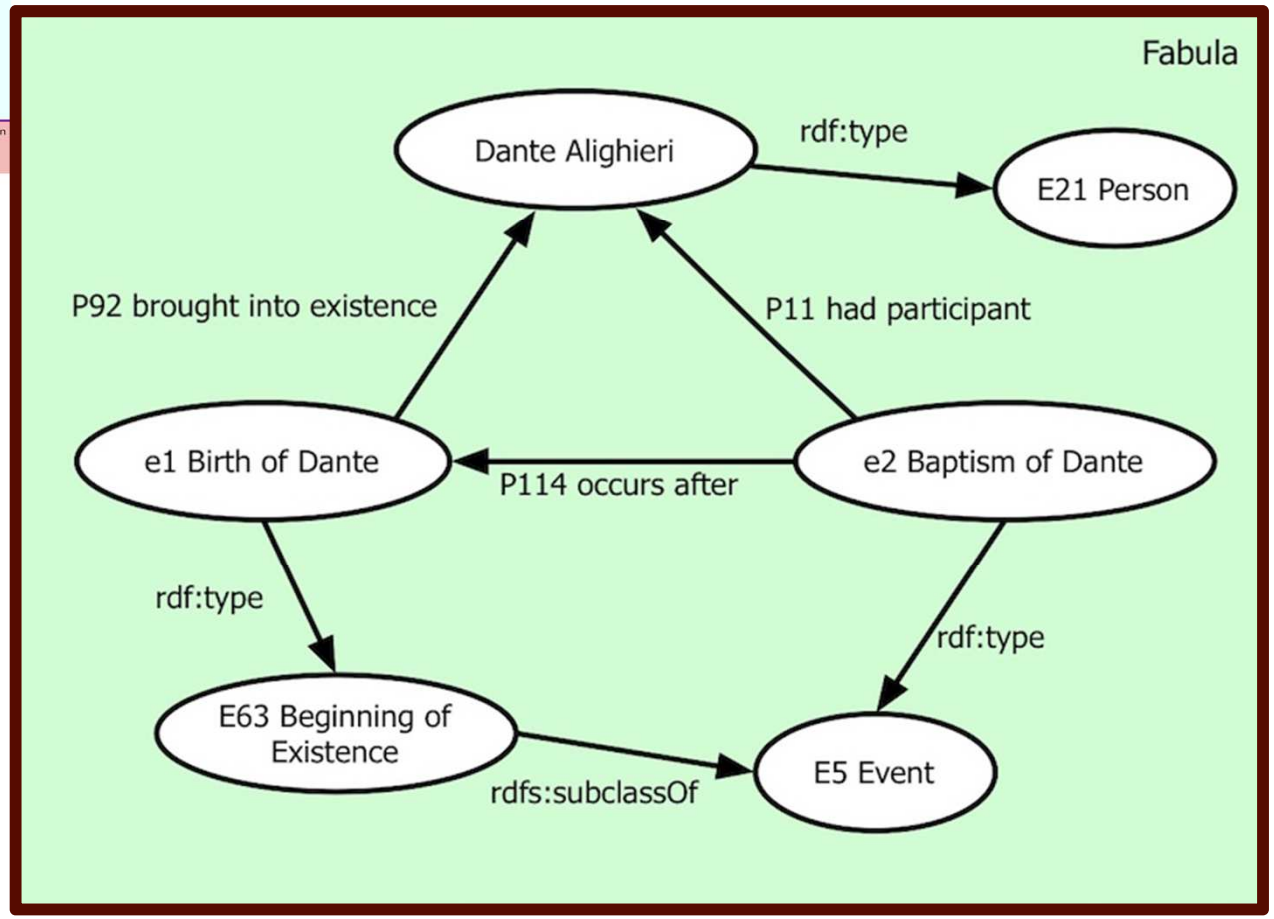
- CRM has been ***extended*** with notions that are specific for narratives

An Ontology for Narratives 2/2

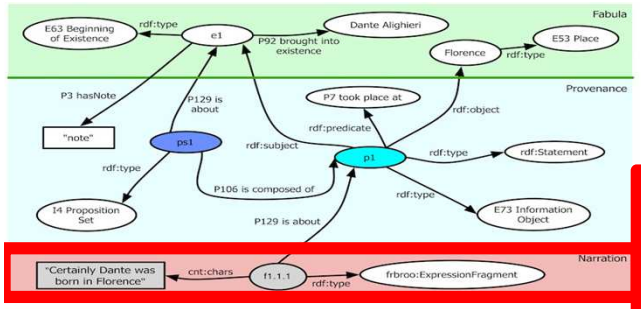


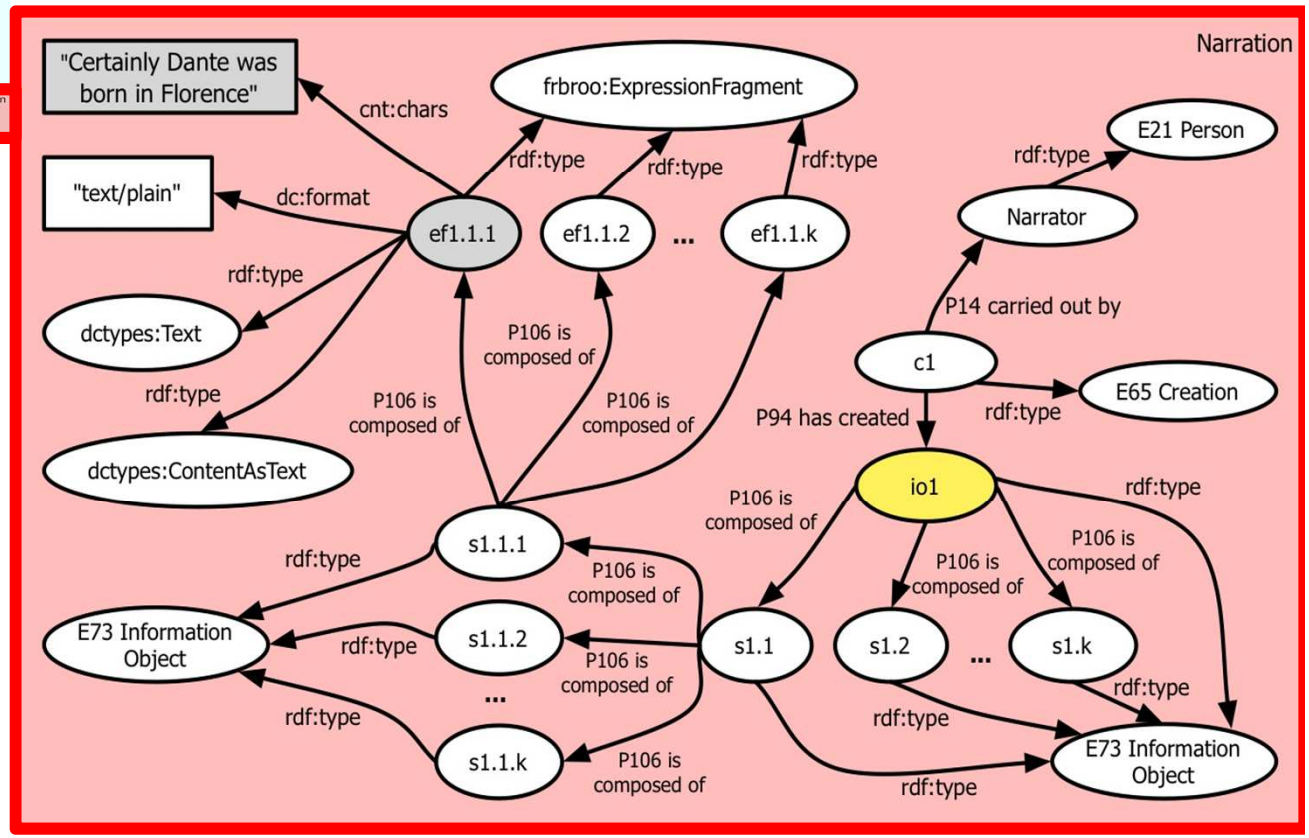
Fabula


Fabula
E63 Beginning of Existence
rdf:type
e1
P92 brought into existence
Dante Alighieri
Florence
E53 Place
rdf:type
Provenance
P3 hasNote
P129 is about
P7 took place at
rdf:object
rdf:predicate
"note"
ps1
rdf:subject
p1
rdf:type
rdf:Statement
rdf:type
P106 is composed of
rdf:type
I4 Proposition Set
P129 is about
E73 Information Object
Narration
"Certainly Dante was born in Florence"
cnt:chars
f1.1.1
rdf:type
frbroo:ExpressionFragment

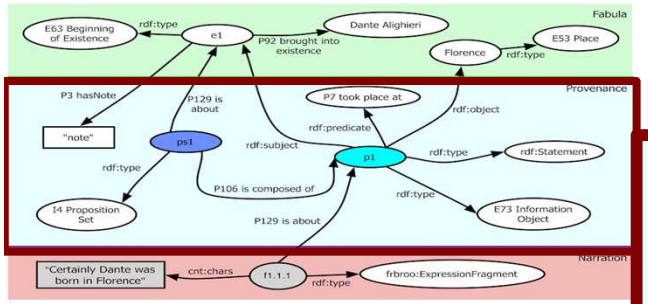

Fabula
Dante Alighieri
rdf:type
E21 Person
P92 brought into existence
P11 had participant
e1 Birth of Dante
P114 occurs after
e2 Baptism of Dante
rdf:type
rdf:type
E63 Beginning of Existence
E5 Event
rdfs:subclassOf

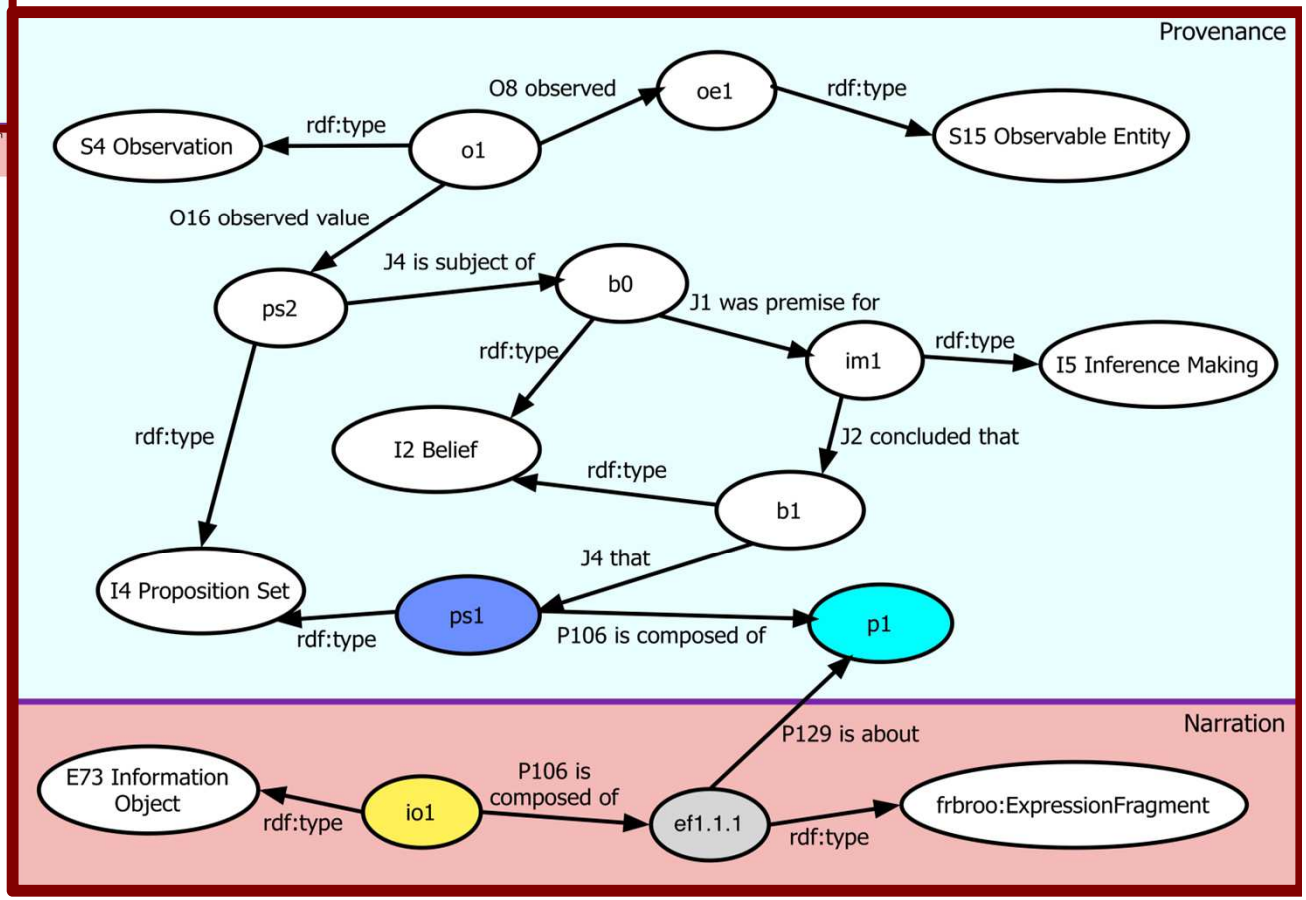
Narration


Fabula
E63 Beginning of Existence
rdf:type
e1
P92 brought into existence
Dante Alighieri
Florence
E53 Place
rdf:type
Provenance
P3 hasNote
P129 is about
P7 took place at
rdf:predicate
rdf:object
"note"
ps1
rdf:subject
p1
rdf:type
rdf:Statement
rdf:type
P106 is composed of
rdf:type
I4 Proposition Set
P129 is about
E73 Information Object
Narration
"Certainly Dante was born in Florence"
cnt:chars
f1.1.1
rdf:type
frbroo:ExpressionFragment


Narration
"Certainly Dante was born in Florence"
cnt:chars
frbroo:ExpressionFragment
"text/plain"
dc:format
rdf:type
rdf:type
rdf:type
ef1.1.1
ef1.1.2
...
ef1.1.k
rdf:type
dctypes:Text
rdf:type
dctypes:ContentAsText
P106 is composed of
P106 is composed of
P106 is composed of
E21 Person
rdf:type
Narrator
P14 carried out by
c1
E65 Creation
rdf:type
P94 has created
io1
rdf:type
P106 is composed of
P106 is composed of
P106 is composed of
rdf:type
s1.1.1
P106 is composed of
E73 Information Object
rdf:type
s1.1.2
P106 is composed of
s1.1
s1.2
...
s1.k
rdf:type
...
rdf:type
rdf:type
s1.1.k
P106 is composed of
E73 Information Object
rdf:type

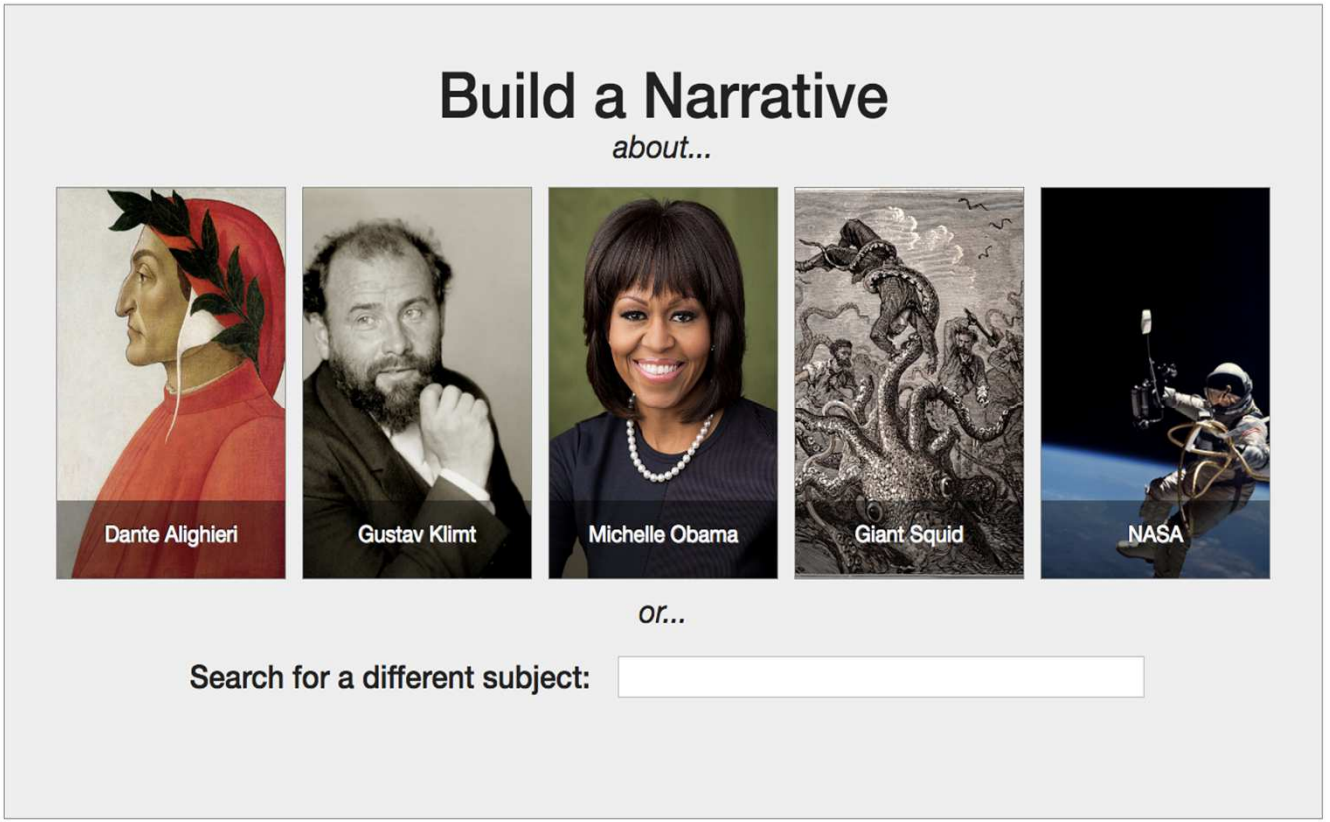
Provenance


Fabula
E63 Beginning of Existence
rdf:type
e1
P92 brought into existence
Dante Alighieri
Florence
E53 Place
rdf:type
Provenance
P3 hasNote
"note"
P129 is about
ps1
rdf:type
P7 took place at
rdf:predicate
rdf:subject
rdf:object
p1
rdf:type
rdf:Statement
P106 is composed of
I4 Proposition Set
P129 is about
rdf:type
E73 Information Object
Narration
"Certainly Dante was born in Florence"
cnt:chars
f1.1.1
rdf:type
frbroo:ExpressionFragment


Provenance
O8 observed
oe1
rdf:type
S15 Observable Entity
S4 Observation
rdf:type
o1
O16 observed value
J4 is subject of
b0
J1 was premise for
ps2
rdf:type
im1
rdf:type
I5 Inference Making
rdf:type
I2 Belief
J2 concluded that
rdf:type
b1
J4 that
I4 Proposition Set
ps1
P106 is composed of
p1
rdf:type
Narration
P129 is about
E73 Information Object
P106 is composed of
rdf:type
io1
ef1.1.1
rdf:type
frbroo:ExpressionFragment

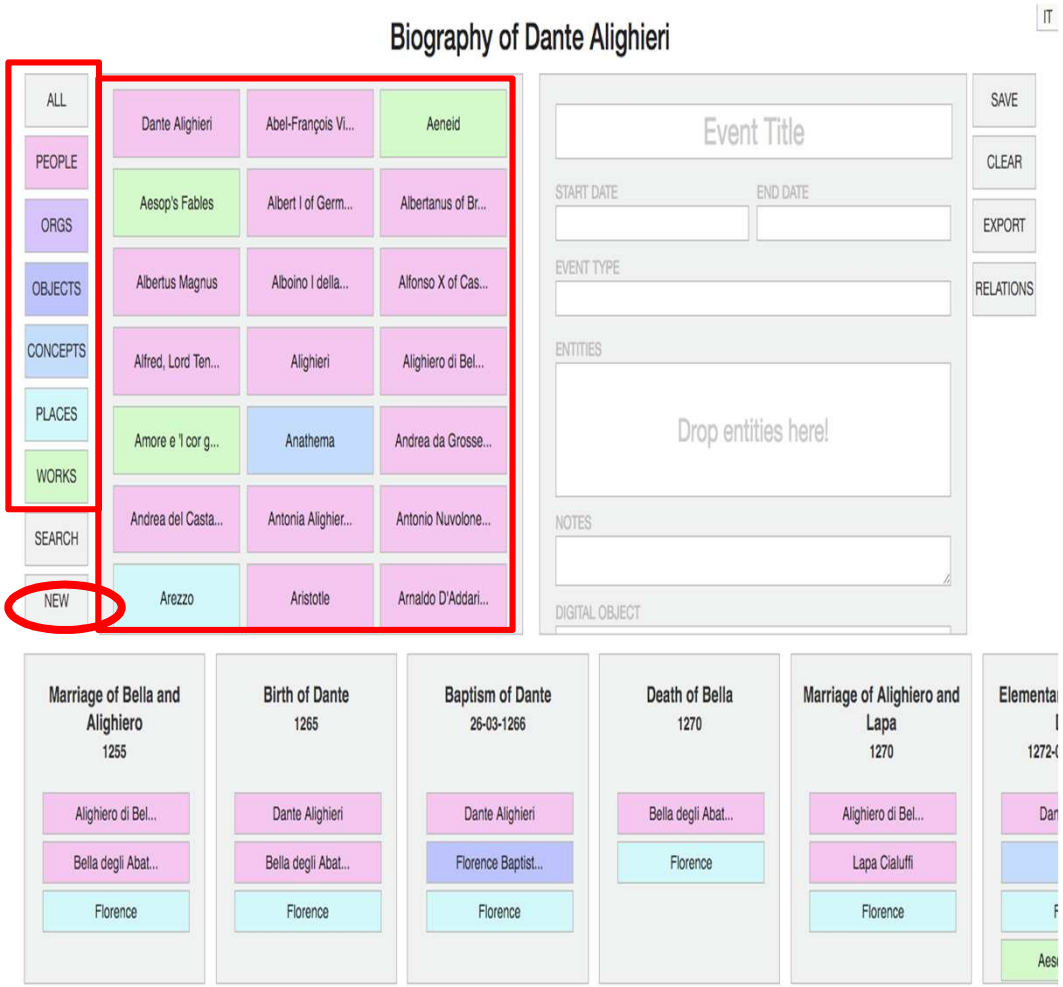
Creating Narratives

On top of the ontology, we created a **semi-automatic tool** for building narratives, since we found no suitable tool to the needs of our research



Tool Interface

- The tool **retrieves** and **assigns URIs** to the instances of the classes using **Wikidata** as **resource**
- The user can introduce **new entities** not present in Wikidata
- The instances are **automatically organized** in the classes of our ontology (e.g. place, people, organization etc.)

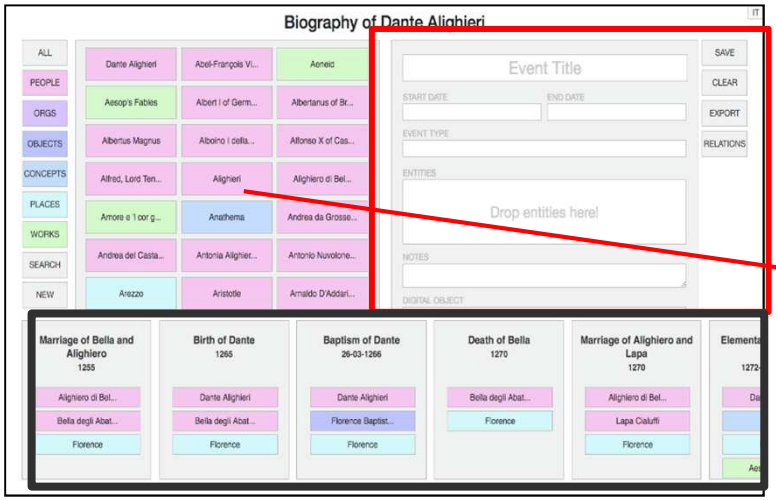


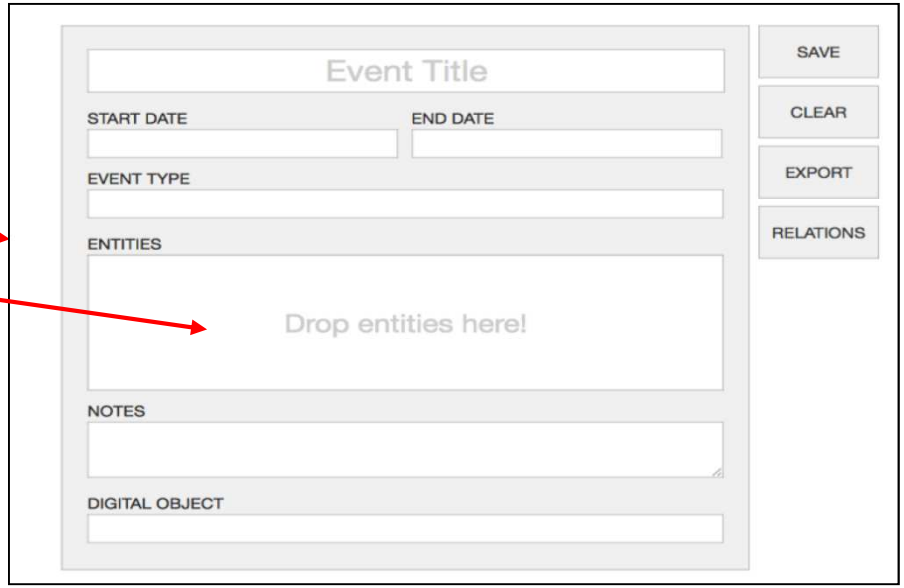
Mapping Between Classes of Wikidata and Our Ontology

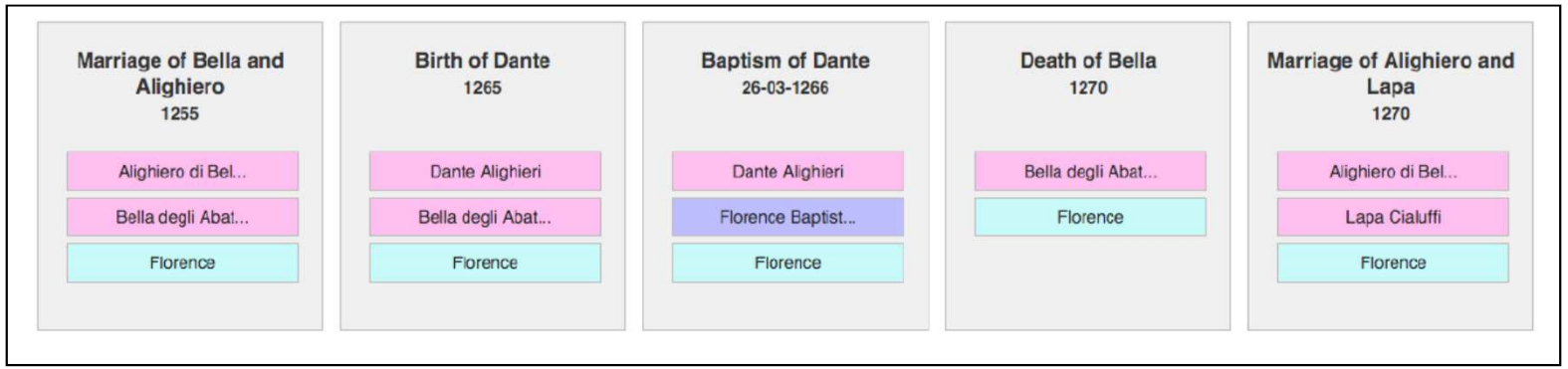
- **Mapping** between **classes of our ontology** and those of **Wikidata**

Wikidata	*Our ontology*
Q5 human	E21 Person
Q16334295 group of humans	E74 Group
Q7184903 abstract object	E89 Propositional Object
Q223557 physical object	E19 Physical Object
Q17334923 location	E53 Place
Q234460 text	E73 Information Object
Q478798 image	
Q340169 media	
Q186081 time interval	E52 Time-Span
Q1190554 event	E5 Event

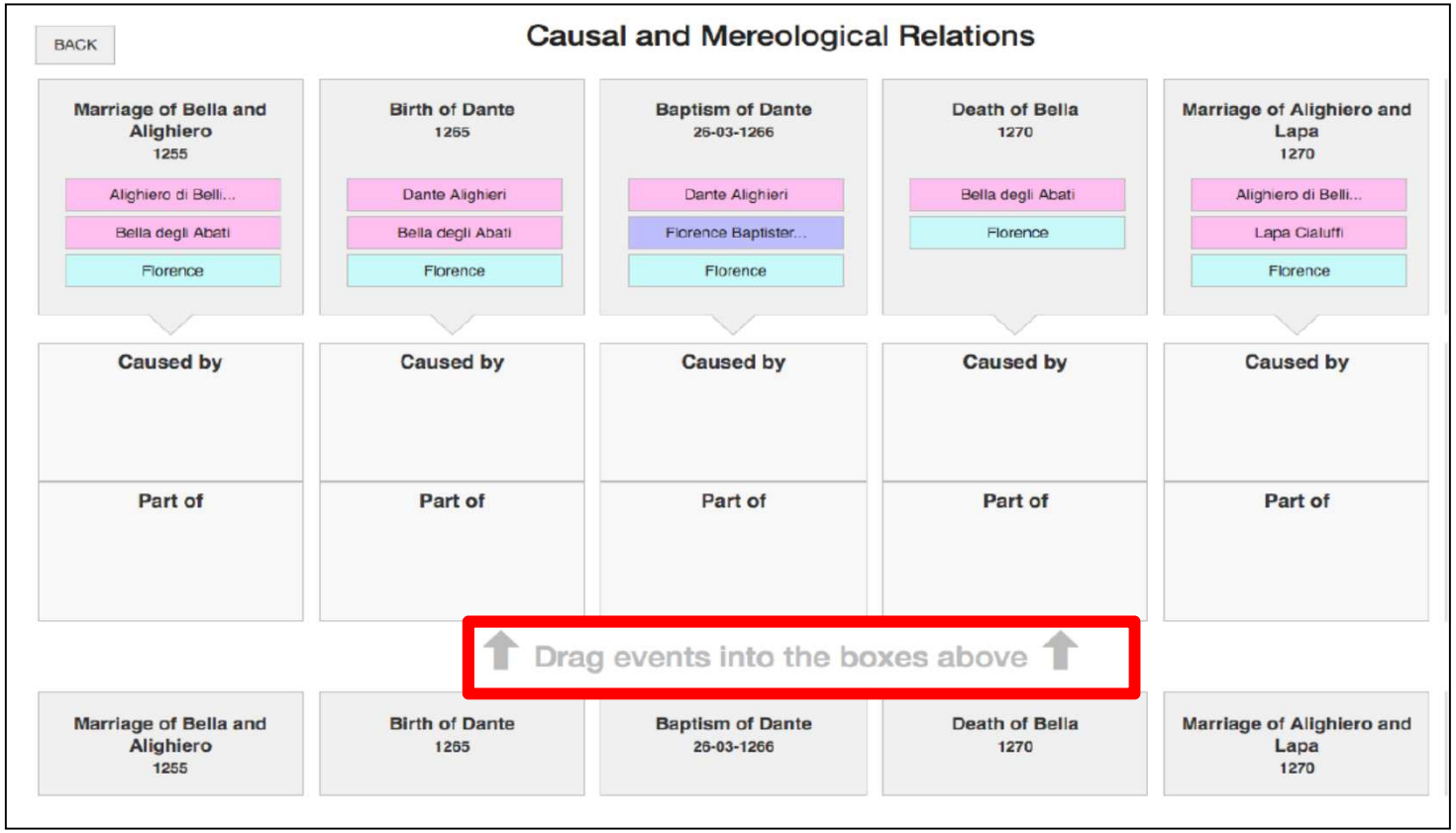
Event Creation and Timeline



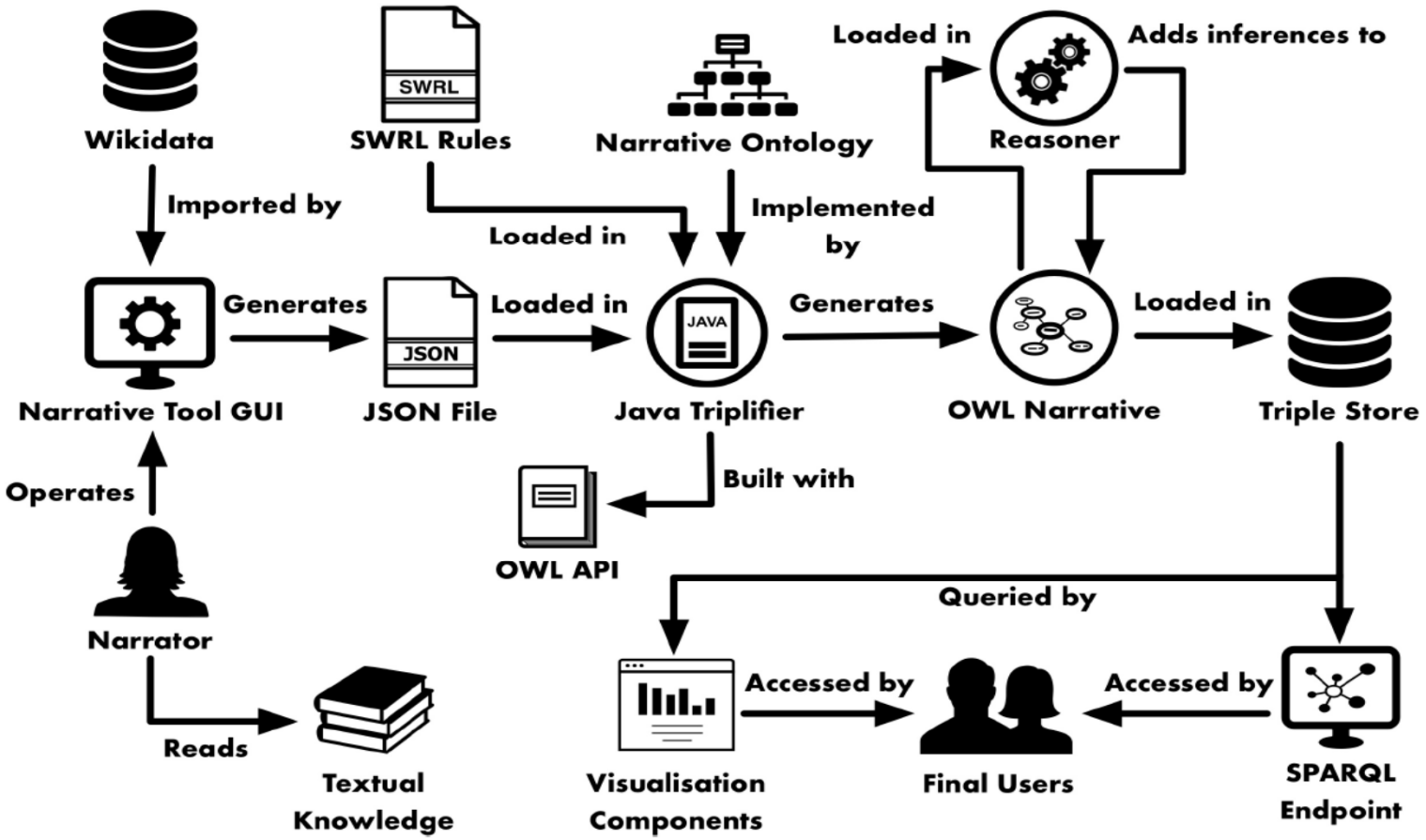




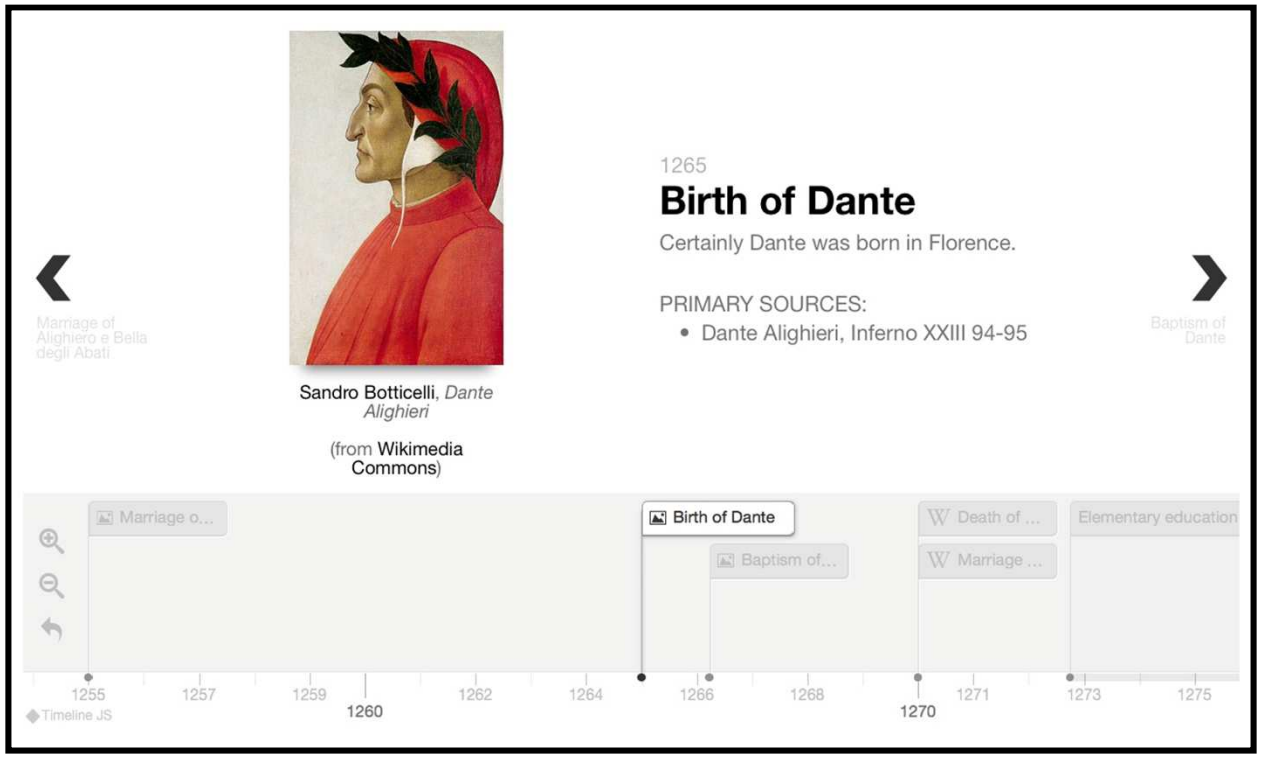
Causal and Mereological Relations



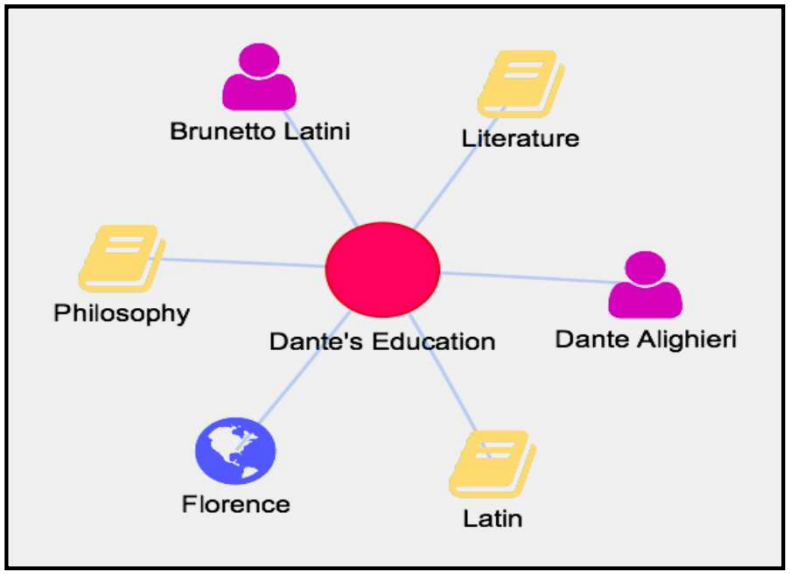
The SW Architecture

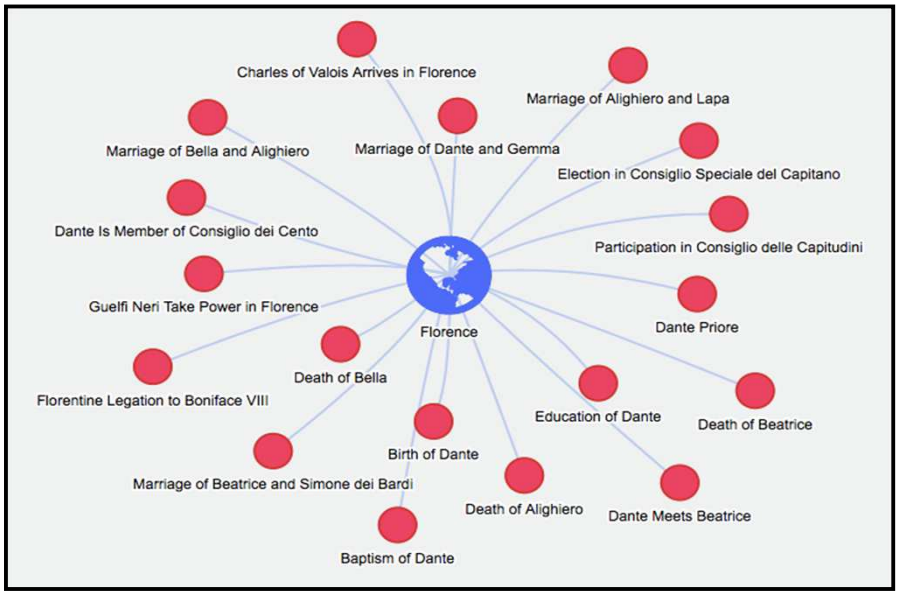


Visualisation: Timeline



Visualization: graphs





Visualization: Tables

Event	Date	Primary Source	Author	Text Fragment	Reference Fragment
Birth of Dante	1265	Inferno	Dante Alighieri	"io fui nato e cresciuto sovra 'l bel fiume d'Arno, alla gran villa"	XXIII 94-95
Baptism of Dante	1266-03-26	Inferno	Dante Alighieri	"il mio bel San Giovanni"	XIX 17
Baptism of Dante	1266-03-26	Paradiso	Dante Alighieri	"in sul fonte del mio battesmo"	XXV 8-9
Marriage of Dante and Gemma	1277-02-09	Instrumentum Dotis	Uguccione di Baldovino	"Domine Gemme vidue, uxori olim Dantis Allagherii et filie condam domini Manetti domini Donati, pro iure sue dotis librarum CC florenorum parvorum ut de instrumento dotis constat manu ser Ranaldi filii condam Oberti Baldovini de Florentia notarii, ex inbreviaturis ser Uguiccionis Baldovini notarii, facto in anno Domini M° CCLXXVI, indictione VJ, die VIIIJ° mensis februarii"	Capitani di Parte Guelfa, Numeri Rossi, 42, c. 3r
Dante's First Travel to Bologna	1286–1287	Trattatello in laude di Dante	Giovanni Boccaccio	"Egli li primi inizii, sì come di sopra è dichiarato, prese nella propia patria, e di quella, sì come a luogo più fertile di tal cibo, n'andò a Bologna"	

Event	Date
Baptism of Dante	1266-03-26
Death of Bella	1270
Marriage of Alighiero and Lapa	1270
Elementary Education of Dante	1272–1277
Death of Alighiero	1276
Marriage of Dante and Gemma	1277-02-09

Evaluation 1/3

Questionnaire to a historian who is an expert of Dante Alighieri's life. We asked him about:

1. The **representational adequacy of the ontology**
    - The historian explored the narrative on the timeline and looked at the contextualization of the individual events of the fabula
    - The use of external resources to enrich the narrative
    - The representation of the provenance
    - Different narratives of the same topic, created by different scholars

    Result: The historian gave *positive* evaluation on the points above (4.3/5)

2. The **effectiveness of a tool** to build narratives

    Results:
    - It is possible to formalize a significant narrative from 40 pages (104,013 characters) of text in a matter of a few hours (80 events/7 hours)
    - 70% of the entities used in the narration of the life of Dante are those defined in Wikidata

Evaluation 2/3

3. The **satisfaction of the historian's requirements**, defined at the beginning of the study → SPARQL queries and visualisation of the knowledge

- Extracted knowledge

- Events happened in a certain given range of time
- Events involving certain given entities (e.g. place, person)
- Events inter-linked by certain given relations (e.g. causal or mereological)

- Knowledge visualisation

- Visualising the fabula of a narrative on a timeline
- Visualising events along with their primary sources in table format, exportable in CSV format
- Primary sources in table format, exportable in CSV format
- Using network graphs to visualise an event and its related entities
- Using network graphs to visualise a certain given entity and its related events

Evaluation 3/3

In order to validate the ontology on **different domains**, representing **different types of narratives**, the narrative building tool and the underlying ontology are currently being used by:

1. A Computer Science professor at the University of Pisa to represent the **history of Informatics**, focussing on the Turing Award

2. A researcher in Computational Biology at the Italian National Research Council (CNR) to narrate the discoveries related to the **giant squid**

3. A Digital Humanities researcher at the CNR to create the narrative of the life and works of the Austrian painter **Gustav Klimt** and of the Italian poet **Dante Alighieri**

Conclusions 1/3

- The **primary research question** of this research was: *could an ontology be created to reasonably represent narrative*s?

- The **hypotheses** were that *written natural language of narratives, or oral narrative*, could be interpreted and analysed in terms of *common and shared interests* through digital libraries and these shared interests should be highly *generalised* and abstracted to an *ontological representation*

Conclusions 2/3

The **primary results** obtained with the work of this thesis are that:

1. A **methodological approach** can exist that advocates an interpretative analysis of written natural language narratives; more formally, the interpretative analysis can be caught by **knowledge engineering** and the process of ontological modelling as an act of **ontology engineering**

2. An **ontological representation** of the narrative has been proposed. This uses the **CRM ISO standard** as reference ontology

3. A **semi-automatic tool** has been built to populate the ontology

4. A **Web application** has been designed and implemented to **visualize** the **knowledge** in a simple form, endowed with tables and graphs

Conclusions 3/3

In order to introduce the narrative as new search functionality for digital libraries:

- Our **ontology** can be used as vocabulary for inserting narratives in digital libraries. Many digital libraries include texts that can be used to create narratives
- Our **tool** may be used to **build narratives** and import them in a digital library, since we used standard technologies to represent them
- The Web application can be used to **visualise narratives** within digital libraries

Future Work

- Developing an **expansion** of the search functionality of digital libraries for retrieving **narratives** in response to users' queries
- A **collaboration** was started with the **Special Interest Group of the CIDOC CRM** Ontology targeted at the development of an **extension of the CRM** for the representation of **narratives**
- extending the narrative ontology in order to
  - **represent and reason about the temporal relations** between events;
  - represent in a richer way the **text of narratives**, called narrations, and the relation between narration fragments and their semantic counterparts, events and objects;
- The methodology and the tools are disseminated as **Open Source Software** for research aims
- As future work, a **user test** to evaluate the **usability** and **accessibility** of the tool for collecting and visualising knowledge has been planned

Thank You

https://dlnarratives.eu