ISTITUTO DI SCIENZA E TECNOLOGIE
DELL'INFORMAZIONE "A. FAEDO"

Populating Narratives Using Wikidata Events: An Initial Experiment
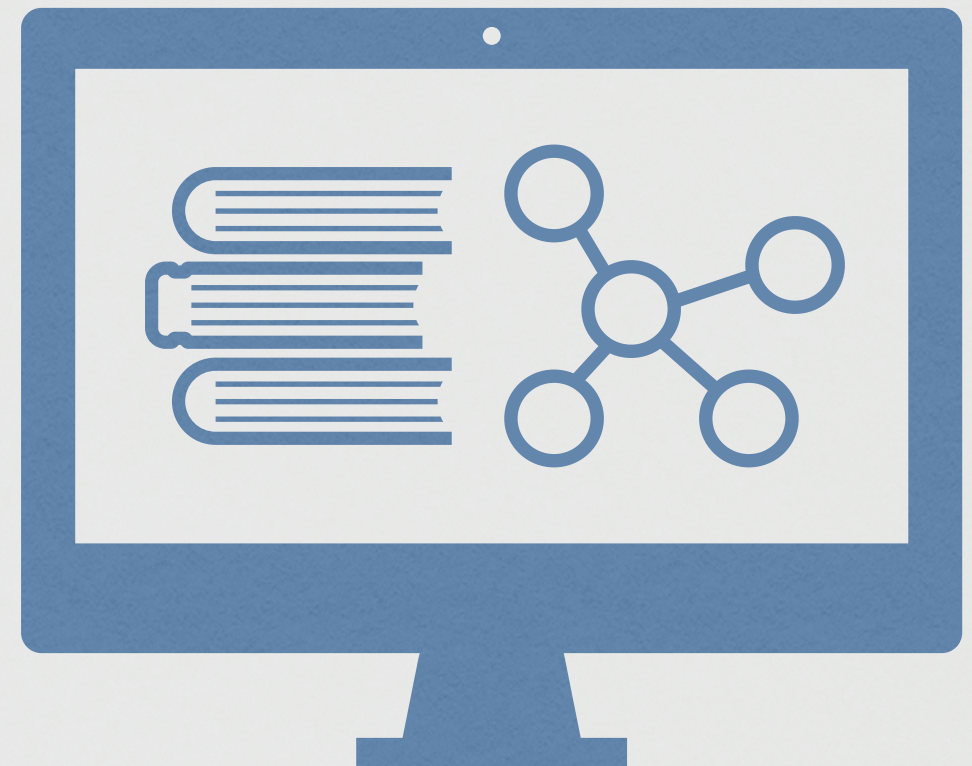
Daniele Metilli, Valentina Bartalesi, Carlo Meghini, Nicola Aloia

IRCDL 2019, Pisa

Narratives in Digital Libraries

Our research aims to **introduce narratives into Digital Libraries**

We believe that narratives would allow DLs to provide **more sophisticated information services** to their users, going beyond the current state

An Ontology for Narratives

We have developed a **formal ontology to represent narratives**

- Based on **Semantic Web** technologies

- Built on top of the **CIDOC CRM** standard ontology

- Expressed in **OWL**

We intend a **narrative** as a network of **events** defined by a **narrator**, endowed with **factual aspects** (who, what, where, when) and **semantic relations**

Narrative Building and Visualising Tool (NBVT)

We have developed a **tool to build and visualise narratives (NBVT)**

- Populates the **ontology**

- Semi-automated, imports knowledge from the **Wikidata knowledge base**

- Four **case studies**:
https://dlnarratives.eu/narratives.html
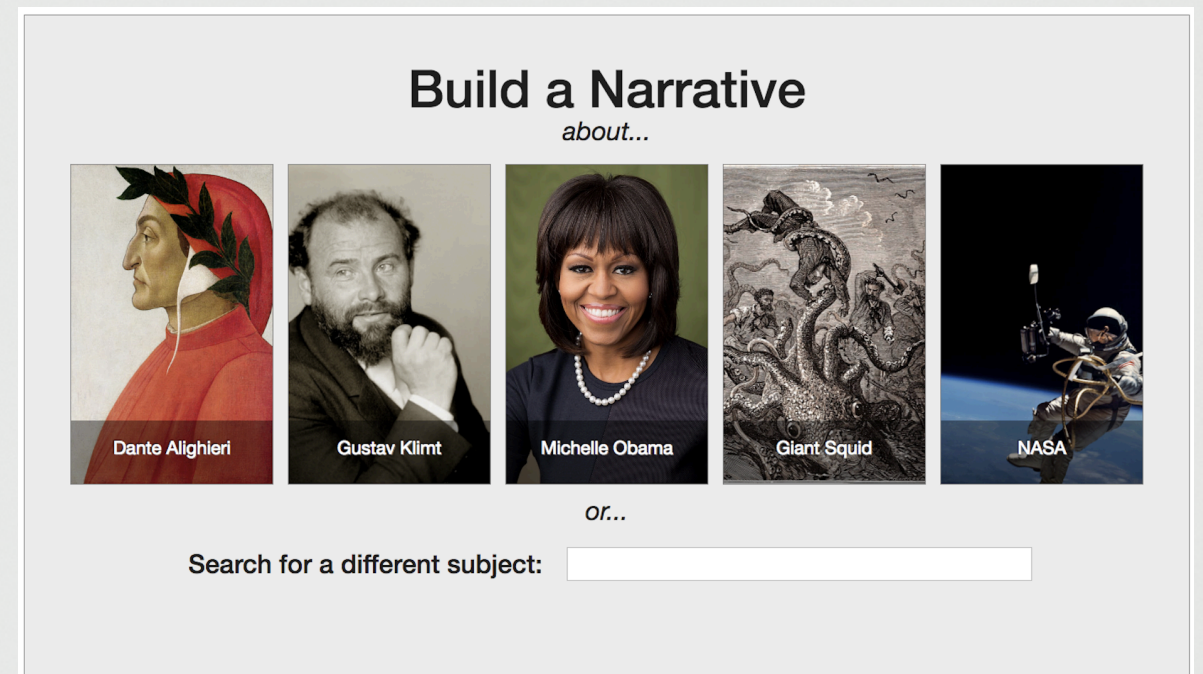
Integration with Wikidata

Wikidata is an **open collaborative knowledge base** containing more than 54 million entities

The user of NBVT is able to **import any Wikidata entity** into a narrative

The **narrative can be linked** to the much bigger **Wikidata graph**, and through it to related projects such as **Wikipedia** and **Wikimedia Commons**
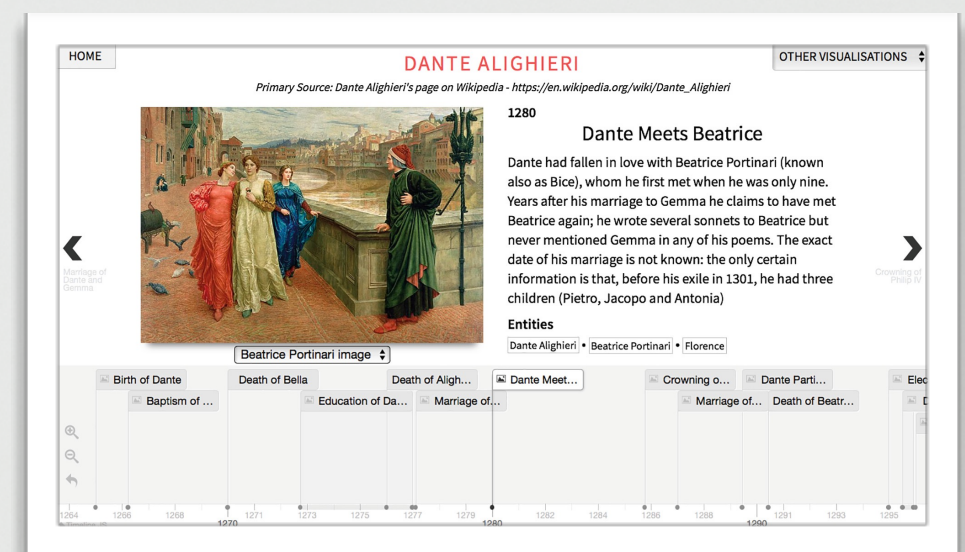
WIKIDATA

A Narrative of Dante Alighieri's Life

As case study we built a **narrative about Dante Alighieri** using NBVT

Composed of **53 events** describing the life of the poet, each connected to one or more **related entities** (e.g. people, places, objects…)
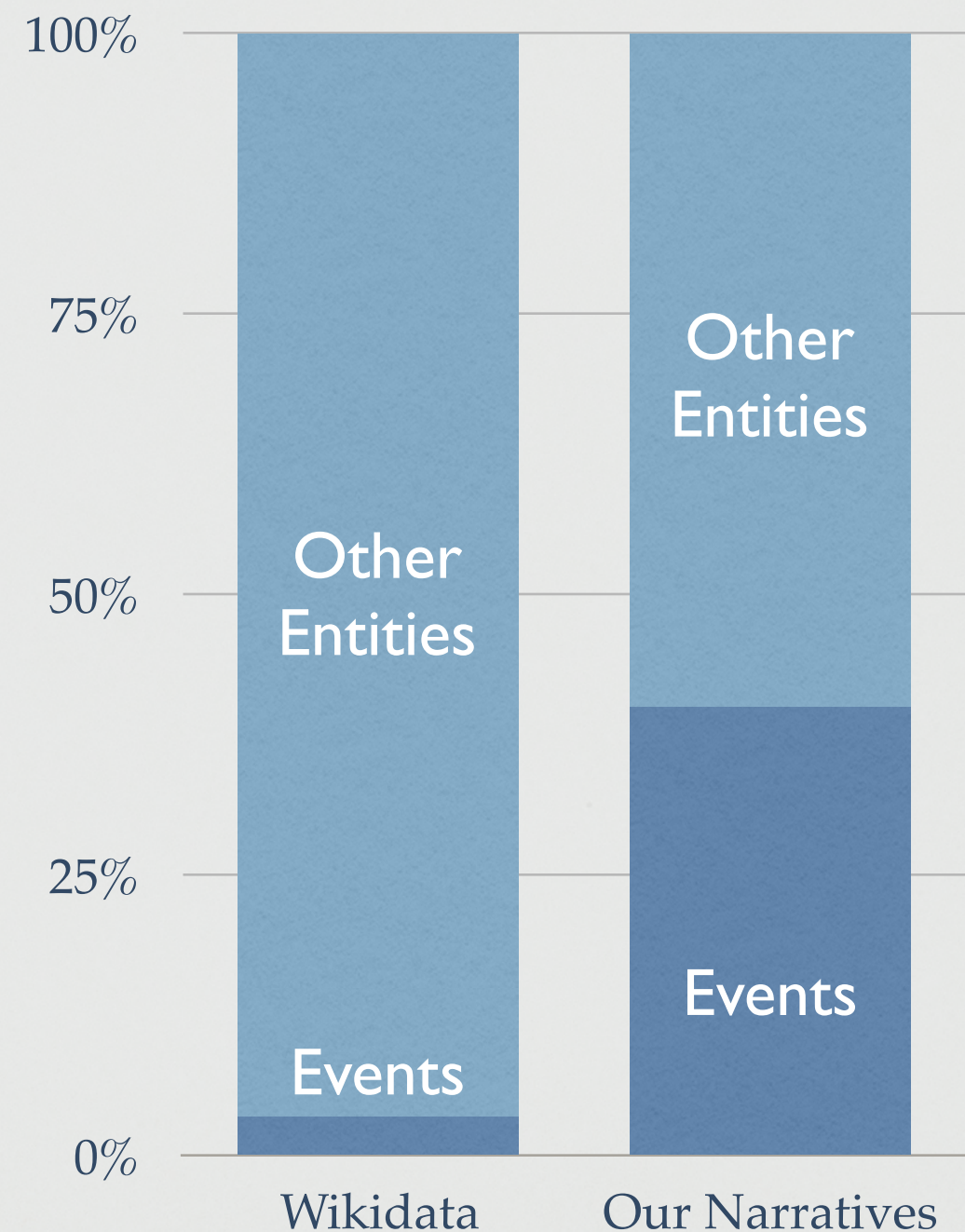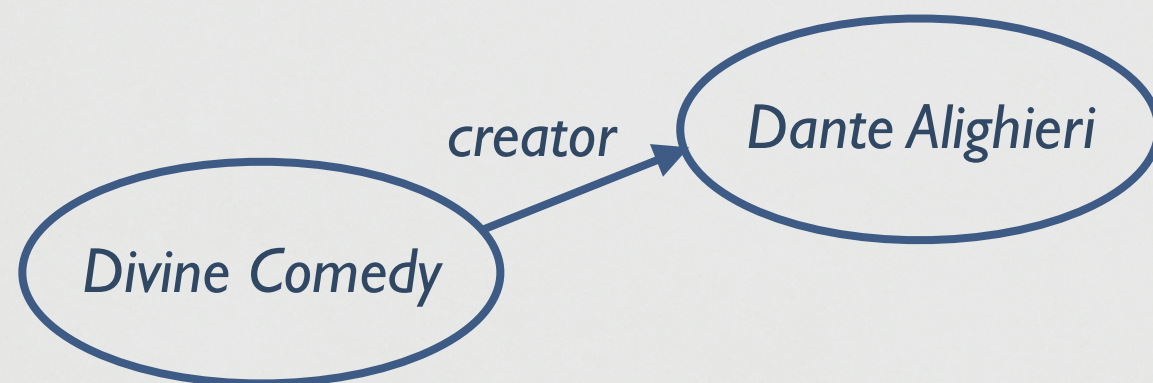
69% of these entities are present in Wikidata… but **only one event!**

Importing Events from Wikidata

Wikidata contains **more than 54 million entities**, but just **3.5% of these are events**

Since Wikidata's ontology is **not event-based**, most events are represented **implicitly**
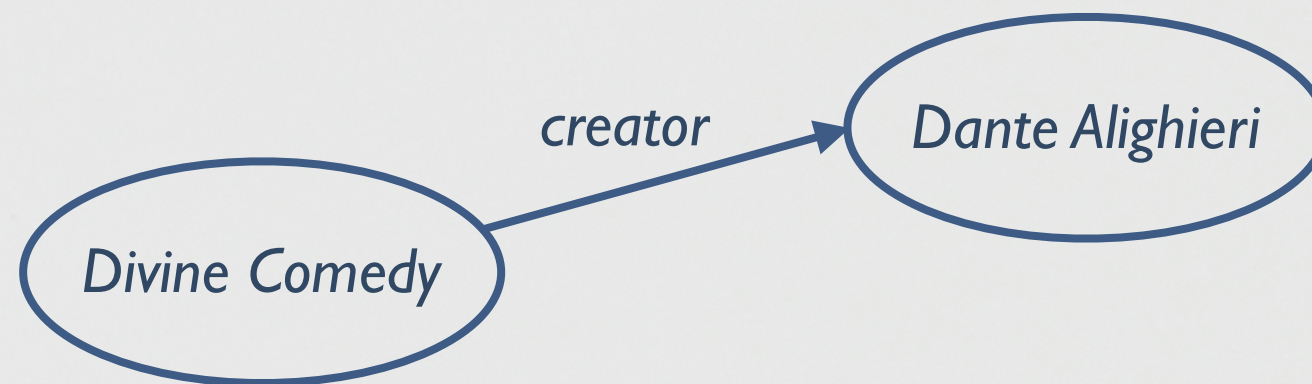
Generating the Wikidata Event Graph

We generated the **Wikidata Event Graph (WEG)** containing **all events** found in Wikidata, both implicit and explicit

- We developed **an algorithm to identify Wikidata properties** that were likely to **express events**

- In this initial experiment we focused on a **subset of 50 properties expressing events about people's lives**

- We generated an **explicit event** for each usage of each property (more than **11 million events** in total)

- We evaluated the results on our **narrative about Dante**

Wikidata Event Graph – Example

In Wikidata

- -

In the WEG
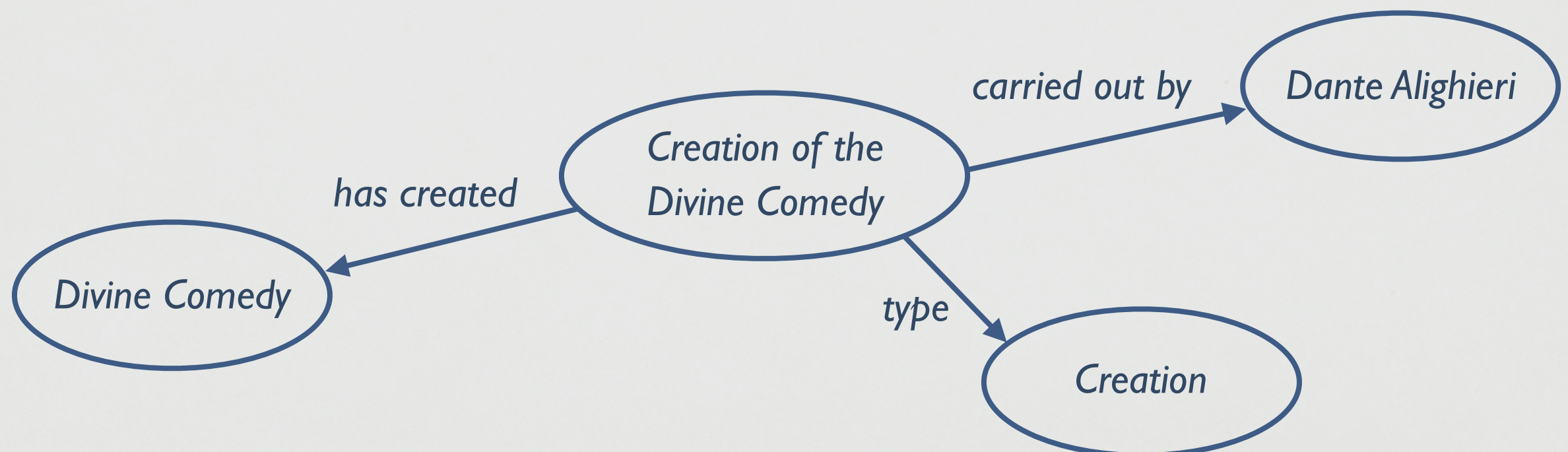
Results of the Experiment

Before the experiment, only one of the **54 events** of our case study was **explicitly present in Wikidata**

Now, **34 events** (60% of the total) can be found in the **Wikidata Event Graph**

Similar to the percentage of **related entities** that we had already linked to Wikidata (**69%**)

Not Linked
40%

Linked
60%

Events

Not Linked
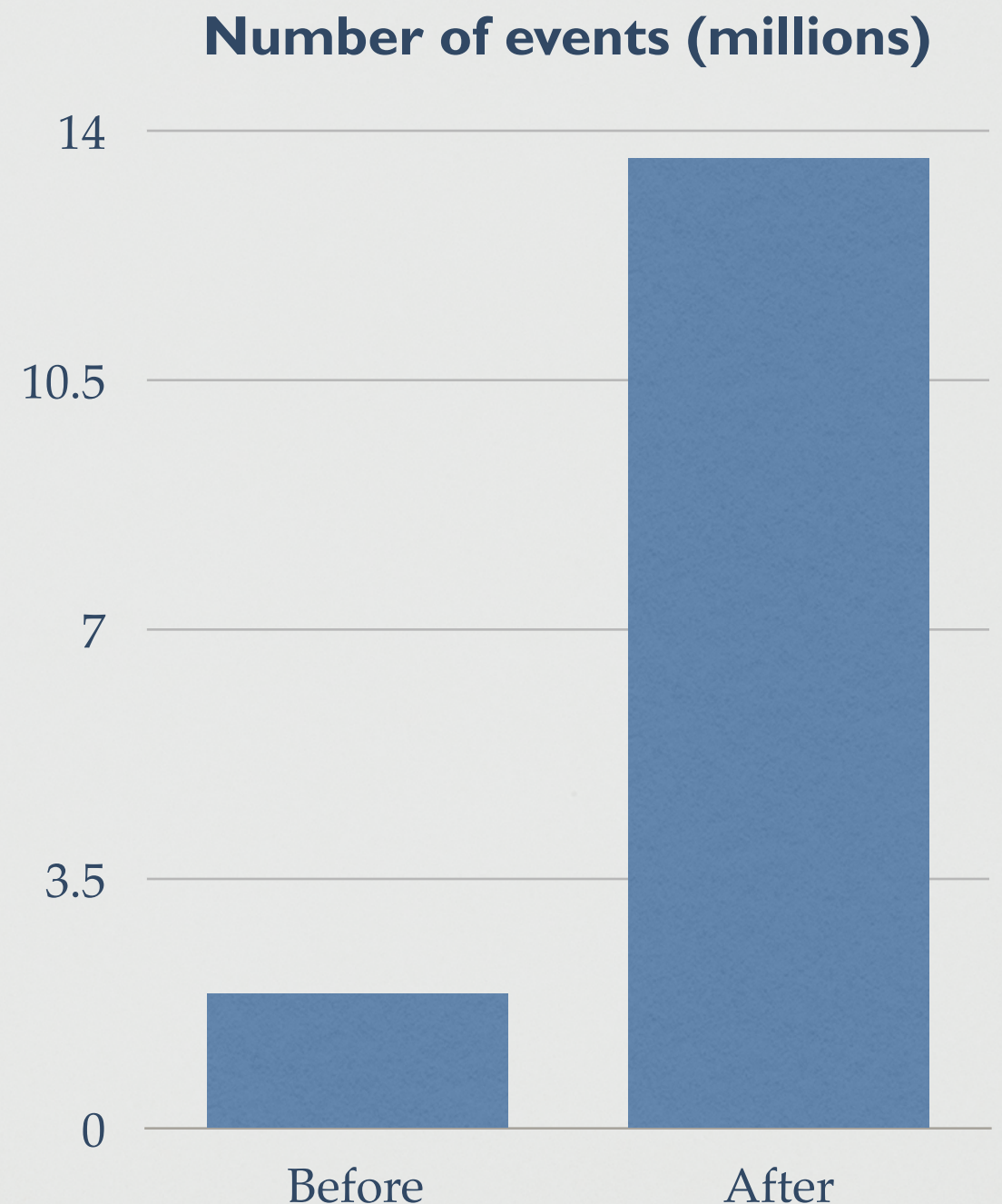31%

Linked
69%

Related Entities
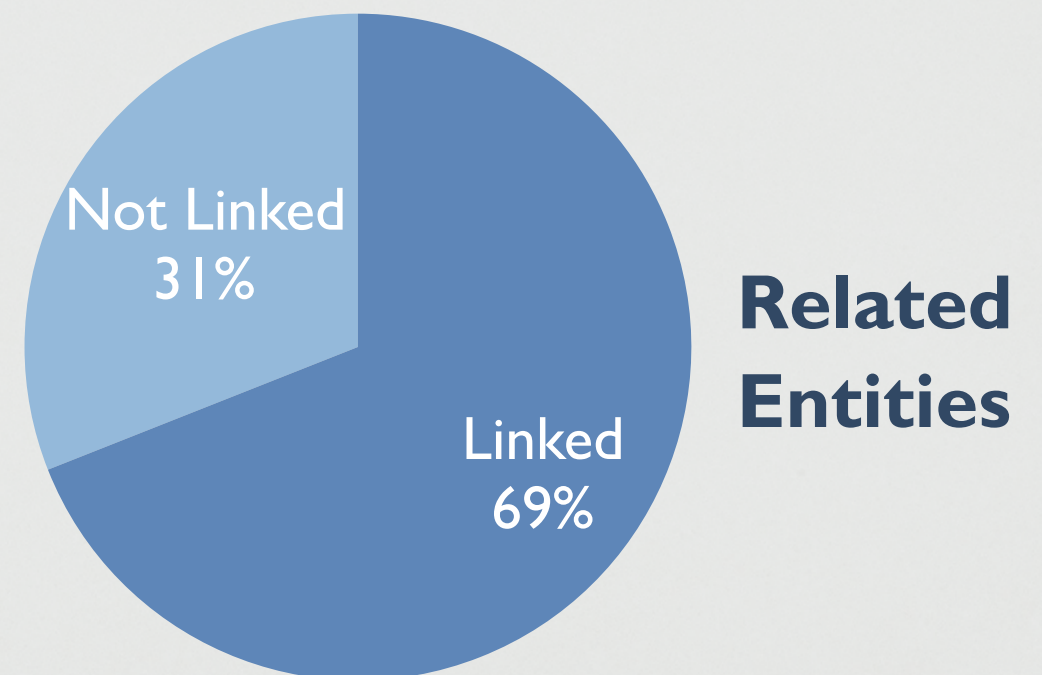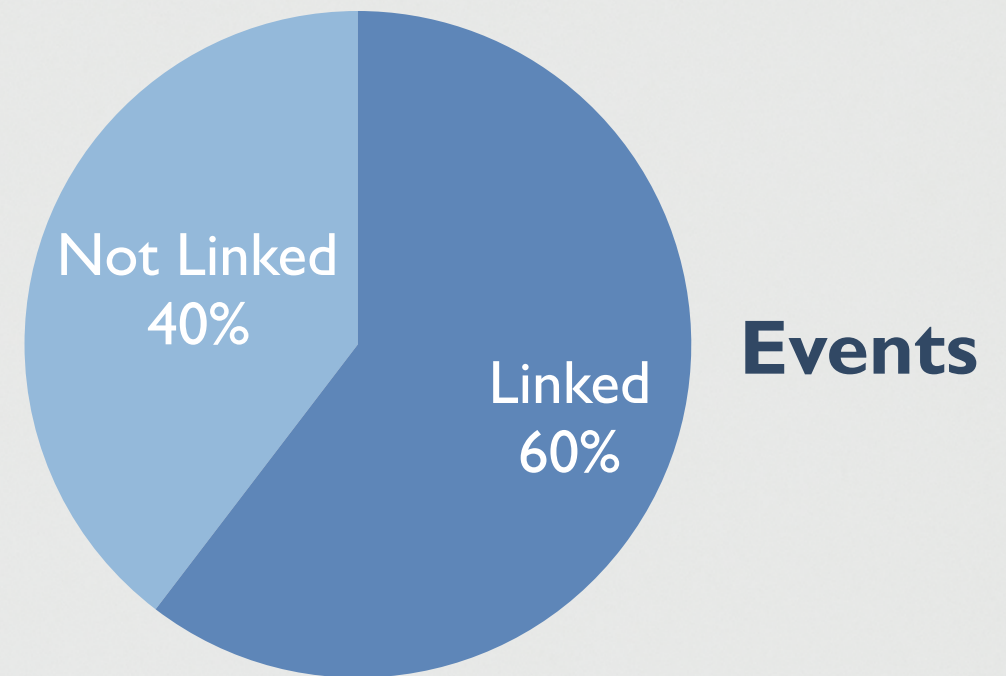
Results of the Experiment

We increased the **number of events** we can import into our tool by **more than 600%**

We started from **1.9 million** explicit Wikidata events

We added **11.7 million** events that were implicit

We now have access to a graph of **13.6 million** events

Number of events (millions)

Conclusions

We have presented an **experiment** on the population of the **narrative** of the life of **Dante Alighieri** using the events of the **Wikidata Event Graph (WEG)**

We have **generated a subset of the WEG** starting from a set of **50 Wikidata properties**

We have been able to **link to the WEG 60% of the events** of our narrative about Dante
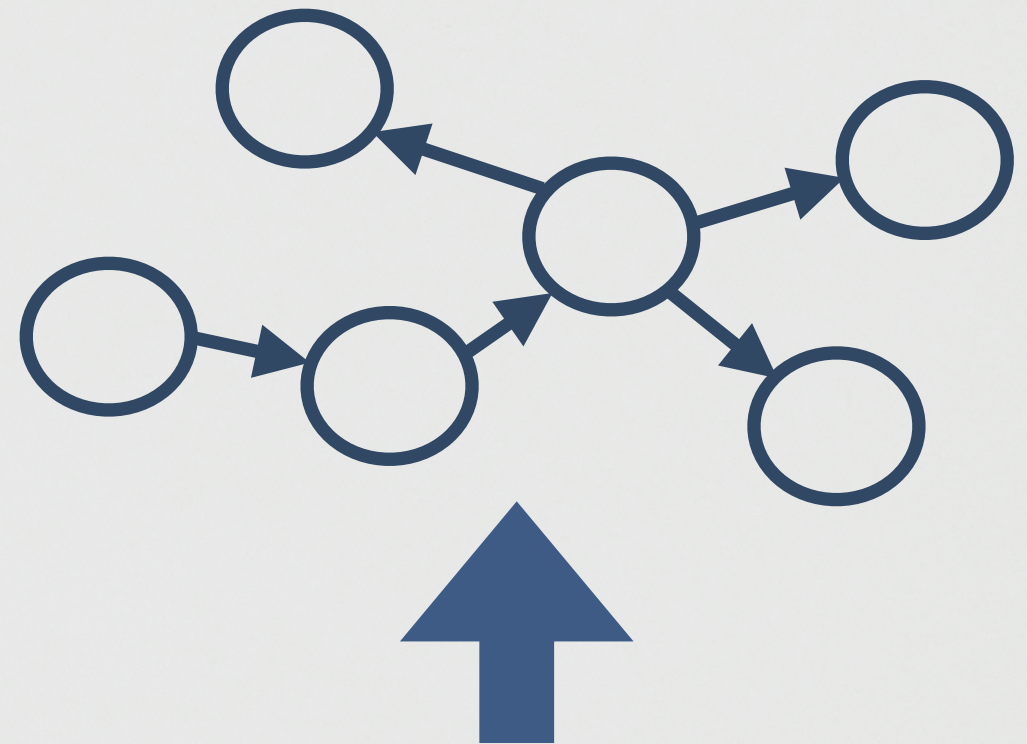
We now have access to a graph of **13.6 million events** which we can import into our **NBVT tool**

Future Works

Analyse more properties to **extract more events**

Study how to **suggest events from the WEG** to the users through the interface our tool

Work on **narrative extraction from text** in natural language

Thank you

https://dlnarratives.eu