



# Technical Deep Dive

Aggregazione di dati e annotazione semantica in  
ambienti culturali collaborativi

Alessandro Bertozzi

NET7 26/10/2023

# Indice

- Presentazioni
- Il progetto in breve
- Problema
- Una possibile soluzione
- Muruca e Pon
- Il workflow
- Considerazioni finali



# Chi sono?

Sono Alessandro Bertozzi. Attualmente sto concludendo i miei studi presso la sede dell'università di Bologna, nel corso di laurea di Digital Humanities and Digital Knowledge. Sono assunto presso l'azienda Net7 come software developer.



# Dove lavoro?

Net7 è un'azienda attiva nel settore dello sviluppo web. Al suo interno conta varie divisioni, tra cui quella dedicata alla Digital Humanities, sezione in cui lavoro.



# Il Progetto: informazioni generali

- Progetto dottorale PON fondi FSE REACT-EU DM 1061/2021, azione IV.4 – tematiche dell'innovazione.
- Area tematica e traiettoria di sviluppo della Strategia Nazionale di Specializzazione Intelligente (SNSI): *Turismo, Patrimonio culturale e industria della creatività* - Tecnologie e applicazioni per la conservazione, gestione e valorizzazione dei beni culturali, artistici e paesaggistici.
- Il progetto può essere riassunto secondo i seguenti punti chiave:
  - Riutilizzo di dati, metadati e standard di metadati da fonti eterogenee.
  - Creazione di una piattaforma popolata da risorse ricercabili e interoperabili.
  - Apertura del progetto verso un pubblico non specialistico.
  - Arricchimento dei contenuti attraverso strumenti open-source.

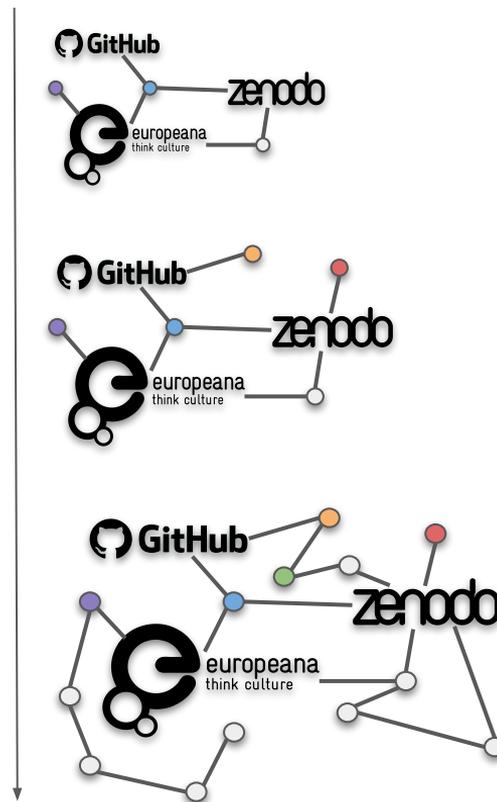


# Problema: una grande quantità di risorse

Con il passare degli anni, sempre più risorse, attinenti al patrimonio culturale, vengono metadotate e digitalizzate.

Questo comporta una serie di problematiche da tenere conto nella gestione e valorizzazione del patrimonio culturale.

1. Eterogeneità della modellizzazione.
2. Eterogeneità nella scelta dei servizi per memorizzare e condividere i dati raccolti.
3. Difficoltà nella reperibilità delle risorse.
4. Difficoltà nel maturare e condurre analisi quantitative su un corpus di dati esaustivo.

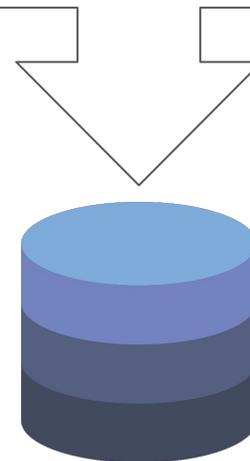
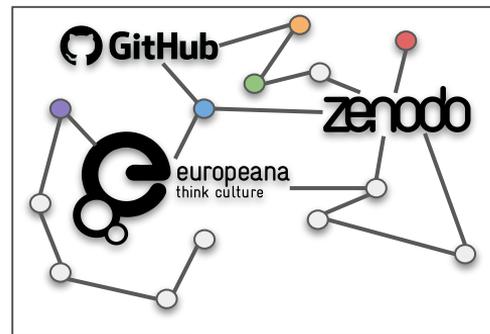


# Aggregatori

Un possibile approccio al problema è la progettazione e realizzazione di un aggregatore. Un singolo spazio dove poter reperire e consultare le risorse desiderate.

Lo scopo di un aggregatore può essere riassunto nei seguenti punti:

1. Normalizzazione dei dati in ingresso.
2. Semplificazione della ricerca di risorse.
3. Unico punto di accesso a una grande quantità di risorse.
4. Possibilità di risalire alla risorsa primaria da cui sono stati recuperati i dati.
5. Possibilità di condurre analisi quantitative su un grande corpus di dati.



# La soluzione di Net7: MURUCA e Pundit



Una stack di tecnologie che consente la veloce pubblicazione di contenuti culturali sul web. In particolare si compone di un *backend*, di un servizio *serverless* e di un *frontend*.



*pin the web*

Pundit è un insieme di tecnologie, sempre sviluppate da Net7, con lo scopo di fornire un gruppo di strumenti per annotare le pagine web.



# Muruca nel progetto PON

## Backend

- La parte adibita allo memorizzazione e modellizzazione dei dati.
- Soprattutto fornisce un ambiente di lavoro utile all'inserimento di dati e alla definizione di particolari politiche di recupero dati da altre fonti.

## Serverless

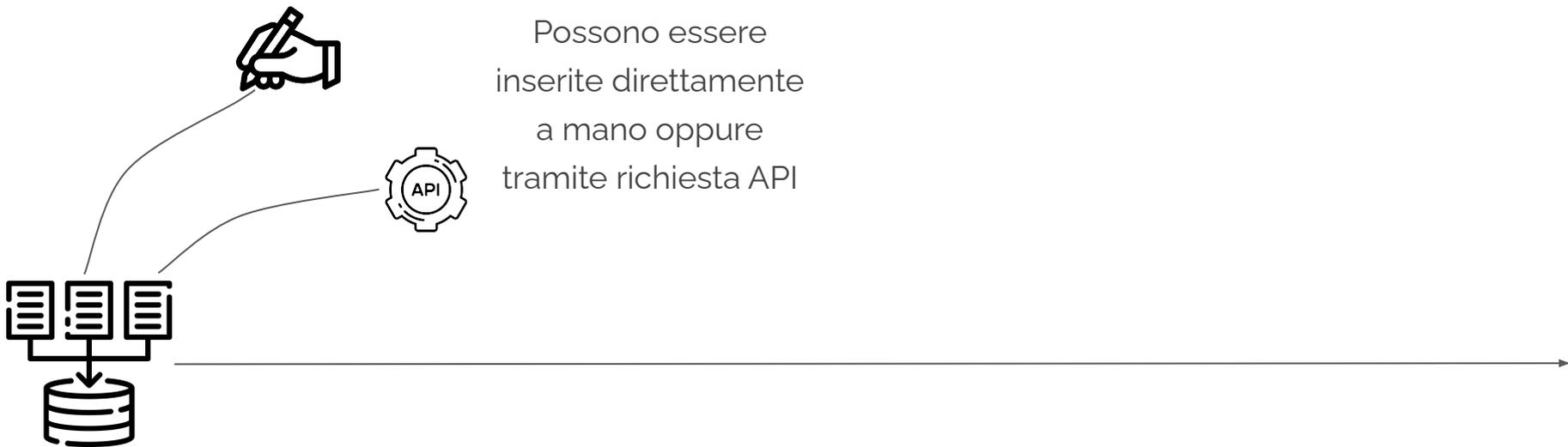
- La parte centrale di tutta l'applicazione.
- Si occupa di recuperare le risorse non gestite direttamente dal backend.
- Si occupa di rimappare secondo i modelli dati definiti dall'utente.
- È lo strato di comunicazione tra il backend e il frontend

## Frontend

- Questa parte è quella con cui interagisce l'utente finale.
- Qui è possibile visualizzare le risorse recuperate tramite il serverless.
- È possibile in questa fase lanciare ricerche utilizzando metadati specifici.



# Muruca: il workflow



Possono essere  
inserite direttamente  
a mano oppure  
tramite richiesta API

## Harvesting

Si parte scegliendo e  
collezionando le  
risorse da riutilizzare  
all'interno del  
progetto

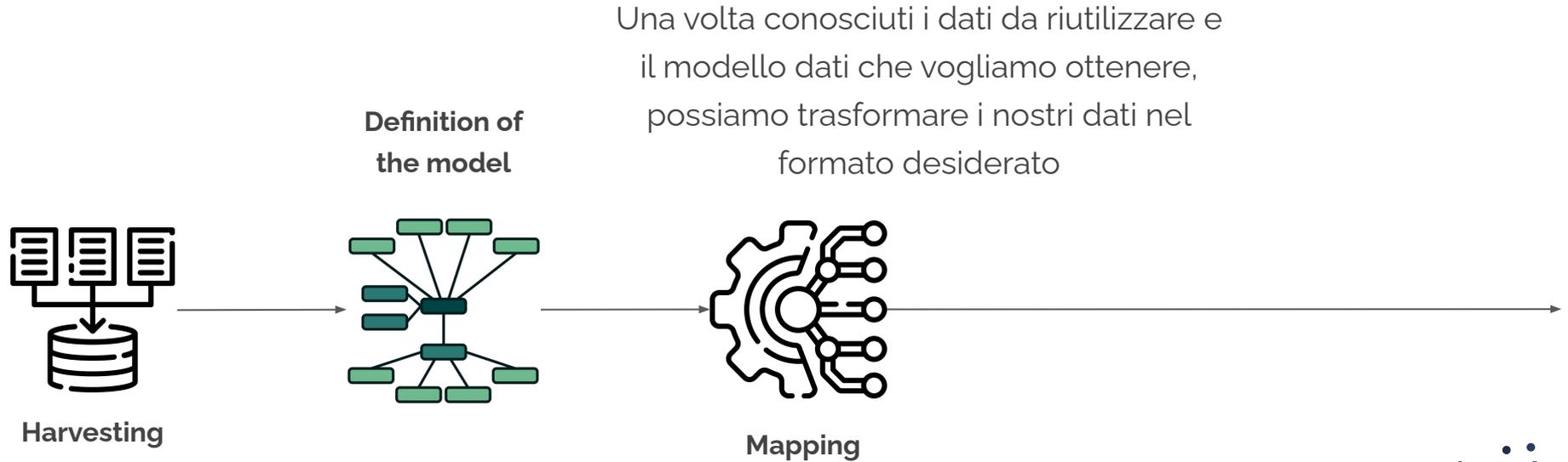


# Muruca: il workflow

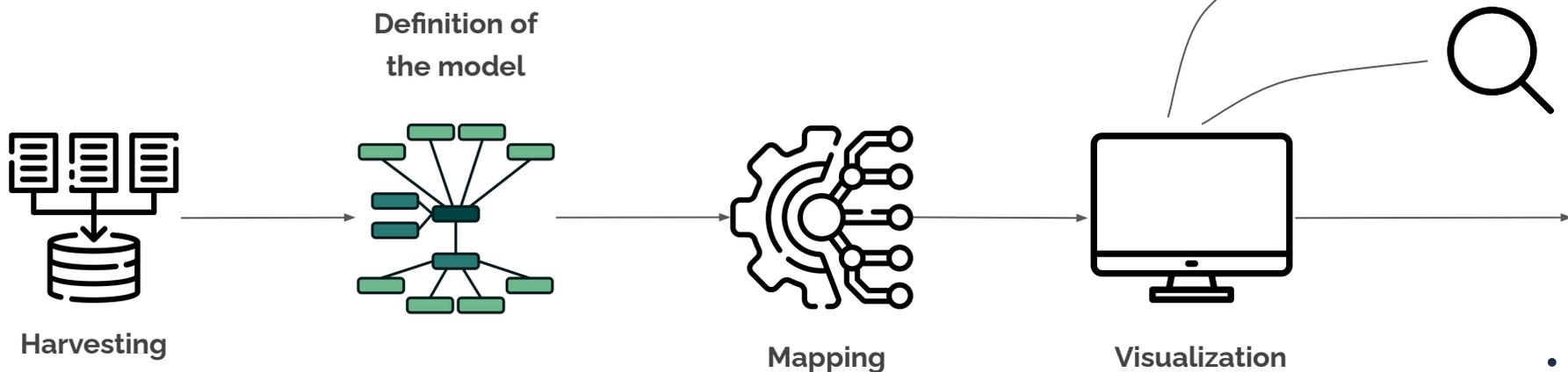


Si definisce un modello dati  
(es. ontologia) per  
normalizzare i dati recuperati  
da fonti eterogenee

# Muruca: il workflow



# Muruca: il workflow



Attraverso questa schermata è possibile ricercare contenuti specifici attraverso metadati oppure arricchirli tramite annotazioni semantiche.



*pin the web*

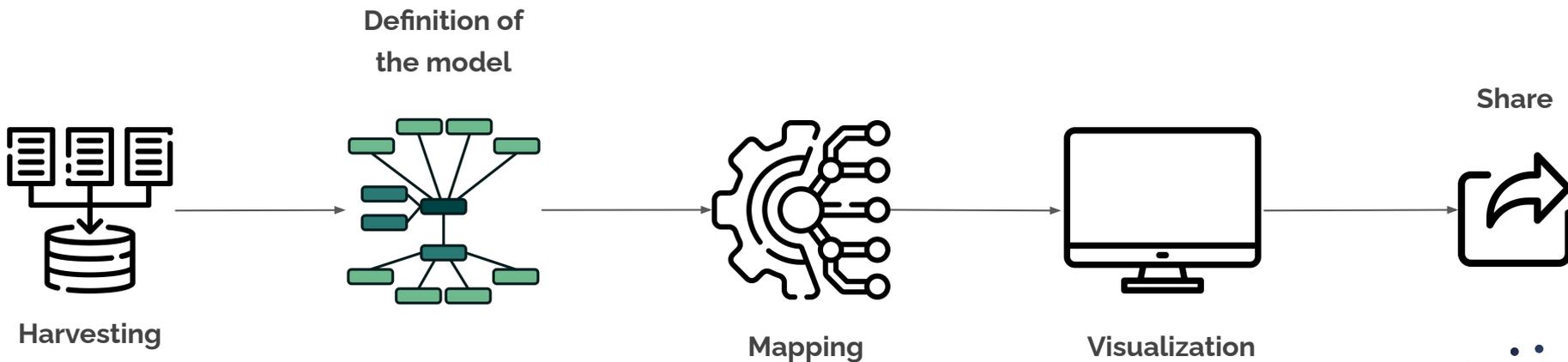


Una volta ottenuti i dati si possono visualizzare.



# Muruca: il workflow

Infine tutti i dati e data model vengono esposti tramite API, per incentivare la riutilizzabilità dei dati prodotti nel corso del progetto.



# Considerazioni finali

- Le scelte a livello infrastrutturale consentono una certa flessibilità:
  - Il backend non invecchia velocemente come le interfacce grafiche. Questo consente di preservare i dati nonostante il cambiamento delle modalità di visualizzazione del contenuto.
  - La possibilità di normalizzare i dati, consente una navigazione più fluida e soddisfacente per l'utente finale.
  - Le ricerche quantitative possono accedere a un quantitativo di risorse sufficiente per analisi accurate.
  - La definizione di un mapping, come un' ontologia, consente agli utenti finali di riutilizzare più efficacemente le risorse disponibili all'interno della piattaforma.

