



UNIVERSITÀ DI PISA

LAUREA MAGISTRALE IN
INFORMATICA UMANISTICA

SEMINARIO DI CULTURA DIGITALE A.A. 2016/17

Linked Data nell'ambito musicale

Rachele Bizzarri

Matricola: 504139

Sommario

Potenza e uso dei linked data in ambito musicale: la connessione delle informazioni per una distribuzione più efficace e efficiente sul web, per ricerche più specifiche e per strutturazioni di siti o web application volti all'indagine dei panorami musicali

INDICE

1. INTRODUZIONE.....	3
2. IL SEMANTIC WEB.....	4
2.1 Web 3.0 – Semantic Web.....	4
2.2 Linked Data.....	5
2.3 Le ontologie nei Linked Data.....	6
2.3.1 Music Ontology.....	7
3. I LINKED DATA MUSICALI.....	9
3.1 MusicBrainz.....	9
3.3 Last.fm.....	11
3.4 BBC.....	12
3.5 The Guardian.....	12
3.6 Linked Jazz.....	13
3.6.1 Fasi del progetto.....	13
3.6.2 Gli strumenti di realizzazione.....	14
3.6.3 API.....	15
4. CONSIDERAZIONI FINALI.....	16
5. BIBLIOGRAFIA E SITOGRAFIA.....	17

1. INTRODUZIONE

Quando pensiamo alla musica e a tutto l'ambiente che la circonda, non pensiamo mai a quante informazioni possono celarsi dietro ad un artista, un album, una casa discografica, un genere musicale. Non ci focalizziamo mai su questo aspetto perché comunque la musica è suono, è testo, è melodia, è armonia: non pensiamo mai che in realtà anch'essa può essere un dato, così come tutto il resto delle cose che ci circondano.

Pensiamo quante informazioni potrebbe fornirci la musica anche solo sul modo di vivere di un gruppo di persone. Sì, perché anche se siamo nell'era della globalizzazione in cui generalmente la tradizione viene in un certo modo soppiantata dal fantomatico mainstream, il tipo di musica che ascoltiamo non solo definisce noi come persona, ma definisce anche il mondo che ci circonda. Non va dimenticato quanto essa sia parte fondamentale della vita di ogni individuo (chi più chi meno), di quanto sia basilare per riallacciarsi alle proprie radici: la musica è un asse portante della cultura e della tradizione di un popolo e il caso maggiormente eclatante di questa presa di potenza della musica dentro un ambiente culturale è quello dell'Africa, dove tutt'oggi ogni rito viene accompagnato da ritmiche ben definite.

Questa relazione non vuole però focalizzarsi sull'aspetto della musica tradizionale di un paese e quanto questa segni ogni individuo che si ritiene appartenete ad essa, vuole però mostrare quanto anche il mondo della musica sia intricato di relazioni, sia intriso di ogni tipo di informazione. Per esempio, è noto che si fa risalire la nascita del punk nella Gran Bretagna degli anni Settanta, ma questo come può essere spiegato? Con i vari documenti online, certo. Però può venire esemplificato ancor più dettagliatamente tramite l'utilizzo della nuova frontiera del Web: il Semantic Web, la versione 3.0 della rete che tutti noi ormai conosciamo fin troppo bene, una rete costituita non più da documenti collegati tra loro mediante link ipertestuali, ma da dati messi in relazione tra di loro: una sorta di grande database accessibile da ogni dove che permette di trovare anche le più impensabili relazioni tra i vari elementi che lo compongono. Quindi, se torniamo all'esempio di prima, se riuscissimo a convertire in dati tutte le informazioni riguardanti la Gran Bretagna in quel dato periodo storico, sicuramente troveremo connessioni con l'insorgere del movimento sopracitato per determinati motivi, senza dover ricorrere a ricerche aggiuntive; questo è solo un possibile scenario di utilizzo del mondo dei linked data e del Web 3.0 e del grande potenziale che avranno.

Al momento non sono state fatte ricerche del genere, ma questi dati sono già largamente utilizzati da applicazioni web o in progetti che mirano al rappresentare al meglio la conoscenza, lo scibile e le informazioni che sono state caricate sul web. Per l'ambito musicale verranno presentati dei progetti e dei siti web che catalogano le loro informazioni musicali sotto forma di linked data e verrà presentato il Linked Jazz Project, il progetto che ha suscitato la realizzazione della seguente relazione, in cui sono mostrate tutte le relazioni che intercorrevano tra i vari artisti Jazz, ottenute grazie a delle estrapolazioni dai documenti online, trascrivendo interviste e cercando i dati negli archivi. Questo progetto mostra al meglio la potenza di questo approccio di pubblicazione dati e permette di visualizzare la conoscenza in ambito di musica jazz, cosa praticamente impossibile nel caso in cui fossero stati utilizzati documenti anziché dati.

2. IL SEMANTIC WEB

La seguente parte verterà sul chiarire cosa sia il Web 3.0 (o Semantic Web), cercando di specificare tutti gli elementi che lo compongono e provando a spiegare le migliorie che potrebbe portare alla rete globale grazie alla sua effettiva e totale diffusione.

Successivamente saranno spiegati i Linked Data, presentando le direttive proposte da Tim Berners-Lee nel 2009 (anno di lancio del progetto), arrivando a trattare delle ontologie web che permettono la rappresentazione della conoscenza nel “linguaggio dei computer”, focalizzando sull’ontologia utilizzata in ambito di dati musicali.

2.1 Web 3.0 – Semantic Web

Il Web 3.0, o Semantic Web, è la trasformazione del World Wide Web da “semplice” rete costituita da documenti collegati tra di loro tramite link ipertestuali a una rete costituita da dati che tipicamente troviamo catalogati e salvati su dei comuni database¹. Questo passaggio da documenti a dati permette di separare il contenuto di un qualsiasi artefatto in più parti che potranno essere collegate a informazioni contenute nello stesso artefatto e con altre contenute altrove, avendo a disposizione così molte più informazioni.

Come propone Tim Berners-Lee, possiamo vedere i documenti come semplici file da leggere e visualizzare, mentre i dati, essendo un’entità più grezza, permettono di essere processati in modi diversi riuscendo a ricavare molte (e nuove) informazioni rispetto ad un “semplice” documento.

L’obiettivo di questa nuova frontiera della rete è quello di permettere interrogazioni molto più efficaci ed efficienti rispetto a quelle che effettuiamo ogni giorno, in quanto si baseranno su delle parole chiave presenti all’interno del documento/risorsa e sulle varie connessioni che intercorrono tra le risorse presenti nella rete, connessioni molto più mirate rispetto a quelle che oggi sono realizzate dai vari collegamenti ipertestuali.

Il Web Semantico permetterà agli utenti di creare e accedere a delle banche dati, realizzate tramite delle regole che permetteranno di gestire e manipolare questi dati:

- RDF (Resource Description Framework), il modello standard di interscambio sul Web²;
- SPARQL, che permette di interrogare dati codificati con RDF³;
- OWL, che è un linguaggio basato sulla logica computazionale che permette di rappresentare la conoscenza complessa di risorse, gruppi di risorse e le varie relazioni tra queste⁴;
- SKOS, un metodo standard che permette di rappresentare sistemi di organizzazione della conoscenza tramite l’uso di RDF⁵.

Per rendere possibile il tutto, è necessario disporre di una grande quantità di dati, tutti disponibili mediante un modello standard che possa essere ricercato e manipolato dal Web 3.0.

¹ W3C, Semantic Web

² W3C, Semantic Web, RDF

³ W3C, SPARQL 1.1 Overview

⁴ W3C, Semantic Web, OWL

⁵ W3C, Semantic Web, SKOS

Importantissime diventeranno le relazioni che intercorrono tra i dati presenti in rete, perché permetteranno la realizzazione di quella vasta e ancor più connessa rete che è il Web.

Nella pagina dedicata al Semantic Web del W3C⁶, è riportato che quest'ultimo è costituito da due parti: i formati comuni per l'interrogazione e la combinazione dei dati che provengono da fonti tra loro diverse, permettendo quindi agli utenti e ai computer di partire da un database in cui sono riportate delle informazioni circa una risorsa fino ad approdare ad un altro in cui risiedono altre informazioni che sono connesse alla nostra ricerca.

2.2 Linked Data

I Linked Data sono una modalità di pubblicazione di dati strutturati che possono essere collegati tra di loro e utilizzati tramite delle interrogazioni semantiche.

Questa modalità si basa su delle tecnologie e degli standard web aperti e comuni. Il formato di codifica comune è RDF, che permette di convertire dati o accedere a diverse tipologie di banche dati già esistenti. Viene utilizzato anche lo standard di HTTP e quello URI (che racchiude sia URL-Uniform Resource Locator che URN-Uniform Resource Name), estendendone però l'applicazione così da poter fornire informazioni che possono essere lette e comprese dal computer: quindi i Linked Data possono essere visti come una collezione di set di dati correlati sul Web.

Per permettere la realizzazione e creazione di Linked Data, Tim Berners-Lee ha fornito quattro regole fondamentali⁷:

1. le risorse dei Linked Data sono nominate mediante delle URI;
2. le URI devono essere di tipo HTTP, così da permettere alle persone di cercare le risorse;
3. tutte le informazioni collegate a quel tipo di risorsa/URI devono essere fornite rispettando gli standard RDF e SPARQL;
4. collegare le varie URI così da poter scoprire ancora più cose dalla ricerca di una risorsa.

I Linked Data sono quindi la base della realizzazione del Web 3.0, dove sarà possibile realizzare un ambiente in cui qualsiasi web application potrà effettuare ricerche (query) su questi dati, elaborare delle conclusioni tramite queste grazie all'uso dei cosiddetti vocabolari. Questi vengono utilizzati per definire concetti e relazioni legati ai dati di una particolare area d'interesse, per classificare i vari termini che potranno essere successivamente utilizzati per realizzare una particolare applicazione web, permettendo l'integrazione tra i vari dati presenti in rete.

Come già detto, per poter realizzare questo web costituito da dati, sarà indispensabile la presenza di numerosi dati: ciò è stato più volte ripetuto anche da Tim Berners-Lee nella sua conferenza TED del 2009, in cui ha pronunciato (e fatto pronunciare al pubblico) "ROW DATA NOW!", dati grezzi ora⁸.

La potenza dei Linked Data e del Semantic Web di conseguenza, risiede nel fatto che sarà possibile accedere a una più vasta conoscenza dell'area di interesse delle applicazioni che si vorranno realizzare, grazie proprio alle relazioni che intercorreranno tra i dati.

⁶ W3C, Semantic Web Activity

⁷ W3C, Linked Data

⁸ Tim Berners-Lee, Conferenza TED sui Linked Data, 2009

2.3 Le ontologie nei Linked Data

In informatica, l'ontologia è una rappresentazione formale, condivisa ed esplicita di una determinata concettualizzazione di un particolare dominio d'interesse; si utilizza per classificare oggetti ed anche in intelligenza artificiale per permettere una più accurata rappresentazione della conoscenza.

Le ontologie consentono di definire delle relazioni contestuali seguendo un vocabolario e la sintassi formale è OWL (Web Ontology Web), la quale è un'estensione di RDFS (RDF Schema).

OWL è definita come una famiglia di linguaggi volti alla rappresentazione della conoscenza e definisce la struttura della stessa in vari domini d'interesse: i nomi rappresentano le classi degli oggetti/risorse e i verbi sono l'anello di congiunzione che permette di realizzare le relazioni tra questi oggetti.

I dati che sono descritti tramite un'ontologia della famiglia OWL vengono interpretati come un set di individui e un set di proprietà che sono connesse tra loro. Un'ontologia consiste, infatti, in un set di assiomi che contengono un gruppo di individui, detti classi, e un set di tipi di relazioni che intercorrono tra i vari individui. Le ontologie OWL possono importare altre ontologie, aggiungendo così informazioni ai dati codificati con una data ontologia.

I linguaggi di ontologia OWL permettono di creare classi, proprietà e definiscono istanze e operatori di congiunzione/relazione:

- un'istanza è un oggetto che corrisponde a un individuo logico;
- una classe è una collezione di oggetti che corrisponde a una descrizione logica di un concetto. Può avere molte istanze e queste possono appartenere a una, nessuna o più classi di riferimento;
- una proprietà è una caratteristica di una classe, rappresentata tramite una relazione binaria (ricordiamo che il Semantic Web è composto da triple, connesse tra loro con una relazione 1:1), che specifica degli attributi che sono veri per la classe di riferimento;
- gli operatori utilizzati dai linguaggi OWL sono di tipo unione, intersezione e complemento.

Fondamentale quindi per la codifica dei dati che andranno a costituire i linked data è il vocabolario di riferimento, che deve essere uno definito come uno specifico standard comune nel contesto d'uso.

Le ontologie portano con sé enormi vantaggi:

- permettono una navigazione coerente;
- permettono di creare delle connessioni tra le varie risorse, così da scoprire informazioni importanti che però sono sotto un'altra fonte;
- permettono la rappresentazione di qualsiasi forma di dato;
- permettono di descrivere una stessa cosa con dei concetti diversi, mantenendo le corrispondenze;
- permettono la deduzione di concetti correlati ad un concetto di partenza, grazie all'uso delle relazioni tra le risorse.

Quando si vuole pubblicare dei Linked Data sul web è quindi opportuno creare un nuovo vocabolario, nel caso in cui il mio dominio di interesse non ne abbia associato alcuno, oppure fare riferimento a dei vocabolari già esistenti.

2.3.1 Music Ontology

Per quanto riguarda i linked data in ambito musicale, l'ontologia di riferimento è denominata, appunto, Music Ontology. L'ontologia di riferimento si basa sul database di MusicBrainz e cerca di essere il più vicino possibile alle denominazioni lì utilizzate ma, allo stesso tempo, cerca di estendere il suo utilizzo e le sue strutture per potersi collegare anche a altri dati presenti in altri sistemi di catalogazione, come ad esempio MySpace, Pandora, etc.

La sua prima pubblicazione risale al 2006 e fornisce le regole per poter creare dei dati strutturati che siano pubblicabili sul semantic web.

Questo "vocabolario musicale" fornisce la possibilità di strutturare al meglio i dati, permettendo di costruire un sito musicale che si basa sui dati raccolti, l'aggregazione di dati provenienti da più fronti e arricchendo i risultati dei motori di ricerca per quel che riguarda le tracce, gli artisti e le opere musicali, disponendo il web di una sorta di database legato alla musica che raccolga tutte le informazioni inserite dagli utenti (nel caso degli open data)⁹.

La Music Ontology si basa su altre ontologie:

- FOAF (Friend Of A Friend), che è un vocabolario che viene utilizzato per descrivere le persone, i gruppi di persone e le organizzazioni;
- Event Ontology, un vocabolario che permette di codificare le informazioni riguardo la descrizione di eventi, permettendo anche di definire quale parte di una canzone costituisce il suo ritornello, ad esempio;
- Timeline Ontology, che permette di codificare e descrivere intervalli di tempo e intervalli su delle linee temporali, quindi anche intervalli presenti in un segnale audio;
- FRBR (Functional Requirements for Bibliographic Records), che è un'ontologia che permette di descrivere opere e le loro relazioni.

Come tutte le ontologie utilizzate per codificare e descrivere dei linked data, anche la music ontology è costruita su RDF, quindi composta da triple soggetto-predicato-oggetto, permettendo ai dati codificati di poter essere collegati con qualsiasi altro tipo di dato codificato utilizzando un linguaggio basato su RDF, permettendo il "linking" a dati di altro tipo.

Un esempio di come viene descritta una traccia tramite la music ontology :

- in HTML

```
<html prefix="mo: http://purl.org/ontology/mo/
        foaf: http://xmlns.com/foaf/0.1
        dc: http://purl.org/dc/elements/1.1/">
[... ]
<span typeof="mo:Track" about="#track-1">
  <span property="dc:title">Turnover</span>
  is a track by
  <a
    rel="foaf:maker"
    href="http://musicbrainz.org/artist/233fc3f3-6de2-465c-
    985e-e721dbabbace#_">
    <span property="foaf:name">Fugazi</span>
  </a>
</span>
</html>
```

⁹ FAQ Music Ontology, Why use the Music Ontology?

- in JSON-LD

```
{
  "@content":{
    "mo": "http://purl.org/ontology/mo",
    "foaf": http://xmlns.com/foaf/0.1,
    "dc": "http://purl.org/dc/elements/1.1/"
  },
  "@id": "http://example.org/#track1",
  "@type": "mo:Track",
  "dc:title": "Turnover",
  "foaf:maker": {
    "@id": "http://musicbrainz.org/artist/233fc3f3-6de2-465c-985e-e721dbabbace#_",
    "@type": "mo:MusicGroup",
    "foaf:name": "Fugazi"
  }
}
```

- in RDF/Turtle

```
@prefix mo: <http://purl.org/ontology/mo/> .
@prefix dc: <http://purl.org/dc/elements/1.1> .
@prefix foaf: <http://xmlns.com/foaf/0.1> .

<#track-1>
a mo:Track ;
dc:title "Turnover" ;
foaf:maker < http://musicbrainz.org/artist/233fc3f3-6de2-465c-985e-e721dbabbace>
a mo:MusicGroup ;
foaf:name "Fugazi" .
```

Nella pagina del sito è riportata una sezione dedicata alla spiegazione delle classi e delle proprietà che possono essere utilizzate con la music ontology.

3. I LINKED DATA MUSICALI

L'approccio Linked Data è arrivato anche nell'ambiente musicale, portando alla creazione di tanti siti di riferimento in cui sono raccolti dati riguardo artisti, album, etichette discografiche e quant'altro.

Il settore è molto vasto, ma solo alcune piattaforme e siti forniscono libero accesso ai propri dati tramite licenze open. Il testo che seguirà focalizzerà solo su quelli con libero accesso, in quanto effettivamente, solo quelle piattaforme, enciclopedie o siti di tale stampo, saranno quelli che permetteranno la reale riuscita del progetto partito nel 2009 da Tim Berners-Lee: la creazione del Web 3.0.

Dopo un'accurata ricerca, ho ritenuto opportuno focalizzarmi sulla vasta enciclopedia musicale MusicBrainz, in cui mi ero imbattuta in passato per la realizzazione di un progetto di esame. Successivamente ho passato in rassegna i maggiori e più famosi siti o piattaforme che si appoggiano al dataset offerto dall'enciclopedia, in quanto potrà essere più chiara la forza che i linked data possono avere se utilizzati con licenze libere.

Navigando in rete, sono arrivata alla pagina del Linked Jazz, un progetto che utilizza i linked data per mostrare agli utenti le relazioni che legavano i vari artisti jazz: ritenendolo un esempio di utilizzo di linked data molto valido, ho deciso di farne una breve introduzione

3.1 MusicBrainz

MusicBrainz è un'enciclopedia musicale aperta che raccoglie tutti i metadati connessi alla musica e li rende disponibili al popolo del web. Quest'enciclopedia permette di ottenere le più recenti informazioni in ambito musicale, permettendo a chiunque di contribuire al loro rilascio in rete, sempre tramite una licenza libera, fornendo una sorta di lingua franca musicale che permette di identificare le risorse musicali sia da parte dell'utente che dalla macchina.

MusicBrainz è una sorta di Wikipedia della musica, in quanto è mantenuta in vita da una comunità globale di utenti che cura le nuove informazioni, modifica le vecchie e controlla che tutto sia il più veritiero possibile, per permettere un'accurata catalogazione dei dati musicali sotto forma di linked data¹⁰.

MusicBrainz è costituito da tre sezioni:

- l'interfaccia web con il backend, che è il database di supporto;
- la documentazione, per permettere agli utenti di ampliare il database e apportare eventuali modifiche;
- le applicazioni e i servizi che mette a disposizione degli utenti.

Nella Beginners Guide¹¹ troviamo subito un monito volto proprio ai principianti:

"If you're like most users, you will probably start to tag your files with information that was already in the database. But remember that MusicBrainz is user-edited, so if you find that something is missing, you can add it, and if it is wrong, you can

¹⁰ Peter Haase, Mapping, Interlinking and Exposing MusicBrainz as Linked Data

¹¹ MusicBrainz, Beginners Guide

fix it! These changes are not always applied instantly, because they pass through our voting system.”

“Se siete come la maggior parte degli utenti, probabilmente inizierete con il codificare i vostri file utilizzando informazioni che saranno già presenti nel database. Ma ricordate che MusicBrainz viene curato dagli utenti, quindi se vi accorgete che manca qualcosa, potete aggiungerla, e se notate che qualcosa è sbagliato potrete correggerla! Questi cambiamenti non saranno sempre applicati immediatamente, perché passando dal nostro sistema di votazione.”

Se prestiamo bene attenzione a questo monito, ci rendiamo conto di come le risorse catalogate con i Linked Data sono effettivamente trattate come un qualsiasi valore presente all'interno di una tabella di database: vengono fortemente scoraggiate le ridondanze, le ripetizioni. Effettivamente ci troviamo di fronte alla costruzione di un grande archivio dati che contiene tutte le informazioni possibili; in questo caso, informazioni riguardanti l'ambito musicale.

In MusicBrainz, ogni CD-album è contraddistinto da un ID, un identificatore che contiene tutte le informazioni necessarie alla sua identificazione univoca, che permette il riconoscimento automatico della risorsa (di nuovo, proprio come accade in un database relazionale con le chiavi primarie). Ovviamente questa specificazione dell'album non è esclusiva: questo formato di identificazione vale anche per gli altri tipi di risorsa.

Per quanto riguarda le applicazioni e i servizi, vediamo che offre vari tipi di strumenti che permettono di lavorare con i dati di MusicBrainz:

- Desktop software
 - MusicBrainz Picard, un'applicazione scritta in Python (fruibile su Linux, Mac OS X e Windows), che supporta la maggior parte dei formati audio e che è ritenuta l'applicazione ufficiale di tagging offerta da MusicBrainz;
 - Magic MP3 Tagger, un tagger di Windows per i file .mp3 che sponsorizza MusicBrainz e utilizza dati fondamentali di MusicBrainz;
 - Yate Music Tagger, un tagger di Mac OS X che supporta l'integrazione tra MusicBrainz, Discogs, Beatport, AcustID, AcousticBrainz, Fanart.tv e iTunes. Supporta anche i file video .m4v e .mp4.
- Mobile apps
 - MusicBrainz for Android, un'applicazione di Android che permette di vedere informazioni riguardo le release, inviare tag e visualizzare codici a barre per MusicBrainz.
- MusicBrainz services
 - MusicBrainz Server, il frontend al database;
 - MusicBrainz Database, completamente di pubblico dominio, permette di scaricare e utilizzare i dati per qualsiasi scopo;
 - Live Data Feed, permette di accedere tempestivamente ai datai di MusicBrainz, il tutto per un uso non commerciale;
 - FReeDB Gateway.

MusicBrainz utilizza RDF e XML per la definizione dei suoi metadati, per rendere possibili le automatizzazioni via HTTP con i metodi portanti del protocollo: GET e POST. L'uso di queste due

strutture permette il linking ad altri vari dati presenti in rete, riuscendo così ad ottenere il “mondo” linked data dell’ambiente musicale.

Altri esempi di database online che catalogano dati di carattere musicale con una licenza open, sono:

- Discogs, sito web e database che mette a disposizione dell’utente delle API e struttura i dati mediante XML; è noto per essere uno dei maggiori database dedicati alla musica elettronica, nonostante conti al suo interno anche numerosi dati suddivisi per genere musicale;
- The Echo Nest, piattaforma di intelligenza artificiale musicale, che cerca di capire il contenuto degli audio e dei testi dei file musicali registrati, volta a un’analisi degli stessi; mette a disposizione degli utenti sviluppatori delle API
- Quantone, società londinese di musica artificiale che fornisce una ricca collezione di metadati semantici riguardanti la musica; tramite l’approccio linked data, permette profondi collegamenti tra i dati raccolti fornendo informazioni su artisti, opere e relazioni, produttori, addetti ai lavori del settore, musicisti presenti nella fase di registrazione, informazioni riguardanti il copyright e molto altro ancora; offre API agli utenti sviluppatori;
- Musixmatch, fondata a Bologna nel 2010, è la maggiore piattaforma di testi di canzoni al mondo che permette di sincronizzare testo e traccia audio durante la riproduzione del file; offre il libero accesso ai suoi dati.

3.3 Last.fm

Last.fm è una radio su internet il cui sistema permette ad ogni utente di costruirsi un proprio profilo in cui sono riportati i suoi artisti e canzoni preferite, includendo anche le canzoni ascoltate tramite la radio online o tramite l’installazione di un plugin all’interno del player musicale che viene utilizzato dallo stesso per ascoltare la musica.

Il sito permette di visualizzare delle statistiche in tempo reale riguardo artisti e tracce più ascoltate nell’ultima settimana, ultimi tre, sei e dodici mesi.

Last.fm, per promuoversi come sito di pubblicizzazione e radio online, incentiva le etichette discografiche e gli artisti a pubblicare e promuoversi sul sito, garantendo loro la proposta dei loro lavori agli utenti, selezionati in base alle loro preferenze grazie a degli appositi filtri realizzati tramite algoritmi di ricerca.

Il sito mette a disposizione degli utenti sviluppatori l’accesso alle API, permettendo a chiunque di “creare il proprio programma utilizzando i dati di Last.fm”¹². Nella pagina riguardante le API, vengono riportati tutti i metodi che Last.fm mette a disposizione per l’interrogazione dei propri dati, suddivisi per categorie (album, artisti, etc.).

Anche Last.fm poggia molto sulla codifica dei metadati presenti in MusicBrainz, questo per poter realizzare l’obiettivo fondante del Web 3.0: realizzare una rete interconnessa che permetta lo scambio di informazioni provenienti da ogni dove, creando continue relazioni tra le varie fonti.

Per poter accedere alle API è necessaria un’API key, che permetterà di accedere ai dati liberamente, ottenuta dopo aver effettuato l’iscrizione al sito e aver aggiunto un account specifico.

I parametri di ricerca conterranno:

¹² Pagina delle API di Last.fm

- artista, parametro obbligatorio, in cui verrà inserito il nome dell'artista, in mancanza del quale basta l'mbid: l'id di MusicBrainz associato all'artista;
- album, parametro obbligatorio, il nome dell'album, in mancanza del quale basta l'mbid: l'id di MusicBrainz associato all'album;
- parametri opzionali;
- apy_key, parametro richiesto per la visualizzazione della query effettuata sui dati di Last.fm.

I dati ottenuti dalla ricerca saranno visualizzati o in formato JSON o in XML.

3.4 BBC

Anche la British Broadcasting Corporation (BBC) ha iniziato ad utilizzare i linked data. La BBC gestisce numerose stazioni radio e canali televisivi, per questo ha iniziato ad utilizzare tale tecnologia, così da poter tenere collegati i dati, assieme a quelli di DBpedia e MusicBrainz. Utilizza dei vocabolari di riferimento ben controllati, così da poter collegare al meglio i vari contenuti in suo possesso riferiti ad un solito argomento che può trovarsi in repository differenti, permettendo l'aumento del contenuto con varie relazione a dati aggiuntivi che forniscono informazioni riguardo quel dato contenuto¹³.

La BBC utilizza e contribuisce ad arricchire il dataset di MusicBrainz, così come altre grandi aziende del settore: la Universal Music utilizza i suoi metadati per poter arricchire i propri, Amazon ne usufruisce per le sue offerte musicali. Anche Spotify fa largo uso del dataset di MusicBrainz per il suo servizio di streaming musicale. Nel mondo linked data, MusicBrainz e Wikidata sono interconnessi: Wikidata utilizza i dataset MB per arricchire la propria serie di open data, così come fa anche MusicBrainz con i data catalogati in Wikidata, creando così una forte correlazione tra i dati musicali e le informazioni ad essi collegati, come ad esempio la provenienza di un'artista, la descrizione del genere musicale di riferimento e tanto altro ancora.

3.5 The Guardian

Anche il quotidiano britannico The Guardian da qualche anno ha ampliato la propria piattaforma The Guardian – Open Platform servendosi dell'approccio linked data. Dal 2010 circa, ha aggiunto identificatori per gruppi musicali e libri.

Viene definito come una "suite di servizi web" che rendono possibile la realizzazione di web application che utilizzano informazioni e contenuti del quotidiano. Le API messe a disposizione degli utenti sviluppatori permettono di interrogare il database del The Guardian in cui sono contenuti tutti gli articoli pubblicati, in formati adatti alla manipolazione da parte di altre applicazioni web. Seguendo il percorso intrapreso dalla radio online Last.fm, il quotidiano ha scelto di includere nel proprio dataset identificatori esterni: per quanto riguarda i libri, gli identificatori ISBN, per quanto riguarda l'ambito musicale, gli ID di MusicBrainz.

Concentrandoci sui dati a stampo musicale, The Guardian utilizza i metadati di MusicBrainz per circa 600 artisti, trattati nei suoi contenuti (questi dati sono riferiti al 2010, molto probabilmente adesso

¹³ Linked Data – The Story So Far, C. Bizer, T. Heath, T. Berners-Lee, BBC Programmes And Music

il numero sarà largamente cresciuto). A motivazione di questa scelta, vediamo sempre sulla pagina del blog del quotidiano dedicata ai linked data, che è stata fatta perché il modello utilizzato dal The Guardian per trattare artisti e band è quello di considerarli come oggetti primari a cui vengono associate altre informazioni. Proprio come viene strutturato il dataset di MusicBrainz: soggetto/artista/band musicale – predicato – oggetto/informazione aggiuntiva.

A concludere l'articolo del blog viene scritto che:

"[...] our requirement was to facilitate accurate joining to and from our content [...]"

"[...] il nostro intento era quello di facilitare la correlazione da e per i nostri contenuti [...]"

abbracciando totalmente l'approccio linked data e la visione della rete supportata dal Web 3.0.

Come per gli altri siti che mettono a disposizione i propri dati tramite delle API, è necessaria l'iscrizione che permetta di avere associata una propria API Key, tramite la quale effettuare ricerche sul vasto dataset libero che il quotidiano mette a disposizione di tutti.

3.6 Linked Jazz

Linked Jazz è connesso a un progetto di ricerca che indaga le capacità delle tecnologie Linked Open Data per quanto riguarda i materiali digitali culturali. Gli obiettivi di questo progetto sono la scoperta delle connessioni che sono riportate nei documenti e nei dati relativi alle vite sia personali che professionali degli artisti jazz e lo sviluppo di strumenti che permetteranno di lavorare con i linked open data, indipendentemente dalla loro natura¹⁴.

I materiali utilizzati in questo progetto sono materiali digitali riguardanti la storia del jazz e, tramite la loro manipolazione e codifica, permettono di visualizzare e rivelare la rete della comunità del jazz, mostrando come queste nuove modalità di collegamento dei dati, nate con l'avvento dei Linked Data, abbiano il potenziale di permettere nuove vie di ricerca.

Il progetto è tutt'ora in corso, e il team che ci lavora amplia le fonti di riferimento, aggiungendo dati, per poter includere altri documenti e set di dati collegati alla musica.

Il team è capitanato dalla professoressa Cristina Pauttelli, professore associato della School of Information Science del Pratt Institute di New York.

3.6.1 Fasi del progetto

Come prima fase del progetto, il team ha sperimentato le proprie idee basandosi sui contenuti delle trascrizioni delle interviste fornite da vari enti, tra cui il Rutgers Institute for Jazz Studies, la più grande libreria e archivio del jazz e di tutto il materiale jazz del mondo, generando triple RDF. Successivamente, per poter realizzare un Linked Open Data (LOD), il team ha sviluppato una serie di strumenti che permettessero di velocizzare i vari lavori, tra cui un analizzatore di trascrizioni e uno strumento di mappatura dei nomi. Questi strumenti, assieme, permettono di risalire e trovare i nomi che vengono pronunciati all'interno di un'intervista, assegnandogli un identificatore in base alla ricerca incrociata della risorsa su DBpedia.

¹⁴ Linked Jazz

Successivamente, viene passato il tutto ad un ulteriore strumento sviluppato dal team: il Linked Jazz 52nd Street, uno strumento di crowdsourcing, attraverso il quale viene chiesto a degli utenti volontari di assegnare dei termini attraverso i quali si sviluppa l'intervista, trovando le parole chiave delle future relazioni tra i dati raccolti preventivamente con gli altri strumenti.

Concluso questo ciclo di elaborazione, il tutto viene mandato ad un visualizzatore online che raccoglie le persone menzionate nelle varie interviste, rappresentata visivamente da una rete fatta di nodi connessi da relazioni: la grandezza del nodo/immagine dell'artista dipende da quante volte viene menzionato nelle trascrizioni.¹⁵

3.6.2 Gli strumenti di realizzazione

Il Transcript Analyzer viene utilizzato per caricare le interviste trascritte, rilasciate dai vari archivi che hanno supportato il progetto, per poter identificare i nomi degli artisti jazz citati durante le interviste sfruttando la cartella in cui sono riportati i nomi. Il tool utilizza l'elaborazione del linguaggio naturale per aggiungere eventuali nomi che non sono già stati salvati nella directory creata dal Linked Jazz Project. L'analizzatore, successivamente, suddivide le interviste in vari segmenti costituiti da domanda-risposta che saranno successivamente implementati nel Linked Jazz 52nd Street, il tool di crowdsourcing¹⁶.

Il Name Mapping and Curator, invece, supporta la creazione della directory di riferimento in cui verranno salvati i nomi degli artisti Jazz trovati tramite il Transcript Analyzer, grazie anche all'aiuto dei linked data di DBpedia, permettendo così una accurata identificazione del nome e rimuovendo il problema delle ridondanze/ripetizioni all'interno della struttura dati che si creerà in ultima istanza¹⁷.

Per permettere questo passaggio di disambiguazione delle istanze/risorse, il team del Linked Jazz Project ha realizzato Ecco!, un'applicazione di LOD che permette di conciliare le entità denominate con le URI provenienti da fonti autorevoli. Ecco! fornisce un'interfaccia utente che supporta il lavoro collaborativo degli utenti, in quanto il team ha deciso di permettere agli utenti di attuare delle raffinerie riguardo ai dati estrapolati dalle trascrizioni mediante i tool sviluppati tramite dei processi di convalida dei vari riferimenti risorsa estrapolata/risorsa autorevole¹⁸.

Come accennato in precedenza, il Linked Jazz 52nd Street è lo strumento sviluppato per il crowdsourcing, attraverso il quale gli esperti di Jazz o gli appassionati del genere aiutano il team nella loro impresa, individuando il tipo di rapporto che intercorreva tra gli individui e controllando i dati estrapolati con gli strumenti presentati in precedenza dalle interviste trascritte. Il tool è una web app che chiede la registrazione al sito per poter effettuare le proprie rifiniture sui dati raccolti nel corso del progetto. Il tutto sarà successivamente convertito in RDF che andrà ad ampliare il dataset dei LOD del Linked Jazz Project¹⁹.

Come ultimo strumento sviluppato dal team, abbiamo quello che permette la visualizzazione dei dati raccolti, il Linked Jazz Network Visualization, anch'esso una web application che offre varie visualizzazioni interattive sui dati. Nel Network Visualization vengono mostrati anche immagini, video e piccole biografie degli artisti jazz presenti nella rete dei linked data estrapolati dai vari processi sopra elencati²⁰.

¹⁵ Linked Jazz – About The Project

¹⁶ Linked Jazz – Tools – Transcript Analyzer

¹⁷ Linked Jazz – Tools – Name Mapping and Curator

¹⁸ Linked Jazz – Tools – Ecco!

¹⁹ Linked Jazz – Tools, Linked Jazz 52nd Street

²⁰ Linked Jazz – Tools, Linked Jazz Network Visualization Tool

3.6.3 API

Per rimanere conformi all'approccio LOD, Linked Jazz mette a disposizione degli utenti sviluppatori delle API²¹ che permettono di agganciarsi a tre specifiche dei dati raccolti durante il progetto: people, relationships e text. I dati risultanti dalle richieste di tipo GET sono in formato JSON e la URL di base a cui effettuare le richieste è <http://linkedjazz.org/api/>, a cui far seguire lo specifico endpoint sul quale effettuare la ricerca:

- people, i cui risultati verteranno su:
 - name, in cui sarà riportato il nome dell'artista;
 - uri, in cui ci sarà la URI di riferimento per quella data risorsa;
 - uriEncoded, l'URI cifrato utilizzato dall'API;
 - image, l'eventuale immagine associata alla persona cercata;
 - comment, il commento sull'artista nel caso in cui siano riportate delle informazioni aggiuntive estrapolate dai vari tool utilizzati per la realizzazione del Linked Jazz Project;
- relationships, i cui risultati saranno sulle relazioni che ha una determinata persona e tutte le relazioni che collegano la rete degli artisti;
- text, le cui richieste restituiranno la trascrizione delle interviste in base all'ID passato nella GET.

Oltre a poggiare sui dataset di DBpedia, Linked Jazz fa largo uso del dataset di MusicBrainz, confermando così la potenza delle relazioni dei linked data sul web e mostrando i grandi vantaggi che questo approccio di pubblicazione online porta con sé: una vasta rete di dati provenienti da più fonti totalmente collegata per organizzare e strutturare meglio la conoscenza.

²¹ Linked Jazz – API

4. CONSIDERAZIONI FINALI

Dopo questa panoramica sui LOD musicali, è possibile fare delle considerazioni.

Partendo dal Linked Jazz Project e tralasciando il fatto che ha permesso una maggior conoscenza del jazz non solo agli addetti ai lavori, ma anche a chi si è imbattuto nel progetto, a chi lo ha utilizzato per delle ricerche, scoprendo relazioni tra artisti di cui forse non era a conoscenza, e che abbia permesso una più accurata diffusione di educazione riguardo il genere musicale, il progetto ha mostrato la potenza dei LOD, la loro enorme capacità di rappresentare una rete composta esclusivamente da materiale umanistico, o comunque culturale. Ciò apre in qualche modo la visione di un web composto da dati interconnessi tra loro, dati di qualsiasi tipo. Pensiamo ad esempio un'applicazione di queste tecniche di Linked Data ad archivi storici, agli artisti d'arte: potremmo scoprire e riprodurre una fitta rete di relazioni che potrebbero segnare un'epoca, una corrente, un filone di pensiero.

Sfruttando la potenza dei linked data potremmo estrapolare informazioni che molto probabilmente non credevamo possibili, potremmo scoprire relazioni di causa effetto a cui non avevamo mai pensato, potremmo letteralmente visualizzare la conoscenza, basandoci sui datai grezzi: potremmo compiere le più disparate operazioni con l'utilizzo dei linked data e del web 3.0, basta continuare a chiedere "ROW DATA NOW!", in linea con Berners-Lee.

5. BIBLIOGRAFIA E SITOGRAFIA

- Christian Bizer, Tom Heath, Tim Berners-Lee, Linked Data – The Story So Far
<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>
- W3C, Standards, Semantic Web, Linked Data
<https://www.w3.org/standards/semanticweb/data>
- Linked Open Data Italia
<http://www.linkedopendata.it/semantic-web>
- W3C, Semantic Web
<https://www.w3.org/standards/semanticweb/>
- W3C, Semantic Web, RDF
<https://www.w3.org/RDF/>
- W3C, SPARQL 1.1 Overview
<https://www.w3.org/TR/2013/REC-sparql11-overview-20130321/>
- W3C, Semantic Web, OWL
<https://www.w3.org/OWL/>
- W3C, Semantic Web, SKOS
<https://www.w3.org/2004/02/skos/intro>
- W3C, Semantic Web Activity
<https://www.w3.org/2001/sw/>
- W3C, Linked Data
<https://www.w3.org/DesignIssues/LinkedData.html>
- Tim Berners-Lee, Conferenza TED su Linked Data, 2009
https://www.ted.com/talks/tim_berners_lee_on_the_next_web
- Linked Data – The Story So Far, C. Bizer, T. Heath, T. Berners-Lee, BBC Programmes And Music
<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>
- Peter Haase, Mapping, Interlinking and Exposing MusicBrainz as Linked Data
https://www.slideshare.net/phaase/mapping-interlinking-and-exposing-musicbrainz-as-linked-data?qid=f5ed703d-1c68-493e-8a08-0133c5f50843&v=&b=&from_search=12
- MusicBrainz, Documentation, Beginners Guide
https://musicbrainz.org/doc/Beginners_Guide
- Music Ontology, FAQ
<http://www.musicontology.com/docs/faq.html>
- Music Ontology, Getting started
<http://musicontology.com/docs/getting-started.html>
- Last.fm – API
<https://www.last.fm/api>
- The Guardian, Blog – Linked Data at The Guardian
<https://www.theguardian.com/open-platform/blog/linked-data-open-platform>
- ProgrammableWeb – Guardian Makes Music and Literary Apps Easier with Linked Data
<https://www.programmableweb.com/news/guardian-makes-music-and-literary-apps-easier-linked-data/2010/10/20>
- Linked Jazz
<https://linkedjazz.org/>
- Linked Jazz - About The Project
<https://linkedjazz.org/about-the-project/>

Linked Jazz – Tools

<https://linkedjazz.org/tools/>

Linked Jazz – API

<https://linkedjazz.org/api/>