LOD IT BE

1 - STUDY OF THE DOMAIN

Originally named "Get Back", "**Let It Be**" went down in history as one of the first examples of a multimedia project in modern entertainment. Since its original release on May 8th 1970, new projects and materials were constantly released, enriching the already existing resources and bringing new life to its legacy.

When the Beatles started working on the "Let It Be" sessions, towards the end of 1968 and mostly in January of 1969, they were approaching the crisis that eventually would have led to their breaking up. That's why they decided to try something new, to "Get back" to the original enthusiasm that they used to live in their first years. To do that, they started working on an ambitious project that would have included:

- A new **record** with an authentic raw sound;
- A **movie** that would have told the audience how it was like to "live" and "work" with The Beatles: to make sure to capture the real daily life moments of a beatle, they started being recorded almost 24/7;
- A **book**, with transcripts of conversations, lyrics and interviews;
- A live **concert** that would have brought The Beatles in front of a live audience for the first time after almost 4 years.

The rest is history. The project, instead of lighting up The Beatles, contributed to bring them down, showing to the public the difficulties in their relationships. The concert, which was supposed to take place in a majestic location like the Sahara, was held in the quickest way possible: on the rooftop of their recording studios.

But right after the project's conclusion and publication, it instantly became an icon in pop culture: referenced as the last act of a legendary career, during the years "Let it Be" and the "Rooftop concert" have been tributed in many ways, expanding the story that lies behind this event:

- The **instruments** that were used for it were put in museums;
- Studies and remixes of the original **tapes** were carried out;

- **Photographic exhibitions** about them were held;
- **Videogames** that could let anyone experience it in first person the Rooftop Concert have been developed.

Finally, in 2021 the highly-awarded movie director Peter Jackson released its own version of the story, working on a new release of the original film material that was shot for the 1970 movie, adding a new piece to the puzzle.

Hence, such a project represents a great and rich source to work with in a **Linked Open Data** environment, exploiting the **semantic web** tools in order to conceptualize this evolving network of artists, events, movies, places, instruments, books and music.

These resources, though, come from a vast range of different institutions and their linkings are often hidden or never declared. That's why we decided to work on the alignment and the semantic connection of **10 items** that could best represent the "Let it Be" experience:

1) **Get Back (song)** was released as a single on 11/4/1969.
2) **Let It Be (vinyl)** is the original format through which the album was released in 1970.
3) **The Beatles: Rock Band** is a videogame released in 2009 that revolves around the career of the rock band.
4) **Savile Row, 3.** Here The Beatles recorded most of the record and performed their last concert on the rooftop of this building.
5) **Photo of Paul and John.** Paul McCartney and John Lennon during the recording session of “Let It Be”.
6) **Paul’s bass.** The bass was played by Paul McCartney during his career with The Beatles.
7) **The Beatles: Get Back** is a book published in 2021 that represents an account of the creation of Let It Be.
8) **Let It Be (movie).** This is a documentary film about the recording sessions and the Rooftop concert.
9) **Rooftop Concert** is the last Beatles’ concert: It was performed on 30th January 1969 on the roof of the Apple Corps building.
10) **Len’s tape.** Recording tape containing “Dig a Pony” and “I’ve got a Feeling”.

2 - KNOWLEDGE ORGANIZATION: ELABORATING MODELS

CONCEPTUAL MAP

After having found and defined 10 items connected to the main topic of our project, the next goal was to describe the idea that lies behind the “Let It Be” project through a conceptual map that could represent their relations, to give a first description of the chosen objects. During this step we complied with metadata and information already provided by the institutions, connecting only a few of the items and describing them using natural language.

The result is a first conceptual map that highlights the main features connected to “Let It Be”.

E/R MODEL

Starting from the conceptual map, we devised the first E/R model. Here we summarized the general relations between objects describing each triple in order to understand how we could formalize our network.

To do so, we defined the objects as a set of entities and attributes, linked to each other thanks to primary keys, foreign keys and associations with different grades and cardinalities (0,n; 1,n; n,n): this allows to better define which kind of linkings between the entities and attributes are required and/or permitted, abstracting the view towards a more general level.

METADATA ANALYSIS

Then, we analyzed data taken from the providers to understand which kind of standard they used and which general guidelines were followed. Almost all the institutions clearly expressed the standard they referred to as reported in the table below.

- The song and the concert are described on “MusicBrainz”, an open music encyclopedia that collects music metadata, and it uses its own metadata schema based on XML Schema: **MMD XML Schema**[1]. It serves a generic but

[1] https://github.com/metabrainz/mmd-schema .

efficient purpose in describing musical objects, providing metadata useful to describe responsibilities, WEMI parameters and musical information.

- We chose to use “Discogs” as our provider for the vinyl format of the album because, differently from Musicbrainz, it provides information about the specific formats through which the album was released throughout the years. Discogs is the largest online database for music and it uses its own metadata schema: **DDS XML Schema**[2], and its documentation and endspaces are freely accessible. It provides metadata elements similar to the ones described by MMD XML Schema.

- The videogame is described in “IGDb” (Internet Game Database): IGDb is owned by Amazon, and it follows a proprietary metadata schema described on the official documentation provided for the use of its API, the **IGDb JSON API**[3], which allows to obtain automatically informations about any existing videogame, their creation, their rating, rankings and so on.

- Regarding the location, we collected the related information thanks to “Ordnance Survey”, the institution responsible for mapping the United Kingdom. As described in their guidelines and standards section, the description of geographical data sets is done in accordance to **ISO 19115**[4] metadata schema for the description of geographic entities, geographic data, geographic datasets and individual geographic features and feature properties. It hence defines: mandatory and conditional metadata sections, metadata entities and metadata elements; the minimum set of elements (core metadata) required by the full range of metadata applications (data discovery, determination of relevance for use, data access, data transfer and use of numerical data); optional metadata elements to allow for a more extensive standard description of geographic data, if required; a method of extending metadata to respond to specialist requests.

- We found the photo representing John Lennon and Paul McCartney thanks to the online collection of the “National Portrait Gallery”. This provider does not explicitly declare which standard it uses, so we have chosen **IPTC**[5] for the alignment, since it’s a standard typically applied to image contents: it structures and defines metadata properties that allow users to add precise and reliable data about images and people, locations and products shown in an image. It also supports dates, names and identifiers regarding the creation of the photo, and a flexible way to express rights information.

---

[2] https://support.discogs.com/hc/en-us/sections/360001058453-Full-Submission-Guidelines .
[3] https://api-docs.igdb.com/#endpoints .
[4] https://www.iso.org/obp/ui/#iso:std:iso:19115:-1:ed-1:v1:en .
[5] https://www.iptc.org/std/photometadata/specification/IPTC-PhotoMetadata#about-the-standard .

- The “MET Museum”, that physically holds “Paul’s bass”, uses metadata created *ad hoc* by the institution to describe the item[6].

- The book is described on Worldcat, the world's largest library catalog, that uses **MARCXML[7]** as a standard for the exchange of “machine-readable” information about bibliographic records. It includes formats for authority records, holdings records, classification schedules, and community information, in addition to the format for bibliographic records.

- The movie is described by IMDb (Internet Movie Database), owned by Amazon, and similarly to IGDb it follows a proprietary metadata schema described on the official documentation provided for the use of its API, the **IMDb metadata schema[8]**.

- The tape is held by the “Museum of master tape reels”: it doesn’t provide any information about the standard used, so we’ve chosen the “**Technical Metadata for Audio XML schema[9]**” provided by the Library of Congress: it contains technical metadata that describes either a digital file or a physical source object. It is based upon metadata elements developed by the Library of Congress.

METADATA ALIGNMENT

During the alignment process we mainly maintained the standards chosen by the institutions. In a few cases, though, we decided to adopt standards that better suited our objects and that, at the same time, are more widely used (so as to ensure interoperability).

- For the videogame we used **schema.org[10]**, a set of documentations and guidelines that serve exactly this purpose, built in compliance with RDF rules by Google and Microsoft;
- The standard chosen for the music instrument was **CIDOC-CRM[11]**. The use of this standard is encouraged in the GLAMs domain and by other cultural institutions to enhance accessibility to museum-related information and knowledge for museum documentation;
- Since MARC is commonly implemented along with ISBD, we decided to align the metadata according to the RDA elements in order to focus on a more conceptual level. **RDA (Resource Description and Access)[12]** offers a “set of

[6] https://github.com/metmuseum/openaccess .
[7] https://www.loc.gov/standards/marcxml/ .
[8] https://developer.imdb.com/documentation/bulk-data-documentation/data-dictionary/?ref_=up_next .
[9] https://www.loc.gov/standards/amdvmd/audiovideoMDschemas.html .
[10] https://schema.org/docs/schemas.html .
[11] https://www.cidoc-crm.org/html/cidoc_crm_v7.1.2.html .
[12] https://www.rdaregistry.info/ .

guidelines and instructions to identify entities of Group 1 and Group 2 of FRBR (Work, Expression, Manifestation, Item, Person, Family, Corporate Body): to relate entities of Group 1 and Group 2 of FRBR, by means of relationships; to enable users to represent and discover entities of Group 1 and Group 2 by means of their attributes and relationships.

- The **Ontology for Media Resources**[13] has been used to describe the movie. It is a W3C recommendation that aims to define "a core set of metadata properties for media resources, along with their mappings to elements from a set of existing metadata formats".

The documentation of each standard is linked in the table.

Settled the standards, we focused on the actual alignment. We created a table to show the matches in metadata elements and properties useful to describe our objects.

DC and DCTerms[14] have been used as guides to define the categories, described first in natural language. Also, the alignment has been compiled following four semantic categories (Who? When? Where? What?), in order to better define each metadata value's function. This also proved useful for the following steps' development.

THEORETICAL MODEL

Then, we enriched the previous map elaborating a further level of interpretation that was obtained by the gathering of new information about the chosen items from different resources and considering their potential connections. The new connections tried to answer four specific questions (Who? When? Where? What?) to add new data about the entity and develop different relationships between attributes. Additional information was collected from Wikipedia, Wikidata, MYmovies, GoodReads and other sources.

ENHANCED E/R MODEL

The relationships expressed in natural language in the theoretical model can be furtherly abstracted using once more the entity/relationship approach to create an enhanced version of the previous E/R model, following the same rules in regards to cardinalities and relationships' grades.

---

[13] https://www.w3.org/TR/mediaont-10/#dc-table .
[14] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/ .

CONCEPTUAL MODEL

In conclusion, we moved from the theoretical model to the formal representation of our domain. We represented entities and relationships using classes and properties from the chosen ontologies.

We generally used Dublin Core as it provides interoperable online metadata standards for a broad range of purposes. Wherever DC was insufficient, we used the ontologies that best described our items starting from the ones that we had already chosen in the “Metadata alignment” step and adding the following ones:

- **The Music Ontology[15].**

  It deals with formal representation of concepts related to music. It is built on RDF and it is also based on: **FRBR**, including the WEMI model, that underlines the hierarchy between work, expression, manifestation and item; **The Event Ontology**; **The Timeline Ontology**; **FOAF**.

- **Ontology for Media Resource[16].**

  It has been produced by the Media Annotations Working Group, as part of the W3C Video on the Web Activity, following the procedures set out for the W3C Process. The purpose of the mappings is to provide an interoperable set of metadata, thereby enabling different applications to share and reuse these metadata. The set of properties of the Ontology for Media Resources was selected with respect to the most commonly adopted set of elements from metadata formats currently in use to describe media resources.

- **Videogame Ontology.**

  Aimed at modeling video game related information, the main goal is to capture knowledge about events that happen in video games and information about players. The ontology extends concepts from **foaf**, **DBpedia**, **Dublin Core Terms, schema.org and Good Relations** vocabularies to define the videogame domain specific extension.

- **Performed Music Ontology[17].**

  Developed as an extension for **BIBFRAME**, initiated by the Library of Congress for the development of a new data model for bibliographic records,

[15] http://musicontology.com/ .
[16] https://www.w3.org/TR/mediaont-10/ .
[17] http://performedmusicontology.org/ .

it serves as an ontology for the description of classes and properties related to the performed music context.

- **GeoNames Ontology[18].**

  Included in the W3C recommendations, the GeoNames Ontology makes it possible to add geospatial semantic information to the web. The Features in the *GeoNames Semantic Web* are interlinked with each other: “children”, “neighbors” and “nearby” hierarchical relations are provided.

- **Location Ontology[19].**

  Developed by the **ArCo Network**, it provides models information related to the localization and georeferencing of a cultural property.

The graphical representation of the Conceptual model was created using Graffoo[20].

[18] https://www.geonames.org/ontology/documentation.html .
[19] http://dati.beniculturali.it/lode/extract?url=https://raw.githubusercontent.com/ICCD-MiBACT/ArCo/master/ArCo-release/ontologie/location/location.owl .
[20] https://essepuntato.it/graffoo/ .

3 - KNOWLEDGE REPRESENTATION

ITEMS DESCRIPTION

In the second part of our project, we focused on Knowledge Representation, starting from the items' description using CSV files. First of all, this task was fundamental to start thinking about the formal representation in triples of each item before the actual formalization into RDF. A triple is constituted by the basic elements of a sentence: subject, predicate and object. In the table, the subject and the object are still specified using natural language while the predicate is expressed by the ontology we had already chosen in the Conceptual model.

RDF PRODUCTION

The main goal of our project was to create Linked Open Data, which is a way of linking data with each other using as core elements RDF[21] and URIs. RDF stands for Resource Description Framework and it is defined as a standard model for data interchange on the Web. The scope of RDF is **semantic interoperability** between applications that share the same information on the Web. As a consequence, new information is produced by modeling relationships between URIs (Uniform Resource Identifier).

As for our project, the first thing that we did was to define URIs for each concept that we wanted then to link with others. In the example, we considered the Rooftop Concert as a starting point because it was one the most representative objects of our domain.
RDF can be formally represented through different kinds of serialization. We decided to choose Turtle because, according to us, it was the easiest to understand, visualize and produce. Here we briefly summarize how a Turtle serialization is created.
First, all the namespaces of the ontologies, vocabularies or authority files used in our statements have been specified at the beginning using prefixes. Next, we created the triples: the subject, expressed with an initial URI, is followed by the different predicate-object pairs. The attribute can be both a value (string) or another URI (individual, kind of things) that can be furtherly linked with other objects dealing with our domain. Of course, the production of LOD is a consequence of the point of view from which we analyzed the topic and the boundaries that we set for our network.
Other than using the already specified connections, we enriched the Turtle with other

[21] https://www.w3.org/RDF/ .

kinds of relationships defined by using OWL[22] or SKOS[23]. OWL and SKOS have been published by W3C in order to provide a standard way to represent knowledge organization systems using the Resource Description Framework. Encoding this information in RDF allows it to be passed between computer applications in an interoperable way.

OWL was used mainly to refer to the `owl:sameAs` relationship in order to set the term defined by the URI equal to the controlled term of the Authority Files (i.e. ULAN, VIAF). In combination with OWL, SKOS was used to capture much of the similarities in many Knowledge Organization Systems, such as thesauri, taxonomies, classification schemas and subject heading systems. We mainly used hierarchical links. A hierarchical link between two concepts indicates that one is in some way more general ("broader") than the other ("narrower"). The properties `skos:broader` and `skos:narrower` are used to assert a direct hierarchical link between two SKOS concepts. A triple `<A> skos:broader <B>` asserts that `<B>`, the object of the triple, is a broader concept than `<A>`, the subject of the triple. Similarly, a triple `<C> skos:narrower <D>` asserts that `<D>`, the object of the triple, is a narrower concept than `<C>`, the subject of the triple (i.e. the attribute “The Beatles” can be related to the broader concept of a rock band).

RDF VISUALIZATION

After having created the Turtle file, we used “RDF Grapher”[24] to translate the final result visually.

[22] https://www.w3.org/OWL/ .
[23] https://www.w3.org/2004/02/skos/specs .
[24] https://www.ldf.fi/service/rdf-grapher .