AtoZ
novas práticas em informação e conhecimento

ISSN: 2237-826X
PAPER | ARTIGO | ARTÍCULO

Harmonizing data: discovering "The Girl from Ipanema"

Harmonizando dados: descobrindo "Garota de Ipanema"

John Robert Rutherford [1], Liliana Giusti Serra [2]

[1] *University of Illinois at Urbana-Champaign (UIUC), Urbana-Champaign, IL, Estados Unidos. ORCID: https://orcid.org/0009-0004-9795-6216*

[2] *Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), São Paulo, SP, Brasil. ORCID: https://orcid.org/0000-0001-6788-2376*

Mail to/*Autor para correspondência/Correo a:* John Robert Rutherford, johnrr3@illinois.edu

Submitted/*Recibido*: 21 de abril de 2025; **Approved/*Aceptado*:** 03 de julho de 2025

Copyright © 2025 Rutherford & Serra. All the journal content (including instructions, editorial policies and templates) - except where otherwise indicated - is under a Creative Commons Attribution 4.0 International, since October 2020. When published by this journal, articles are free to share, adapt and you must give appropriate credit, provide a link to the license, and indicate if changes were made. To further information check http://revistas.ufpr.br/atoz/about/submissions#copyrightNotice.

Abstract

Introduction: This paper explores the potential of Linked Data to enhance library catalog records, using the internationally recognized song "The Girl from Ipanema". **Method:** The famous song was used as a case study to illustrate the possibilities of the usage of a knowledge graph and metadata (authority and bibliographic data) enrichment. **Results:** By applying Linked Data principles, the study demonstrates how metadata enrichment can improve access and discoverability in library systems. This is achieved by including International Resource Identifiers (IRIs), the utilization of external sources such as the Virtual International Authority File (VIAF), and the Last.fm application programming interface (API). In addition, a knowledge graph was introduced as a tool for providing structured relationships with bibliographic entities within a catalog. **Conclusions:** The Last.fm open API is used to enrich metadata, offer track listings, and add album art to search results, thus enhancing user experience and information retrieval. The case study includes a knowledge graph in the online library catalog.

Keywords: Knowledge graph; Linked data; Library catalog; Metadata enrichment; RDA.

Resumo

Introdução: *Este artigo explora o potencial do Linked Data para aumentar os registros do catálogo da biblioteca, utilizando a música internacionalmente conhecida "Garota de Ipanema".* ***Método:*** *a famosa música foi utilizada como um estudo de caso para ilustrar as possibilidades de uso de knowledge graph e enriquecimento de metadados (autoridade e bibliográfico).* ***Resultados:*** *ao aplicar os princípios de Linked Data, o estudo demonstrou como o enriquecimento de metadados pode promover o acesso e descoberta em sistemas de bibliotecas. Isto é alcançado ao incluir dados de Identificadores Internacionais de Recursos (IRIs), na utilização de fontes externas como o Virtual International Authority File (VIAF) e a Application Programming Interface (API) do Last.fm. Somado a isso, um knowledge graph foi incluindo no catálogo da biblioteca, proporcionando a estrutura de relacionamento entre entidades bibliográficas.* ***Conclusão:*** *A API aberta do Last.fm utilizada permitiu o enriquecimento de dados, com inclusão de listas de músicas, dados de álbum e capa no resultado de busca, aumentando a experiência do usuário e a recuperação de informações. O estudo de caso incluiu a estrutura do knowledge graph ao catálogo online da biblioteca.*

Palavras-chave: *Knowledge graph; Linked data; Catálogo - Biblioteca; Enriquecimento de metadados; RDA.*

INTRODUCTION

In 1962, the musician Antonio Carlos Jobim and the poet Vinicius de Moraes were at a bar in Ipanema, a beach in Rio de Janeiro, Brazil. They saw a girl, Helô Pinheiro, walking to the beach. This ordinary episode inspired the composition of one of the most well-known songs in the world: *The Girl from Ipanema*. This song has over five hundred versions with various performers in different languages and styles. Written in Portuguese, it was translated to English by Norman Gimbel and recorded by Frank Sinatra, Ella Fitzgerald, Madonna, Amy Winehouse, and many others.

The album Getz/Gilberto was released in 1964 with Stan Getz on the saxophone, João Gilberto on guitar, Tom Jobim on the piano, and Astrud Gilberto singing some songs, including *The Girl from Ipanema*. This album popularized the genre Bossa Nova worldwide and won the Grammy for Record of the Year in 1965.

This paper explores the potential of enhancing sound recording records in a library catalog by applying Linked Data principles. It focuses on incorporating IRIs (International Resource Identifiers) and URIs (Uniform Resource Identifiers) for authoritative data and utilizing an Application Programming Interface (API) to expand bibliographic descriptions. It seeks to observe the possibilities of enriching bibliographic data from external sources, increasing the potential for record discovery. To this end, it uses mixed techniques for data enrichment, illustrated in a single record existing in a library catalog. The text highlights the use of linked data to enrich bibliographic catalog data. Although we mainly focus on sound recordings, the methods and possibilities discussed can also be applied to other types of resources. The selected song serves as an illustrative example of these possibilities.

We chose the song *The Girl from Ipanema* to illustrate this study because it is a good example of exploring Resource, Description, and Access (RDA). Due to its international popularity, this Work has several Expressions, Manifestations, and Items. We selected an English sound recording (Expression) of "*The Girl from Ipanema*" translated from the original Work "Garota de Ipanema" to analyze the possibilities of applying Linked Data features, providing data enrichment and discovery expansion.

Sound recordings are usually described in academic library catalogs at the album level, with basic data, and they do not always contain details of their songs or performers. An album combines a series of songs, each being a Work or an Expression of a Work. Due to the lack of information about the songs in the catalog, it is impossible to relate them to other records in the catalog, such as songs by a specific performer or with the same or similar genres. This situation compromises song discoverability.

Other aspects that hinder discoverability include:

- Locating a specific song version (Expression) by a specific performer (Agent)
- Finding the relationship between the song (Work or Expression) and the performer (Agent)
- Exploring information about the performers (Agents) or discovering other related songs (Works or Expressions)
- Locating other versions (Expressions) of a song in other albums (Works)

This paper explores new ways to overcome two key problems: First, authority data usually lacks details of Agents (Person and Collective agent entities), which could help users identify other data sources or provide connections between related entities, such as Works, Expressions, Manifestations, other Agents, etc. Second, poor metadata quality makes it difficult for users to discover and access a complete and nuanced context of a Work. Enriching metadata and enhancing connections to related Works, Expressions, and Manifestations will ultimately provide users with a more comprehensive and meaningful experience.

This study is original, since no similar studies were found that propose data enrichment using mixed techniques with different sources. Besides enrichment, the paper also suggests the generation of a knowledge graph departing from bibliographic data.

METHOD

We aim to assess the current state of Linked Data in libraries and explore how we could advance its integration into catalog systems. This research offers a high-level perspective on improving access and discovery in catalog systems, using a case study to exemplify some options proposed. Our case study approach provides an in-depth exploration of the existing MARC records and their limitations. We selected this method to analyze the specific challenges and opportunities associated with improving the metadata for a single, well-known Work.

We considered both authority and bibliographic data. To enrich authority data regarding agents (composers, performers, etc.), we selected the Virtual International Authority File (VIAF) for birth and death dates because it is a reliable source, as it contains data provided by national libraries and other institutions. These enriched data will be stored in subfield $1 for fields 100 (main entry), 700 (added entry), and 600 (personal name as a subject) in the bibliographic MARC records, thus describing agents and contextualizing real-world information. The authority data will be stored in subfields 024 $1 (identifiers) and 100 $a (personal name), following identifier descriptions shown in Serra, Schneider, and Santarém Segundo (2020).

For bibliographic data enrichment, we use the open data source Last.fm API and demonstrate the potential of Linked Data in enhancing information discovery and access. This newfound information is used to recommend similar tracks in the selected example. Lastly, we create a graph connecting agents related to the songs in the album. The graph is created from the links already formed from the catalog records.

LITERATURE REVIEW

Resource Description and Access (RDA) is a cataloging standard released in 2010 that provides guidelines and instructions for bibliographic data description. For alignment with the International Federation of Library Associations and Institutions Library Reference Model (LRM), a conceptual model released in 2017, RDA went through significant changes due to the RDA Toolkit Restructure and Redesign (RDA 3R Project) ending in 2020. RDA structures the bibliographic records using the entities Work, Expression, Manifestation, and Item (WEMI), where a Work is realized through an Expression, embodied in a Manifestation, and exemplified by an Item. Work and Expression are conceptual elements; Manifestation is the record effectively cataloged, and Item is the copy that is part of the collection in a library.

Some library resources have Aggregates, which are Manifestations that embody multiple Expressions of Works. In other words, aggregates are:

> [...] sets of multiple independently created expressions which are 'published' together in a single manifestation. [...] Examples include journals (aggregates of articles), multiple novels published together in a single volume, books with independently written chapters, CDs (aggregates of individual songs), and various collected/ selected works (International Federation of Library Associations and Institutions (IFLA), 2011, p. 3).

Music albums are identified as Aggregate Collections of Expressions, and several songs (Expressions of Works) are embodied in a single Manifestation (a disc, a CD, etc.). In aggregates, the selection of Expressions can occur for varied reasons, such as similar genres or themes, songs related to a particular composer, artist, or performer, or also being a random collection of Expressions.

Libraries build their descriptions using the MARC 21 format, a complex and specialized format used only by libraries, which makes interoperability with the Web challenging. Although it is known that MARC cannot allow a semantic description and will eventually be replaced, libraries mostly keep using it or are considering data conversion to other structures. The fact is that many libraries have a data legacy in MARC format, and any changes or transitions to another format will depart from MARC. The format fills fields and subfields with textual descriptions or terms from controlled vocabularies. Authors and subjects are controlled terms (authority data), usually in tables, while notes, titles, and other fields are based on text placed in fields and subfields (bibliographic data).

We know that the MARC format does not have a structure that is fully prepared to encompass linked data. However, we chose to use the format in this research because it is used by most libraries. Other frameworks are under development, such as BIBFRAME or RDF, but we chose to conduct this analysis in MARC because it is the current reality of libraries. Suppose it is possible to create this relationship structure with MARC, which has so many restrictions for linked data. In that case, the application with structures prepared for this type of data should not pose any difficulties.

One prominent problem with MARC is that the process of full-text searching does not compute synonyms or homonyms, variant spellings, other languages or dialects, use of acronyms, etc. (Beall, 2008). Another challenge is that full-text search provides bad retrieval, due to a lack of precision, typically returning several records that are not exactly what the user wants. Although MARC 21 allows the inclusion of IRIs and URIs in some tags, its structure is not aligned with the incorporation of Linked Data, and not all fields can utilize this feature. Some IRI/URI inclusions are in authority data and might reflect on bibliographic data, such as personal names in tag 100 for authority and tag 100 for bibliographic data. However, including the IRI/URI is not enough. If the library system is not ready to convert a URI/IRI into an active link, the linked data features will not be explored in the online catalog.

CASE STUDY

Initially, we selected a physical CD (Item) of the album Getz/Gilberto, released in 2014 by the Verve Music Group (OCLC (OCoLC)877860553). Several parts of the data were sparse, thus rendering the discovery process weak and incomplete. We analyzed the MARC bibliographic data and realized it was possible to enrich data from fields 100 and 700. We noticed further enrichment could be done by adding the agents' birth and death dates. We included URIs manually because it was a single record with just a few agents involved, but it is possible to apply metadata reconciliation tools (Heng, Cole, Tian, & Han, 2022; Carlson & Seely, 2017; Parker & Gray, 2019), including URIs from VIAF, Wikidata, or other datasets.

Figure 1 presents a fragment of a bibliographic record from the album Getz/Gilberto in MARC 21, without control (00X), number, and code (0XX) tags, focusing on the basic data for description, such as main entry (100), title (245), publishing data (260), physical description (300), formatted content note (505), and added entries (700). For this sample, the 650 fields were removed because the subject/genre was not the focus of this study.

100 1# $a Getz, Stan, **$d** 1927-1991, **$e** performer.
245 10 $a Getz/Gilberto / **$c** Stan Getz and João Gilberto; featuring Antonio Carlos Jobim.
260 ## $a Santa Monica, CA : **$b** Verve, **$c** p2014.
300 ## $a 1 audio disc : **$b** digital, stereo/mono ; **$c** 4 3/4 in.
505 00 $t The girl from Ipanema -- **$t** Doralice -- **$t** P'ra machucar meu coração -
- **$t** Desafinado -- **$t** Corcovado -- **$t** Só dançco samba -- **$t** O grande amor -
- **$t** Vivo sonhando -- **$g** Bonus tracks. **$t** The girl from Ipanema **$g** (single version) -
- **$t** Corcovado **$g** (single version).
700 1# $a Gilberto, João, **$e** performer.
700 1# $a Jobim, Antonio Carlos, **$e** performer.
700 1# $a Neto, Sebastião **$c** (Bassist), **$e** performer.
700 1# $a Banana, Milton, **$e** performer.
700 1# $a Gilberto, Astrud, **$e** performer.

Figure 1. Fragment of the record from *The Girl from Ipanema* (UIUC, 2024)
Source: Taken from the University of Illinois at Urbana-Champaign OPAC

Despite the song's prominence, current library cataloging practices face challenges in ensuring the discoverability and accessibility of related information, depending on a textual description inside a note, which can have several pieces of information. In library catalogs, the song descriptions inside an album are usually informed by a formatted contents note, the field 505 on MARC 21. Below, the field 505 is highlighted (Figure 2).

505 00 $t The girl from Ipanema -- **$t** Doralice -- **$t** P'ra machucar meu coração -- **$t** Desafinado -- **$t** Corcovado -- **$t** Só dançco samba -- **$t** O grande amor -- **$t** Vivo sonhando -- **$g** Bonus tracks. **$t** The girl from Ipanema **$g** (single version) -- **$t** Corcovado **$g** (single version).

Figure 2. Fragment of a MARC record describing the song "*The Girl from Ipanema*"

In this record, the songs that are part of the album have descriptions on subfields "$t" (title) and "$g" (miscellaneous information), but they lack details about the performers, genres, or other related data. The subfield $r could be used to collect the statement of responsibility, indicating the authorship of each song, but this example does not have this subfield. The lack of clarity illustrates the fragility of the MARC format, which does not provide clear definitions for its descriptive elements. In this field, it is impossible to distinguish the relationship between the subfields $t with $g or $r, or even to which $t the data of $g or $r are related. This situation makes machine interpretation difficult since it is impossible to identify the relations between the subfields because it depends on the filling sequence used during the cataloging process and on patron interpretation.

In note 505, the subfields are usually separated by "—" (space-dash-dash), which may enhance human readability but hinder computer or machine processing, making data recognition and extraction challenging. According to J. T. Clark (2024), this punctuation is familiar to librarians and is used in examples in AACR2 and RDA and for the Library of Congress (LC) Program for Cooperative Cataloging (PCC) interpretation of RDA. Also, J. T. Clark (2024) notices that the double hyphen is presumably used because the dash is not a standard character of the American Standard Code for Information Interchange (ASCII). Figure 3 illustrates the presence of punctuation separating data from a content note.

Contents	The girl from Ipanema -- Doralice -- P'ra machucar meu coração -- Desafinado -- Corcovado -- Só dançco samba -- O grande amor -- Vivo sonhando -- Bonus tracks. The girl from Ipanema (single version) -- Corcovado (single version).

Figure 3. User interface exhibiting the contents note of the album Getz/Gilberto

Kirkland and Wald (2018) reported research showing that content notes increased the circulation of library books, thus proving that the data in notes was almost as useful as the titles themselves. Looking at the music scenario, content notes provide information about musical Works, which are identified mainly by their titles. However, data retrieval from text notes is inefficient and imprecise because it relies on matching words in a search query with words in a record.

Recognizing the need for more complete data, we connected the data to external sources, enriching authority and bibliographic data. URIs from the VIAF were added to authority data to identify real-world entities. VIAF

provides basic data related to an author, such as date of birth and date of death, 'see from' and 'see also' tracing references, related Works, etc. Table 1 lists the agents present in the record enriched with dates and identifiers from VIAF.

Enriched authority data using VIAF
024 7# $a http://viaf.org/viaf/113597520 $2 Viaf 100 1# $a Getz, Stan, $d 1927-1991
024 7# $a http://viaf.org/viaf/85530975 $2 Viaf 100 1# $a Gilberto, João, $d 1931-2019
024 7# $a http://viaf.org/viaf/51875584 $2 Viaf 100 1# $a Jobim, Antonio Carlos, $d 1927-1994
024 7# $a http://viaf.org/viaf/5133102 $2 Viaf 100 1# $a Tião Neto, $d 1931-2001
024 7# $a http://viaf.org/viaf/35673323 $2 Viaf 100 1# $a Banana, Milton, $d 1935-1999
024 7# $a http://viaf.org/viaf/44485145 $2 Viaf 100 1# $a Gilberto, Astrud, $d 1940-2023

Table 1. Enriched authority data using VIAF

While enriching the case study's data, we realized that the authority 'Neto, Sebastião (Bassist)' was inappropriate according to the rules related to Portuguese names, even though it exists in this format on the Library of Congress Authority Data records. For this study, we relied on the form available in the Brazilian National Library in the VIAF and not the Library of Congress (LC) authority description because of the author's nationality.

These authority additions result in richer bibliography data, allowing for appropriate record linking. Table 2 compares the original and the enriched bibliographic data after these additions.

Original data	Enriched data with VIAF
100 1# $a Getz, Stan, $d 1927-1991	100 1# $a Getz, Stan, $d 1927-1991 $1 `http://viaf.org/viaf/113597520`
700 1# $a Gilberto, João	700 1# $a Gilberto, João, $d 1931-2019 $1 `http://viaf.org/viaf/85350975`
700 1# $a Jobim, Antonio Carlos	700 1# $a Jobim, Antonio Carlos, $d 1927-1994 $1 `http://viaf.org/viaf/15758584`
700 1# $a Neto, Sebastiao Sc (Bassist)	700 1# $a Tiao Neto, $d 1931-2001 $1 `http://viaf.org/viaf/5133102`
700 1# $a Banana, Milton	700 1# $a Banana, Milton, $d 1935-1999 $1 `http://viaf.org/viaf/35763323`
700 1# $a Gilberto, Astrud	700 1# $a Gilberto, Astrud, $d 1940-2023 $1 `http://viaf.org/viaf/44485145`

Table 2. Original and enriched bibliographic data

A different approach was applied to bibliographic data to increase the discoverability of songs based on data from agents (composers, performers, musicians, etc.) and related songs. This approach can facilitate the discovery of songs in an album or from a specific author, and similar songs or genres.

Last.fm[1] is a music service that tracks people's music listening habits, helping them discover and share music through recommendations, exploration, and integration with social networking tools. They offer an open API[2] that can be used to retrieve information about music in general. The API offers data on artists, album art, albums and song names, genres, and descriptions.

We created a simple Python script that interfaces with the Last.fm API to fetch and display information about artists, albums, and tracks. The API makes use of the requests library to send HTTP GET requests to the Last.fm API, retrieving data in JSON format. The script consists of several functions that handle different types of queries:

a) **search_artist**: Esta função constrói uma URL para consultar a API do Last.fm usando o nome do artista fornecido. Se a requisição for bem-sucedida (código HTTP 200), ela obtém as informações do artista e imprime os nomes de até cinco artistas similares.

b) **search_album**: Semelhante a *search_ artist*, esta função consulta a API do Last.fm para obter informações sobre um álbum, utilizando o nome do artista e do álbum fornecidos.

[1] https://www.Last.fm/
[2] https://www.Last.fm/api/

c) **search_track**: Esta função realiza duas consultas: uma para obter informações sobre uma faixa específica e outra para buscar faixas similares. Se ambas as requisições forem bem-sucedidas, a função formata ("prettifies") os dados para torná-los legíveis e imprime as informações da faixa juntamente com uma lista de faixas similares e seus respectivos artistas.

The script's main function prompts the user to enter the type of search they want to perform (artist, album, or track) and then gathers the necessary inputs (artist name, album name, track name) to call the appropriate function. The results are printed in a user-friendly format, which can be incorporated into a front-end experience.

The Last.fm API offers a vast array of methods to call data from the API. These allow for varied enhancement of user discovery interfaces. For this project, we specifically selected information that aligned with our conceptual design of the discovery section and user interface. The focus was on enriching the database with enhanced metadata and offering suggested tracks to users.

In the "search artist" API call, we requested similar artists but only received a single album by the same artist; therefore, we disregarded that data and focused on tracks. While the artist information received was minimal, consisting only of "id," "name," and "URL," this was acceptable because we have other authoritative sources to supplement artist information.

The retrieved track information included more detailed data such as duration, listeners, play count, streaming status, tags, URL, and a robust wiki about the track. Although not all the track information was essential for the project, it is helpful to understand the range of data available. Even using only three API methods out of dozens, we were able to retrieve a substantial amount of additional information, which would help users.

Since the Last.fm API provides the ability to retrieve similar tracks, we request five recommendations from the API based on the main search. Figure 4 shows part of the script to retrieve similar tracks.

The code is freely available here: https://github.com/infoqualitylab/lastfm_api_interface

```
def search_track(artist_name, track_name):
    url =
f"http://ws.audioscrobbler.com//2.0/?method=track.getinfo&api_key={API_KEY}&artist={a
rtist_name}&track={track_name}&format=json"
    similar_url =
f"http://ws.audioscrobbler.com//2.0/?method=track.getsimilar&artist={artist_name}&track={
track_name}&api_key={API_KEY}&format=json"
    response = requests.get(url)
    similar_response = requests.get(similar_url)

    if response.status_code == 200 and similar_response.status_code == 200:
        data = response.json()
        similar_data = similar_response.json()
        pretty_data = prettify_data(data)

        # Get similar tracks
        if 'similartracks' in similar_data:
            sim_track_data_list = similar_data['similartracks']['track']
            print("\nSimilar Tracks:")
            for i, track_data in enumerate(sim_track_data_list[:5], 1):  # Top 5 similar tracks
                print(f"{i}: {track_data['name']} by {track_data['artist']['name']}")

        return pretty_data
```

Figure 4. Section of code drawing in 5 recommended, similar tracks, and additional information about those tracks

Applying this script to the case study, we get as a response some data, such as artist and track name, similar tracks (Figure 5).

```
/Users/johnrr3/PycharmProjects/LastFMData/lastfmdata/bin/python /Users/johnrr3/PycharmPr
Enter the type of search (artist, album, or track): track
Enter the artist name: Stan Getz
Enter the track name: The Girl From Ipanema

Similar Tracks:

1: Corcovado (Quiet Nights Of Quiet Stars) by Stan Getz
2: Para Machuchar Meu Coracao by Stan Getz
3: Desafinado by Stan Getz & João Gilberto
4: Chega de Saudade by João Gilberto
5: Wave by Antônio Carlos Jobim
{
    "track": {
        "album": {
            "@attr": {
                "position": "14"
            },
            "artist": "Stan Getz",
            "image": [
                {
                    "#text": "https://lastfm.freetls.fastly.net/i/u/34s/21d2a064f60b4865
                    "size": "small"
                },
                {
                    "#text": "https://lastfm.freetls.fastly.net/i/u/64s/21d2a064f60b4865
                    "size": "medium"
                },
                {
                    "#text": "https://lastfm.freetls.fastly.net/i/u/174s/21d2a064f60b486
                    "size": "large"
                },
                {
                    "#text": "https://lastfm.freetls.fastly.net/i/u/300x300/21d2a064f60b
                    "size": "extralarge"
                }
            ],
```

Figure 5. Response from the script

Starting with our current campus search results page, we superimposed a discovery section into our search results section to ingest the data from the API. To access the Last.fm API, a Python script, connects and returns the API results for this discovery section. By leveraging the data from the API, it was possible to incorporate the album artwork, "suggested tracks" section, and fill in other missing metadata. Figure 6 presents the data retrieved in this discovery section.

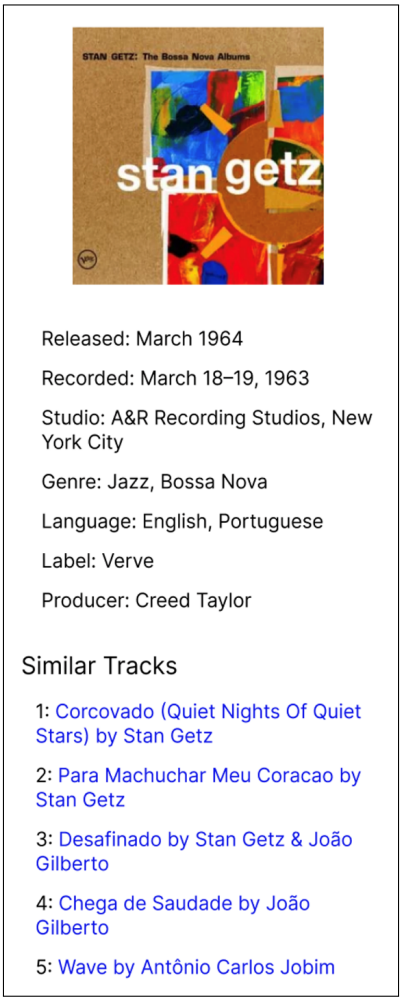

Figure 6. Artwork and album information

Knowledge graphs are structured representations of information, often used to model relationships between entities in a way that is both human-readable and machine-interpretable. They integrate data from various sources, providing a semantic context that enhances the ability to search, analyze, and visualize complex information. They have proven to be effective tools for enhancing User Experience design by organizing and visualizing structured information and semantic queries (J. A. Clark & Young, 2017).

To illustrate the relationship between agents and offer other possibilities for discovery, we propose knowledge graphs for this example, providing user navigability (Figure 7).

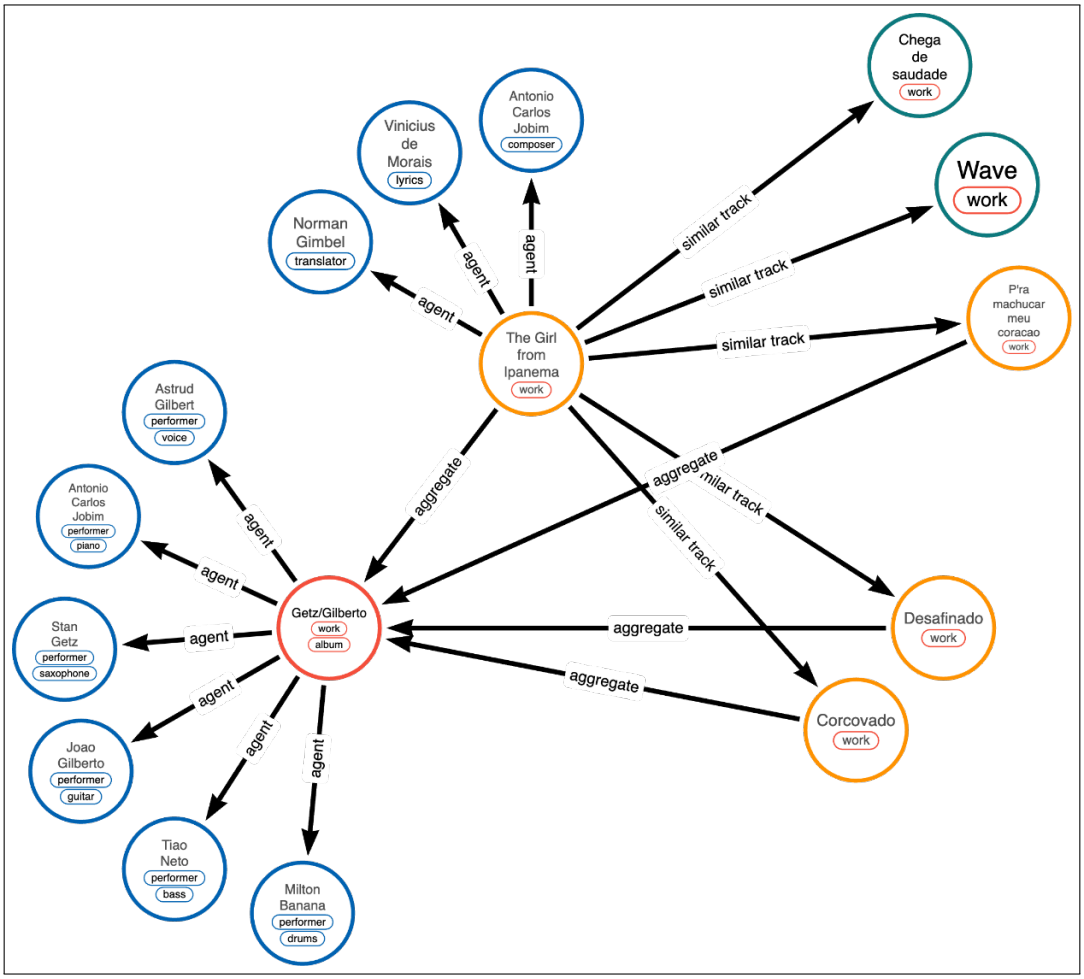

Figure 7. Knowledge graphs presenting relationships between entities

By analyzing the graph, it is possible to identify that some similar tracks suggested by the Last.fm API are also aggregated in the album Getz/Gilberto. The graph showcases the various agents related to the album and the song *The Girl from Ipanema*. For each Work, its Agents. The API can be expanded to include more attributes for each song, such as duration, date of creation, genre, etc.

The information extracted by the API and the knowledge graphs can be added to the OPAC in a box with the search results, providing users with other possibilities for discovery. Figure 8 suggests these features are embedded in the OPAC.

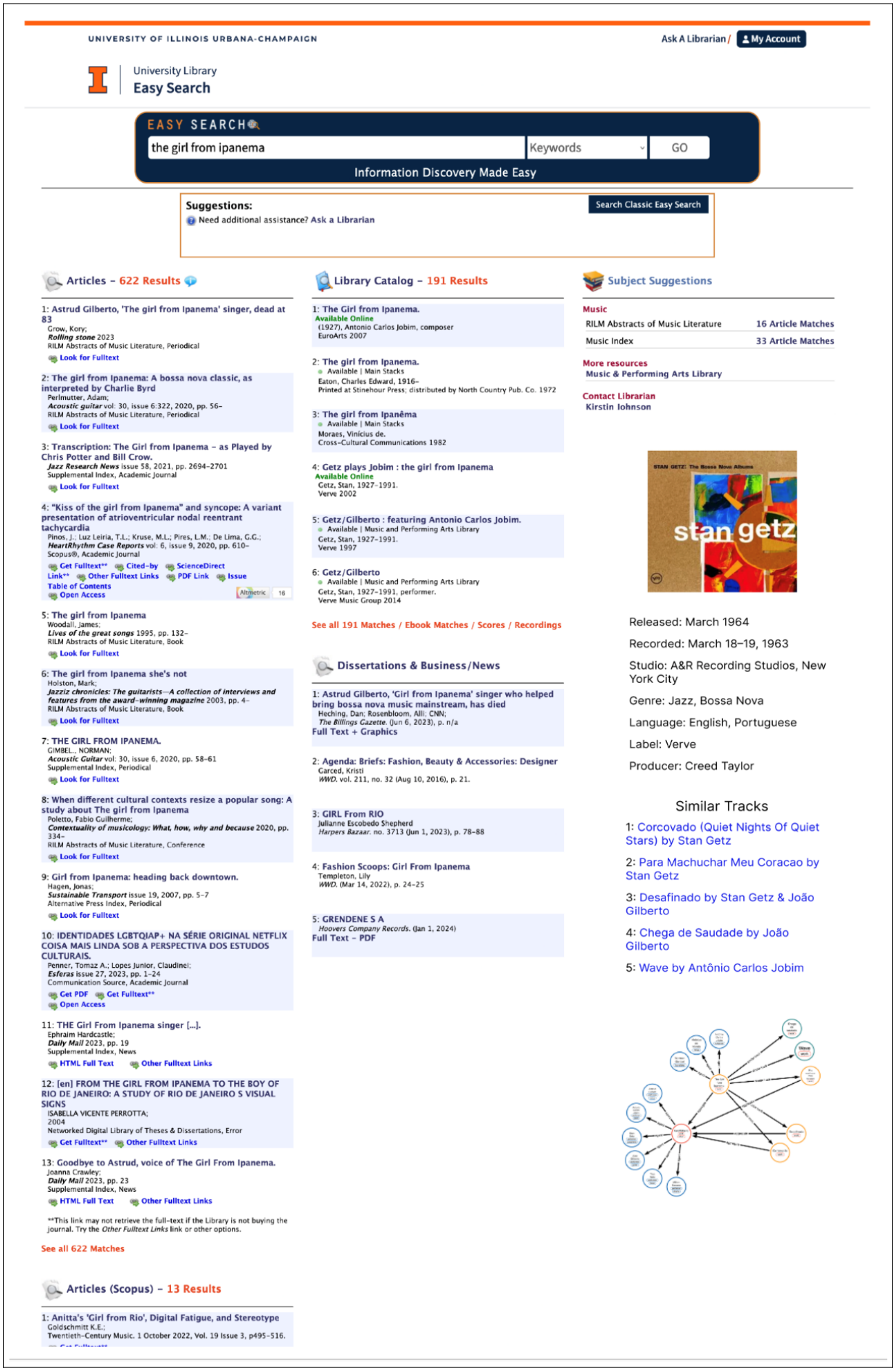

Figure 8. Data from the API, graph in the OPAC
Source: Added data from the API and knowledge graph at the University of Illinois at Urbana-Champaign OPAC

CONCLUSION

This paper emphasizes the significant potential of Linked Data to improve the discoverability and accessibility of catalog records, using "*The Girl from Ipanema*" as an illustrative example. By connecting catalog data with reliable external sources like VIAF, and Last.fm API, we demonstrated how metadata enrichment can provide users with a deeper and more nuanced understanding of a Work and its broader context.

The integration of visual tools, such as knowledge graphs, can significantly improve user experience by providing structured, visually accessible information and recommendations. A graph constructed from catalog records, allowing filtering and basic linking, is an essential enhancement. Users now expect basic metadata and recommendations in modern contexts.

There is a transformative potential of Linked Data in library cataloging. We advocate for continued efforts in this direction, believing that these enhancements can significantly improve library services and ultimately enhance the user experience. We aim to contribute meaningfully to the ongoing dialogue within library and information science, promoting the use of Linked Data to create richer, more connected catalog records.

REFERENCES

Beall, J. (2008). The weaknesses of full-text searching. *Journal of Academic Librarianship*, *34*(5), 438–444. Retrieved from https://tefkos.comminfo.rutgers.edu/Courses/e530/Readings/Beal%202008%20full%20text%20searching.pdf

Carlson, S., & Seely, A. (2017). Using openrefine's reconciliation to validate local authority headings. *Cataloging & Classification Quarterly*, *55*(1), 1–11. doi: 10.1080/01639374.2016.1245693

Clark, J. A., & Young, S. W. H. (2017). Linked data is people: building a knowledge graph to reshape the library staff directory. *The Code4Lib Journal*(36). Retrieved from https://journal.code4lib.org/articles/12320

Clark, J. T. (2024). The format and display of the marc 505 contents note: some recent practical developments. *Cataloging & Classification Quarterly*, *62*(2), 188–208. doi: 10.1080/01639374.2024.2350467

Heng, G., Cole, T. W., Tian, T., & Han, M.-J. (2022). Rethinking authority reconciliation process. *Cataloging & Classification Quarterly*, *60*(1), 45–68. doi: 10.1080/01639374.2021.1992554

International Federation of Library Associations and Institutions (IFLA). (2011). *Final report of the working group on aggregates.* IFLA. Retrieved from https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbrrg/AggregatesFinalReport.pdf

Kirkland, L. N., & Wald, J. (2018). Assessing the benefits of content notes enhancements for digital audio resources. *Cataloging & Classification Quarterly*, *56*(4), 287–297. doi: 10.1080/01639374.2017.1398196

Parker, B., & Gray, A. (2019). Rethinking the university of maryland authority file for the linked data environment. *Journal of Library Metadata*, *19*(1-2), 69–81. doi: 10.1080/19386389.2019.1589699

Serra, L. G., Schneider, J. A., & Santarém Segundo, J. E. (2020). Person identifiers in marc 21 records in a semantic environment. *Cataloging & Classification Quarterly*, *58*(5), 505–519. doi: 10.1080/01639374.2020.1771499

University of Illinois at Urbana-Champaign. (2024). *Library catalog.* Retrieved from https://i-share-uiu.primo.exlibrisgroup.com/permalink/01CARLI_UIU/gpjosq/alma99753766612205899

How to cite this article (APA):
Rutherford, J. R. & Serra, L. G. (2025). Harmonizing data: discovering "The Girl from Ipanema". *AtoZ: novas práticas em informação e conhecimento, 14*, 1 – 13. Retrieved from: http://dx.doi.org/10.5380/atoz.v14.99319

NOTAS DA OBRA E CONFORMIDADE COM A CIÊNCIA ABERTA

CONTRIBUIÇÃO DE AUTORIA

Papéis e contribuições	John Robert Rutherford	Liliana Giusti Serra
Concepção do manuscrito	X	X
Escrita do manuscrito	X	X
Metodologia	X	X
Curadoria dos dados	X	X
Discussão dos resultados	X	X
Análise dos dados	X	X

EQUIPE EDITORIAL

Editora/Editor Chefe

Paula Carina de Araújo (https://orcid.org/0000-0003-4608-752X)

Editora/Editor Associada/Associado Júnior

Karolayne Costa Rodrigues de Lima (https://orcid.org/0000-0002-6311-8482)

Editora/Editor de Texto Responsável

Fabiane Führ (https://orcid.org/0000-0002-3723-050X)

Seção de Apoio às Publicações Científicas Periódicas - Sistema de Bibliotecas (SiBi) da Universidade Federal do Paraná - UFPR

Editora/Editor de Layout

Felipe Lopes Roberto (https://orcid.org/0000-0001-5640-1573)