

RECURSOS SONOROS E LINKED DATA: ABORDAGEM SOBRE METADADOS

GT 4 – Gestão e organização da informação e do conhecimento

Modalidade da apresentação: comunicação oral

SANTOS, Amanda Azevedo dos¹
SIMIONATO, Ana Carolina²

Resumo: O aperfeiçoamento de Tecnologias da Informação e Comunicação (TIC) e sua aceitação por parte da sociedade, em virtude de suas facilidades de uso, colabora com o crescimento do fluxo e volume informacional na *Web*. Crescimento que ocorre de forma acelerada e despreocupada com a organização, representação e recuperação de recursos informacionais, sejam eles fotos, vídeos ou músicas, o que traz dificuldades para a recuperação. Deste modo, a Ciência da Informação busca compreender estes fluxos, com o objetivo de aperfeiçoar a forma de tratamento e representação desses recursos, de forma que propicie eficiência na recuperação e no acesso. No contexto de recursos sonoros podem ser utilizados metadados, que identificam e propiciam a organização e representação desse recurso. Visto isso, este trabalho possui como objetivo investigar o *dataset* *BBC Music* disponível no projeto *Linking Open Data*, para extrair formas que contribuem com a representação e organização de recursos sonoros em catálogos. A pesquisa caracteriza-se como exploratória e bibliográfica. Os resultados apresentaram como as formas de representação utilizadas pelo *BBC Music* podem colaborar com catálogos, assim como novas perspectivas para Ciência da Informação.

Palavras-chave: Linked data. Linking Open Data. Recursos sonoros.

SOUND RESOURCES AND LINKED DATA: METADATA APPROACH

Abstract: The improvement of Information and Communication Technologies (ICT) and their acceptance by society, in reason its facilities, collaborat with the growth of the flow and volume of information on the *Web*. Growth that occurs in an accelerated and disinterested way with an organization, representation and retrieval of information resources, be they photos, videos or music, that bring the difficult to a recovery. In this way, Information Science seeks to understand, with the objective of improving the form of treatment and the forms of resources, so that exist efficient in the recovery. In the context of resources, indicators can be used as metadata, identifiable and conducive to an organization, and an increase in resources. This project aims to investigate the *BBC Music* data set in the *Linking Open Data* project to extract forms that contribute to a representation and organization of sound resources in catalogs. The research characterize as exploratory and bibliographic. The results show how

¹Graduanda (Discente), Universidade Federal de São Carlos (UFSCar), amandahazevedo19@gmail.com;

²Doutora (Docente), Universidade Federal de São Carlos (UFSCar), simionato.ac@gmail.com;



BBC Music's representation forms collaborate with catalogs, as well as new perspectives for Information

Keywords: Linked data. Linking Open Data. Sound data.

1 INTRODUÇÃO

Diante da ascensão do uso e investimento no desenvolvimento das Tecnologias da Informação e Comunicação (TIC), em conjunto com o acesso à *internet*, pôde se observar um novo comportamento social, que consiste em produzir e compartilhar informações e conteúdos digitais em grande escala. O que reflete em um crescimento exponencial de diversas tipologias de recursos informacionais, gerando dificuldades para definir e contabilizar esse crescimento, além de dificuldades para a recuperação de um determinado recurso.

Referente ao volume de recursos sonoros (termo atribuído ao recurso informacional música) e conteúdos musicais (notação musical, música, informações adicionais como artistas, compositores e história), a plataforma *Web* se apresenta como um suporte favorável à sua disseminação, assim, *downloads e uploads* de músicas e álbuns podem ser realizados com facilidade e eficiência.

Frente a isso, a Ciência da Informação pode aperfeiçoar e criar interações de qualidade entre a informação e o usuário, por meio do uso de tecnologias, técnicas e padrões para a representação desses recursos, o que pode influenciar na exatidão da recuperação.

1.1 OBJETIVOS

Tendo em vista o crescimento de recursos sonoros na *Web*, e as dificuldades na recuperação desses recursos, o objetivo geral consiste em investigar o *dataset* *BBC Music* disponível no projeto *Linking Open Data*, para encontrar formas que possam colaborar com novas construções de registros para catálogos.

Os objetivos específicos são:

- Compreender a definição de metadados;
- Discorrer sobre os conceitos de *Linked Data* e *Linking Open Data*;
- Identificar os metadados do catálogo *BBC Music*, utilizando a ferramenta *Open Refine*.

1.2 PROCEDIMENTOS METODOLÓGICOS



Refere-se a uma pesquisa exploratória de caráter qualitativo de cunho teórico que aborda os temas: organização, representação, recuperação, princípios *Linked Data* e o movimento *Linked Open Data*. A partir da análise exploratória e descritiva da literatura disponível, é possível construir o conhecimento teórico acerca de novas perspectivas no que tange dados de recursos sonoros disponibilizados na *Web*.

Para a contextualização teórica foram utilizadas fontes bibliográficas, como embasamento para os resultados. O recorte da pesquisa se restringiu em pesquisas publicadas internacionalmente e no Brasil, a partir do período de 2000, visto o crescimento de publicações sobre o tema, porém alguns clássicos também são citados. As fontes de informações utilizadas foram: Portal de Periódicos da Capes, *Web of Science*, Google Scholar e *World Wide Web Consortium (W3C)*.

Além disso, para verificar e extrair os metadados disponíveis no *BBC Music* foi utilizado o *software* livre *Open Refine*. Ferramenta desenvolvida em 2010 pelo *Google* como (*Google Refine*) com o objetivo de limpar, converter, organizar e reconciliar dados.

2 DESENVOLVIMENTO

A música como arquivo no formato MP3 apresenta dados (anexados) referentes à sua produção que podem ser visualizados no computador quando o usuário acessa as propriedades de um arquivo e seleciona a aba "Detalhes". Esses dados são padrões denominados *Identify a MP3 (ID3)*, como pontua Ferreira (2015, p.13). “Para representação de músicas em formato MP3 existe o padrão ID3, que é um conjunto de metadados incorporados ao próprio arquivo de áudio.”

Grácio (2002, p. 23) define metadados como “Conjunto de elementos que descrevem as informações contidas em um recurso, com o objetivo de possibilitar sua busca e recuperação.” Elementos que devem apresentar informações descritivas, como o título, data de criação, autor ou responsável, em suma as características de um recurso ou item.

Os metadados também podem ser assimilados como descritores ou atributos, que podem enriquecer, identificar e auxiliar na interação entre dados e *softwares*, pois eles podem enriquecer dados e informações podendo preservá-los. (POLLOK, 2011).

Ciente que desde os primórdios o homem possui um senso e necessidade em criar ferramentas para que suas informações possam ser registradas e perpetuadas para que futuramente possam ser utilizadas, e com o aumento da produção e fluxo informacional a necessidade de registro e investimento em desenvolvimento de novas tecnologias de

armazenamento foi elevada de forma a interferir na memória nativa e social (OLIVEIRA; RODRIGUES, 2011).

Frente ao desenvolvimento de estudos sobre *internet*, em 1989 o pesquisador Tim Berners-Lee do *European Organization for Nuclear Research Center* (CERN) criou um ambiente que usava uma rede de comunicação, popularmente conhecida como *internet* que viabiliza a conexão e a partir dela há a *Wide World Web* ou WWW, em que pode existir o compartilhamento de arquivos, textos, áudios, imagens e vídeos. (BERNERS-LEE, 1989).

Arakaki (2016) em seu discurso apresenta as definições segundo Heath e Bizer (2011, não paginado) que afirma que *Linked Data* [...] “o termo *Linked Data* refere-se a um conjunto de melhores práticas para a publicação e interligação de dados estruturados na Web.”.

O diagrama do *Linked Open Data* é sustentado por *datasets* (sendo o principal o DBpedia), divididos de acordo com temáticas e áreas como dados governamentais, ciências da vida, *cross domain*, redes sociais, mídias, dados gerais e gerados pelos usuários que podem ser diferenciados por cores na legenda.

No *Linked Open Data* é possível localizar *datasets* relacionados ao contexto musical, e analisar o modo como esse recurso é oferecido aos usuários, assim como, quais ferramentas, metadados e padrões são utilizados para favorecer e aperfeiçoar a recuperação deste conteúdo. Dois *datasets* sobre música são o *BBC Music* que é um portal de música interativo o qual podem ser encontradas resenhas de eventos, álbuns, cobertura de lançamentos.

A British Broadcasting Corporation conhecida popularmente como BBC é uma famosa rede de rádio e televisão que com o propósito de expandir o alcance e produção dos seus produtos para os usuários, como *BBC Sport*, *BBC Education*, *BBC Music* disponibiliza o seu conteúdo no *Linked Open Data* que em decorrência requer a elaboração de novos modelos capazes de suportar as novas funcionalidades desenvolvidas pela rede, assim como englobar novos e variados domínios de interesse (IWASZKO, 2017)

Além disso, é considerada como um dos exemplos mais famosos, no que se refere a dados conectados, pois a rede BBC está na *Web* desde 1994 e continuou trabalhando para o enriquecimento das páginas *Web*. A partir disso foi se expandindo e começou a publicar dados acerca de informações musicais, adotando o vocabulário do *MusicBrainz*, outro *dataset* relacionado a música., Deste modo, a BBC é capaz de proporcionar aos seus clientes boas experiências com seus produtos. (ISOTANI; BITTENCOURT, 2015)

Deste modo, a BBC utiliza como base ontologias que viabilizam a disponibilização dos seus dados no *Linked Open Data*, afirmando que:

O *Linked Data* nos dá a oportunidade de conectar o conteúdo por meio desses tópicos. Utilizamos ontologias para descrever o mundo que nos rodeia, conteúdo criado pela BBC e gerenciamento, armazenamento e compartilhamento desses dados dentro da Plataforma *Linked Data*. (BBC, 2017, não paginado, tradução nossa).

A rede BBC acredita que o *Linked Open Data* é uma oportunidade de expandir os seus produtos para seus clientes, além de alcançar outros clientes potenciais, também há a ideia de criar suportes para futuros produtos.

Para disponibilizar conteúdos no *Linked Open Data* são utilizadas ontologias específicas para cada produto, e o *CreativeWork* (trabalho criativo) item de conteúdo online que define os metadados necessários “Nós definimos os metadados necessários, como título, breve descrição e um link para o trabalho criativo completo na Ontologia do *CreativeWork*.”. (BBC, 2017, não paginado, tradução nossa).

A BBC utiliza ontologias específicas para cada um dos seus produtos explanadas no quadro 1.

Quadro 1: Ontologias BBC

<i>CMS Ontology</i>	Representação dos sistemas de gerenciamento de conteúdo
<i>Core Concepts Ontology</i>	Para pessoas, eventos, instituições entre outros
<i>Creative Work Ontology</i>	Mínimos metadados necessários para representar o conteúdo criativo
<i>Curriculum Ontology:</i>	modelos de dados e vocabulários para descrever conteúdo do Reino Unido;
<i>Food Ontology</i>	Vocabulário que descreve assuntos sobre gastronomia
<i>Journalism Ontology</i>	usada para descrição da produção jornalística
<i>Politics Ontology</i>	Descrição de termos utilizados para o governo local e em eleições
<i>Programmes Ontology:</i>	Para definição de diferentes gráficos com nome utilizados na loja e viabiliza a associação com metadados que possibilitam o gerenciamento, validação e exposição de dados
<i>Storyline Ontology</i>	Representa notícias
<i>Wildlife Ontology:</i>	Simple vocabulário que descreve espécies biológicas habitats
<i>Business New Ontology</i>	Descrição de assuntos voltados aos negócios

Fonte: Elaborado pelas autoras

O BBC Music encaminha as suas *playlists* para as plataformas *Deezer*, *Spotify* bibliotecas de músicas e o *YouTube* para que possam ser reproduzidas integralmente, e

disponibiliza *links* para *Facebook*, *Twitter*, *Instagram*, *MusicBrainz*, *YouTube* relacionados a uma banda ou artista.

Para verificar os metadados do *BBC Music*, pode-se acessar a *datahub* que é uma plataforma derivada da *Open Knowledge International*, baseada no sistema de gerenciamento CKAN, uma importante ferramenta de gerenciamento e publicação de coleções de dados *open-source* desenvolvida pela Fundação *Open Knowledge* em 2006, Ckan (2017), utilizada por órgãos governamentais, institutos de pesquisa, e diversas organizações que coletam e trabalham com um grande volume de dados.

Visto isso, o *Open Refine software* livre, criado em 2010 para trabalhar com dados considerados confusos, de modo a realizar a limpeza, conversão do seu formato original., viabilizou a verificação dos metadados do *BBC Music* a partir do acesso ao *datahub*.

Visto isso foi possível verificar os metadados do *BBC Music*:

Quadro 2: Metadados *BBC Music*

Arquivo	Neste são inseridos o que é o arquivo e nome dentro do ()
Nome	O respectivo nome do arquivo, uma das formas mais buscadas
Título	O título oficial que assim como nome também é uma das formas mais buscadas
Descrição	Na descrição é apresentado um pouco sobre o <i>BBC Music</i> , um portal específico de áudio e música que compõe grupo BBC, também apresenta nomes de alguns dos responsáveis pelo portal
Palavras-chave	Palavras indexadas e comumente relacionadas com o <i>BBC Music</i> como, <i>bbc</i> , <i>format-foaf</i> , <i>format-rdf</i> , <i>lod</i> , <i>media</i> , <i>music</i>
Recurso – URL	Disponibiliza um <i>link</i> para serviços
Recurso – Formato	Os formatos como SPARQL, RDF/XML e RDF <i>Schema</i>
Recurso – Descrição	A descrição do recurso em seus formatos, SPARQL endpoint, RDF/XML, XML e RDF <i>Schema</i>
Extras – Triplas	As sua triplas (RDF) que no caso são 20000000
Extras – Namespace	<i>Link</i> que encaminha aos namespaces utilizados
Extras – <i>Link</i> Licença	<i>Link</i> que encaminha a licença do portal
Extras – <i>Links</i> DBpedia	<i>Links</i> relacionados a DBpedia somando 2300
Fontes – <i>Web</i>	Encaminha ao portal do <i>BBC Music</i>

Fontes – Nome	Nome de referência
Fontes – E-mail	Encaminha ao <i>link</i> para conhecer mais sobre o nome de referência que no caso é Tom Scott.

Fonte: Elaborado pelas autoras

Em Recursos URL do *BBC Music* três *links* são disponibilizados, o primeiro não é possível acessar na *Web*, o segundo abre uma página do *BBC Music* na *Web* sobre a artista Madonna e o terceiro realiza o *download* de um arquivo com ontologias de música que pode ser aberto no bloco de notas (txt) ou em *softwares* específicos de esquemas XML e RDF como o *XML Copy Editor* que possibilitam a ateração e visualização destes esquemas.

4 CONSIDERAÇÕES FINAIS

Ao que se refere à *BBC Music*, verificou-se que em sua estrutura existe um padrão de metadados, que além de organizar seus recursos, acrescentam informações que podem ser acessadas por meio de *links* que direcionam a conteúdos objetivos e completos, também foi identificado que a rede BBC, busca aprimorar seus produtos, para seus cliente, bem como o utilizar vocabulários de outros *datasets* no *Linked Open Data*, como o *MusicBrainz*.

Além disso, foi possível observar que os metadados podem ser utilizados para a organização, descrição e representação dos recursos sonoros, o que revela que o uso de metadados pode influenciar na eficiência da recuperação de recursos sonoros na *Web*.

Assim, verifica-se que esse tema pode trazer novas formas de representação dos recursos sonoros, e possibilidades, no que tange novas perspectivas de tratamento de recursos informacionais no contexto da *Web*, *Linked Data* e no projeto *Linked Open Data*.

REFERÊNCIAS

ARAKAKI, F. A. **Linked Data**: ligação de dados bibliográficos. 2016.

BBC. **Welcome**. Disponível em: <<http://www.bbc.co.uk/ontologies>>. Acesso em: 14 ago. 2018

BARRETO, A. DE A. A questão da informação. São Paulo. **Perspectiva**, v. 8, n. 4, p. 3–8, 1994.

BERNERS-LEE, T. **Information Management**: a proposal. CERN.1989

BERNERS-LEE, T. Tim Berners-Lee on the next Web. **TED Talks**, 2009. Disponível em: <https://www.ted.com/talks/tim_berniers_lee_on_the_next_Web>. Acesso em: 22 mai. 2018.



CKAN. **BBC Music**. Disponível em: <<https://datahub.ckan.io/dataset/bbc-music>>. Acesso em: 02 ago. 2018.

FERREIRA, J.M. **Recuperação de Informação de Música e dados ID3**: possíveis aplicações. 2015.

GRÁCIO, J. C. A. **Metadados para a descrição de recursos da Internet**: o padrão Dublin Core, aplicações e a questão da interoperabilidade. 2002.

HEATH, T.; BIZER, C. **Linked data**: evolving the Web into a global data space. 2011.

ISOTANI, S.; BITTENCOURT, I. I. **Dados Abertos Conectados**: Em busca da Web do Conhecimento. 2015.

OLIVEIRA, E. B.; RODRIGUES, G. M. **O conceito de memória na Ciência da Informação**. 2011.