



XXII Encontro Nacional de Pesquisa em Ciência da Informação – XXII ENANCIB

ISSN 2177-3688

GT-2 – Organização e Representação do Conhecimento

DEFININDO INTEGRAÇÃO DE DADOS NA WEB A PARTIR DO DESENVOLVIMENTO DOS CATÁLOGOS

DEFINING DATA INTEGRATION BASED ON THE DEVELOPMENT OF CATALOGS

Sarah Rúbia de Oliveira Santos. UFMG.

Célia da Consolação Dias. UFMG.

Modalidade: Trabalho Completo

Resumo: A integração de dados, apesar de ser mencionada no contexto dos ambientes digitais, remonta a uma inquietação que aparece ao longo da história dos padrões de catalogação. A partir da contextualização dessa história, o objetivo deste trabalho é compreender em que consiste essa integração, verificar qual a necessidade e importância de sua realização, e propor uma definição para o termo “integração de dados” no contexto da Ciência da Informação. Busca-se elucidar, no contexto da Ciência da Informação, como surgiu a integração de dados, em que consiste essa integração, e verificar qual a necessidade e importância de sua realização. Este é um estudo qualitativo, exploratório e descritivo, com vistas a verificar sob quais perspectivas a “integração” é abordada na literatura brasileira de Biblioteconomia e Ciência da Informação. Um levantamento bibliográfico foi realizado na Brapci e 14 artigos compuseram a discussão sobre o tema. A integração é uma temática abordada na área, sendo mencionada nos tópicos que abrangem a organização do conhecimento em ambientes digitais e o uso das tecnologias da Web Semântica e dos Dados Interligados, mas não é um termo explicitamente definido. A integração aparece como um objetivo a ser alcançado e tem alguns de seus benefícios mencionados. Como contribuição deste trabalho, propõem-se que o termo seja definido como a união de um ou mais elementos, que pode ocorrer no nível das fontes de informação, dos sistemas e das unidades de informação, e está relacionada com uma visão integrada de um conjunto de recursos de informação.

Palavras-Chave: Integração de Dados. História da Catalogação. Web Semântica. Dados Interligados.

Abstract: Data integration, despite being mentioned in the context of digital environments, goes back to a concern that appears throughout the history of cataloging standards. Contextualizing this history, the objective of this work is to propose a definition of “data integration”. It seeks to elucidate, in the context of Information Science, how data integration came about, what this integration consists of, and to verify the need and importance of its implementation. This is a qualitative, exploratory and descriptive study, seeking to check under which perspectives “integration” is approached in the Brazilian literature on Librarianship and Information Science. A bibliographic survey was performed at Brapci and 14 articles contributed to the discussion on the topic. Integration is a topic addressed in the area, being mentioned in topics that cover the organization of knowledge in digital environments and the use of Semantic Web and Linked Data technologies, but it is not an explicitly defined term. Integration appears as an objective to be achieved and has some of its benefits mentioned. As a contribution of this work, it is proposed that the term be defined as the union of one or more elements,



which can occur at the level of information sources, systems and information agencies, and is related to an integrated view of a set of information resources.

Keywords: Data Integration. History of Cataloging. Semantic Web. Linked Data.

1 INTRODUÇÃO

Os catálogos são instrumentos que permitem às instituições de memórias organizar e divulgar seus acervos. Ao longo de sua evolução, adquiriram diferentes formatos como as listas de títulos, de autores e de assuntos; catálogos de fichas impressas; e, com a chegada dos computadores, os catálogos on-line. Pode-se falar, atualmente, sobre como a Web proporcionou os meios para que os recursos de informação disponíveis sejam localizados de forma integrada, notadamente pelo uso das tecnologias da Web Semântica e dos Dados Interligados.

A partir da história dos catálogos e do desenvolvimento das regras de catalogação, entende-se que é possível identificar uma linha do tempo capaz de apresentar como surgiu a necessidade da integração de dados. O objetivo deste estudo é compreender em que consiste essa integração, verificar qual a necessidade e a importância de sua realização, e propor uma definição para o termo “integração de dados” no contexto da Ciência da Informação.

A estrutura do artigo está como segue: no tópico 2 apresenta-se a metodologia, seguida da história da catalogação, buscando compreender quais acontecimentos influenciaram a ideia de integração no tópico 3. No tópico 4, diferentes tipos de integração são apresentados sendo feitas reflexões sobre seus objetivos. Por fim, são apresentadas as considerações finais.

2 METODOLOGIA

O estudo é qualitativo, exploratório e descritivo. O levantamento bibliográfico foi realizado em duas etapas e em duas bases bibliográficas distintas. Primeiro, contemplando a discussão e a contextualização da integração de dados realizada na seção 3, realizou-se uma busca sobre a história da catalogação na Web of Science e na Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI). A escolha pelos estudos selecionados baseou-se na pertinência dos trabalhos para descrever a história da catalogação. Foram selecionados artigos em inglês e português, publicados após 2006, ano em que se dá a publicação de Tim Berners-Lee sobre os Dados Interligados no contexto da Web Semântica,



procurando averiguar como a história da catalogação e dos catálogos converge para a integração de dados em ambientes digitais.

Com vistas a verificar sob quais perspectivas o termo “integração” é abordado na literatura brasileira de Biblioteconomia e Ciência da Informação, realizou-se um levantamento bibliográfico na BRAPCI. Ao todo, 57 artigos foram recuperados. Dos quais 14 estão relacionados diretamente com a integração de dados e foram selecionados para leitura na íntegra e discussão no presente estudo.

3 CONTEXTUALIZANDO A INTEGRAÇÃO DE DADOS A PARTIR DA HISTÓRIA DA CATALOGAÇÃO

A integração aparece como uma temática fortemente relacionada com as tecnologias da Web Semântica e com os dados interligados. É por meio dessas tecnologias que dados de múltiplos acervos podem ser acessados de forma integrada a partir de um único sistema. No entanto, a integração de dados remonta a uma inquietação que aparece ao longo da história dos padrões de catalogação, com os trabalhos de descrição e representação de materiais bibliográficos, pensados para viabilizar e facilitar o acesso à informação para os usuários.

Nesta seção, acontecimentos importantes para a história da catalogação que influenciaram como os catálogos on-line e a integração são entendidos, atualmente, serão pontuados brevemente. Para fins de organização da seção, optou-se por seguir a divisão temporal que Kalita e Deka (2020) fazem para descrever o desenvolvimento dos metadados usados em bibliotecas. Essa divisão se dá em quatro momentos: antes de 1900, 1900-1950, 1950-pré-web e pós-web.

Antes de 1900, dois eventos se destacaram no que concerne à criação de regras para a organização de catálogos. O primeiro, ocorrido em 1841, foi a publicação das 91 regras para compilação de um catálogo (*Rules for the Compilation of the Catalogue*) de Anthony Panizzi. Suas regras definiam como registrar títulos e nomes de autores, o que fazer com trabalhos anônimos, e assim por diante. O segundo evento, foi a publicação do código de catalogação proposto por Charles Cutter. Ele foi publicado pela primeira vez em 1876 como *Rules for a Printed Dictionary Catalogue*. As regras de Cutter se destacam porque, além de listar os itens de uma biblioteca por autor, listava-os também por título e assunto, todos juntos em uma lista ordenada alfabeticamente. De forma prática, em um sistema de catálogo de fichas, isso significava que todas as fichas seriam arquivadas no mesmo conjunto de gavetas (DENTON,



2007; KALITA; DEKA, 2020). Esses dois eventos demonstravam a crescente preocupação com a padronização dos catálogos e as formas como eles poderiam ser acessados.

Entre 1900 e 1950, o destaque em relação à organização de catálogos ocorre por meio do trabalho colaborativo entre instituições para discutir e elaborar padrões de catalogação. A American Library Association (ALA), que já trabalhava em conjunto com a Library of Congress, fez uma fusão, do que já havia sido desenvolvido por elas, com as regras elaboradas pela Library Association of the United Kingdom. A partir dessa colaboração, o primeiro conjunto de regras de catalogação anglo-americanas foi publicado em 1908. O *Catalog Rules: Author and Title Entries* foi desenvolvido nos moldes das regras de Panizzi e Cutter (DENTON, 2007; DOBRESKI, 2020; KALITA; DEKA, 2020).

Críticas a essas regras eram constantes, tanto pelo formalismo e dificuldade de se aplicar as regras, quanto pela incapacidade do instrumento de lidar com as variações descritivas exigidas para diferentes tipos de itens documentais na catalogação. Assim, outros instrumentos seguiram-se, como uma segunda edição das regras anglo-americanas em 1949, conhecida como Livro Vermelho; e o *Rules for descriptive cataloguing in the Library of Congress*, chamado de Livro Verde, publicado no mesmo ano (DENTON, 2007; KALITA; DEKA, 2020). Nessa época, destacaram-se a colaboração entre instituições com vistas a padronizar o processo de catalogação, diferente dos anos anteriores em que as regras de catalogação foram elaboradas por autores individualmente. Além disso, pontua-se a gradativa atenção dispensada a outros itens documentais disponíveis nas bibliotecas, como periódicos, materiais cartográficos, músicas, filmes, entre outros.

Na era entre 1950 e a pré-web, frente às críticas que as regras criadas pela ALA estavam recebendo, a divisão de Catalogação e Classificação da instituição requisitou à Library of Congress um estudo intensivo sobre as regras de catalogação, trabalho que foi confiado ao professor Seymour Lubetzky. Na ocasião, ele “apontou falhas nas regras de entrada e expôs a necessidade de estabelecer um conjunto de princípios para que um código melhorado pudesse ser construído” (ATTI, 2021, p. 31).

Lubetzky simplificou o uso das regras de catalogação, tornando-as mais objetivas, consistentes e voltadas para o fazer profissional dos catalogadores. Seu trabalho resultou no *Code of Cataloging Rules; Author and Title Entry* e formou a base para a declaração dos 12 Princípios de Paris da International Federation of Library Association (IFLA) na Conferência



Internacional sobre Princípios de Catalogação em Paris em 1961, princípios que tinham por intuito servir de base para uma normalização internacional na catalogação (DENTON, 2007; KALITA; DEKA, 2020; ATTI, 2021).

O Princípio de Paris foi seguido, mais tarde, pelo Código de Catalogação Anglo-Americano (AACR) em 1967, que passou por grandes revisões após a IFLA introduzir a perspectiva do *International Standard Bibliographic Description* (ISBD). O ISBD, apesar de não ser um código de catalogação, fornecia as diretrizes para produzir um padrão que fosse reconhecível e permitisse a fácil troca de registros criados por diferentes instituições. Em 1978 foi lançado o AACR2, código de catalogação amplamente difundido, que se divide em duas partes: parte I - Descrição e parte II - Cabeçalhos, Títulos Uniformes e Referências. A primeira parte baseou-se no ISBD e a segunda parte nos Princípios de Paris (DENTON, 2007; KALITA; DEKA, 2020). O grande destaque dessa época pode ser atribuído à visão de Lubetzky de um catálogo funcional para a biblioteca, que influenciou o desenvolvimento de padrões e diretrizes para normalizar a descrição de itens documentais. Outro ponto relevante foi o foco do ISBD na internacionalização e na troca de informações entre instituições.

No **Pós-web**, surgem as questões em torno da mudança dos catálogos para ambientes digitais. Começam então as atividades de informatização para lidar com a explosão bibliográfica resultante da demanda por pesquisas e crescimento de instituições acadêmicas. As atividades de informatização para catalogação e a participação ativa da Library of Congress, que publicava seu catálogo em fichas até então, resultaram no registro catalográfico legível por computadores, na forma do Machine Readable Cataloging (MARC). O formato foi inspirado nas especificações do ISBD e nas regras do AACR2 (KALITA; DEKA, 2020; MARCONDES, 2021). Além dos desenvolvimentos do MARC, havia o interesse em desenvolver uma abordagem de descrição mais baseada nos princípios de entidade-relacionamento para dados bibliográficos. Interesse seguido pela elaboração do modelo conceitual da IFLA, Requisitos Funcionais para Registros Bibliográficos (FRBR), publicado pela primeira vez em 1998. Esse modelo conceitual, mais tarde, embasaria a criação do *Resource Description and Access* (RDA), código de catalogação organizado em torno das entidades conceituais estabelecidas pelo FRBR e indicado para descrever objetos digitais, dados e catálogos (DOBRESKI, 2020).

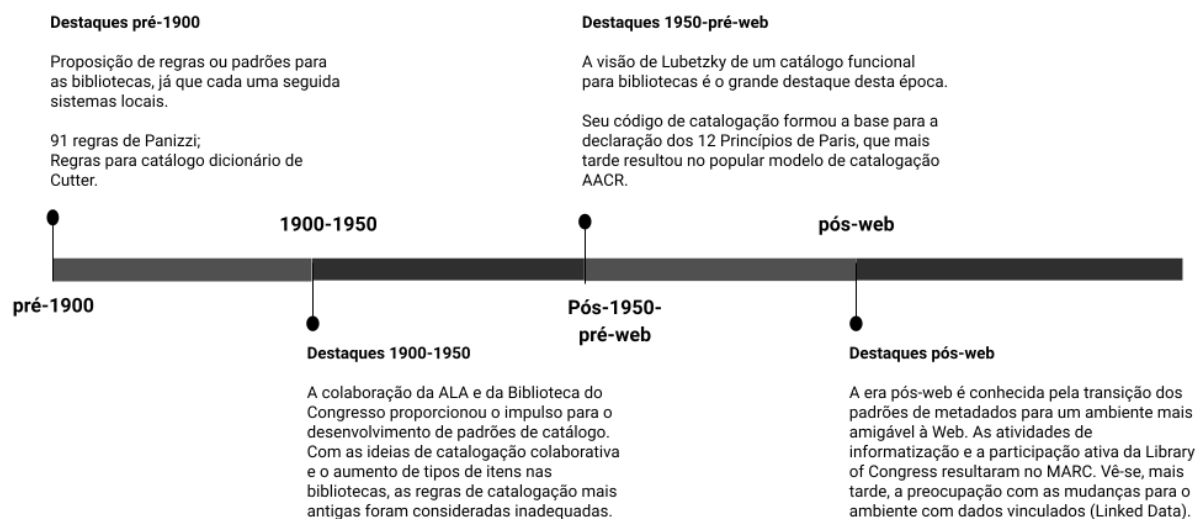


Por fim, com o surgimento da World Wide Web na década de 1990, as bibliotecas tiveram que pensar em como aumentar sua presença no ambiente digital. Nesse contexto, em 1995, o esforço combinado do Ohio Computer Library Center (OCLC) e do National Center for Supercomputing Applications (NCSA) resultou em um esquema de metadados “genérico” denominado Dublin Core (KALITA; DEKA, 2020; MARCONDES, 2021).

Esse esquema rapidamente recebeu a atenção da comunidade global por fornecer um vocabulário capaz de descrever uma ampla gama de recursos de maneira interoperável na Web, isto é, fazendo com que tenham descrições padronizadas e possibilitem a troca de informações entre sistemas. Com o amadurecimento das tecnologias da Web Semântica e, em particular, do formato *Resource Description Framework* (RDF), utilizado como modelo padrão para intercâmbio de dados na Web, a integração de recursos de informação tornou-se algo ainda mais desejável (KALITA; DEKA, 2020; MARCONDES, 2021; DUBLIN..., c2022).

Na Figura 1, apresenta-se um resumo da evolução da catalogação a partir das 4 divisões temporais discutidas.

Figura 1 - Evolução da catalogação



Fonte: Elaborado pelas autoras (2022).

Observa-se que houve evolução na forma como a catalogação foi pensada ao longo do tempo. Melhorias foram agregadas às regras de catalogação à medida que as tecnologias avançaram, mas, mais do que isso, melhorias foram influenciadas pela comunidade de profissionais que se posicionou quanto às necessidades das bibliotecas e de seus usuários. As mudanças mais recentes advindas do surgimento das tecnologias da Web Semântica



possibilitaram pensar um catálogo que vai além das paredes da biblioteca e envolve outras instituições de memória, como os arquivos e os museus. Elas trouxeram a possibilidade de descentralização e integração de diferentes acervos disponíveis em catálogos por todo o mundo. Tendo em vista o exposto, na próxima seção discute-se as noções de integração, buscando conceituá-las a partir de suas especificidades.

4 CONCEITUANDO OS DIFERENTES TIPOS DE INTEGRAÇÃO

Com vistas a alcançar o objetivo de propor uma definição acerca da integração, questiona-se: o que é integração e o que pode ou precisa ser integrado? No levantamento da literatura brasileira, foram identificados os seguintes termos relacionados à integração: a integração de acervos; integração de catálogos; integração de dados; integração de modelos conceituais; integração de sistemas de informação; e integração semântica. Nesses trabalhos, observa-se que a integração aparece associada a noções de cooperação, compartilhamento, interação, (inter)ligação, reunião e unificação entre dois ou mais elementos (CARRASCO; VIDOTTI, 2019; MARCONDES, 2016, 2017; RIBEIRO; BORGES; SOUSA, 2020; SIMIONATO; ARAKAKI; SANTOS, 2017; SOUSA; RIBEIRO, 2021).

O termo integração aparece, geralmente, acompanhado de outros como “interoperabilidade” (ALMEIDA, 2002; CASTRO; SANTOS, 2010; MARCONDES, 2016, 2017) e “compartilhamento” ou “intercâmbio de dados” (ALVES; SOUZA, 2007; CARRASCO; VIDOTTI; SANTARÉM SEGUNDO, 2017; CARRASCO; VIDOTTI, 2019; RIBEIRO; BORGES; SOUSA, 2020).

Marcondes (2021, p. 19) afirma que a interoperabilidade se constitui como a capacidade de diferentes sistemas operarem em conjunto através de “padrões tecnológicos, instrumentos semânticos, acordos ou propostas”, com vistas a intercambiar conteúdos mantendo seu significado original. Para que isso seja possível, ela opera em diferentes dimensões, a saber: sintática, semântica, administrativa/política, e tecnológica. Para Alves *et al.* (2017) a interoperabilidade possibilita, entre outras coisas, a integração. A integração de acervos digitais de bibliotecas, arquivos e museus, especificamente, é “uma preocupação e um desafio conceitual, metodológico e tecnológico” (MARCONDES, 2021, p. 19) que concerne à interoperabilidade.

A integração aparece como um objetivo a ser alcançado por meio da interoperabilidade e recorrendo às tecnologias da Web Semântica e do *Linked Data*. Arakaki, Simionato e Santos (2017, p. 2250) evidenciam essa afirmação quando declaram que “a



integração é fundamental para proporcionar um ambiente ligado e ampliar as possibilidades de navegação dos usuários”. Ela ainda é retratada de duas formas na literatura: como um processo e como uma solução. Como um processo que permite “dinamizar o acesso aos distintos objetos digitais que representam os acervos arquivístico, bibliográfico, museológico e de coleções biológicas” (ALVES *et al.*, 2017, p. 2) da Fundação Oswaldo Cruz. E como uma solução que pode proporcionar uma “recuperação mais precisa, relevante e significativa para o usuário final” (CASTRO; SANTOS, 2010, p. 156).

O mais próximo de uma definição de “integração” é o que Serra *et al.* (2017) definem, com base na obra *The Intellectual Foundation of Information Organization* de Svenonius (2000), como o princípio da integração. Esse princípio diz que a integração depende de um conjunto de regras para descrever diferentes tipos de recursos de informação. Por meio da integração há a aproximação de sistemas bibliográficos com escopos semelhantes, mas existe complexidade em tentar obter um conjunto de regras capaz de descrever recursos de informação de sistemas bibliográficos diversos (SVENONIUS, 2000; SERRA *et al.*, 2017).

Nos trabalhos levantados na BRAPCI não foram identificados conceitos e/ou definições sobre o que, de fato, é a integração. Em consulta ao Dicionário de Biblioteconomia e Arquivologia, o termo “integração” é definido, de forma geral, como o “compartilhamento de dados ou informação entre subsistemas e sistemas” (CUNHA; CAVALCANTI, 2008, p. 208). A obra de referência indica que a integração possui diferentes níveis e se divide em três deles: a integração de dados, de formatos e de sistemas. Esses três tipos baseiam-se, essencialmente, na ideia de que uma diretriz é necessária para que um conjunto comum de regras seja utilizado para descrever diferentes tipos de mídias (SVENONIUS, 2000), utilizando formatos para descrição que permitam que conteúdos disponíveis em sistemas de informação sejam integrados.

Considerando os três tipos de integração, a **integração de sistemas** corresponde a diferentes tipos de computadores interagindo conjuntamente (CUNHA; CAVALCANTI, 2008). Em geral, a integração de múltiplos sistemas visa combiná-los formando um conjunto que dá aos usuários a impressão de que estão interagindo com um único sistema de informação. Marcondes (2016), ao mencionar a integração entre acervos arquivísticos, bibliográficos e museológicos, identifica como “interoperabilidade” o processo em que tecnologias e padrões como o MARC, Dublin Core, o Z39.50 e outros são utilizados para permitir a consulta unificada



a diferentes acervos. Em suma, o tema da integração de sistemas de informação contempla a interoperabilidade entre sistemas no sentido em que eles são colocados para se comunicar e interagir, independente da abordagem tecnológica de suporte.

Para preencher a lacuna, ausente na literatura nacional, referente à definição de termos, sobretudo de “integração de dados” e “integração de formatos”, foi necessário consultar alguns textos da literatura estrangeira (TEEL, 1994; STUDWELL, 1996; GRAHAM; JOHNSTON, 1997; GARDNER, 2005, ZIEGLER; DITTRICH, 2007; HYVÖNEN, 2012; CHEATHAM; PESQUITA, 2017). A literatura brasileira recuperada, apesar de não abordar a integração de formatos, faz alusão a esta ao tratar da integração ou harmonização de modelos conceituais, como será observado em breve.

A **integração de dados** pode ser entendida como a conversão de dados derivados de origens diversas em um conjunto compatível (ao nível de codificação e semântico), visando localizá-los e acessá-los pelos mesmos critérios de busca (CUNHA; CAVALCANTI, 2008). Esse tipo de integração pode ocorrer em diferentes fases da recuperação da informação. De acordo com Hyvönen (2012), existem duas estratégias principais disponíveis: 1) integrar os dados dinamicamente durante o processamento de consultas, isto é, quando os dados estão sendo buscados; ou 2) integrar os dados antecipadamente em um estágio de pré-processamento separado, a partir da descrição e organização deles. Nota-se que o segundo ponto trata da integração de dados a partir de uma perspectiva semântica.

A integração semântica de dados é um processo através do qual dados semanticamente heterogêneos podem ser integrados com perda mínima de informação (CHEATHAM; PESQUITA, 2017). Conforme Ziegler e Dittrich (2007, p. 48, *tradução nossa*),

a integração semântica é a tarefa de agrupar, combinar ou completar dados de diferentes fontes, levando em consideração a semântica dos dados explícita e precisa, a fim de evitar que dados semanticamente incompatíveis sejam estruturalmente mesclados. [...] Um pré-requisito para isso é resolver a ambiguidade semântica em relação aos dados integráveis por metadados explícitos para obter todas as suposições implícitas relevantes e informações de contexto subjacentes.

Segundo Gardner (2005), muitos projetos de integração de dados carecem de uma verdadeira integração semântica. Por mais que se tente vincular os dados com significado, acaba-se criando um “silo” de informações parcialmente vinculadas, que não são muito mais



acessíveis do que eram anteriormente. Geralmente, a integração torna-se estática, já que novas questões e contextos não são adicionados facilmente aos dados.

Existe uma série de desafios que dificultam a integração semântica de dados, a saber: a acessibilidade dos dados; fornecimento de dados com contexto semântico para apoiar sua interpretação; e o estabelecimento de ligações significativas entre os dados (CHEATHAM; PESQUITA, 2017). Observou-se que uma abordagem eficaz para esse tipo de integração é focar explicitamente na organização e representação do conhecimento. A ideia é que se um recurso é bem descrito e representado, ele pode ser conectado com outros recursos. Essa conexão contribuiria com o enriquecimento do recurso inicial, mas também com o enriquecimento dos demais recursos, formando uma estrutura semanticamente consistente (GARDNER, 2005).

Isso é possível a partir das tecnologias da Web Semântica e dos dados abertos interligados, que permitem a construção de um ambiente interoperável. Com a aplicação dessas tecnologias, as descrições de conteúdo podem ser mais precisas baseando-se em identificadores uniformes de recursos (*Uniform Resource Identifier* – URI), adicionados a uma estrutura de descrição interoperável, o RDF. Ao combinar fontes de dados de forma interoperável, os dados de uma fonte podem ser enriquecidos semanticamente com dados vinculados adicionais de outra fonte (HYVÖNEN, 2012; MARCONDES, 2021).

No contexto da descrição dos recursos de informação, a **integração de formatos** possibilita que os mesmos campos, subcampos e designadores utilizados num registro bibliográfico podem ser utilizados na descrição de qualquer tipo de documento, independente de seu suporte ou conteúdo (CUNHA; CAVALCANTI, 2008). Na década de 1990, a integração de formatos era discutida no contexto da catalogação. Com ela, buscava-se integrar os formatos separados, isto é, os elementos descritivos que existem para livros, publicações seriadas, mapas, música, materiais visuais, manuscritos, arquivos de computador e controle de arquivos (GRAHAM; JOHNSTON, 1997).

Com o crescimento das bibliotecas, a evolução das tecnologias e a criação de novos formatos, outras formas de integrar a descrição de formatos específicos foram pensadas. Alguns dos métodos utilizados para esse fim incluíam estender o uso de *tags* válidas para a descrição; excluir e tornar obsoletas algumas *tags* e códigos; adicionar novas *tags* e códigos válidos para todos os materiais; esclarecer nomes de elementos e descrições de campos existentes; e reconciliar diferentes práticas de codificação entre formatos (TEEL, 1994).



Como cada formato tem elementos descritivos específicos, tornava-se um desafio descrever formatos diferentes utilizando um instrumento capaz de descrever apenas materiais bibliográficos (STUDWELL, 1996; GRAHAM; JOHNSTON, 1997). Alguns dos desafios desse tipo de integração incluem lidar com a inexistência de campos para descrever especificidades de um formato e com a incompletude de informações para o mesmo fim. Já havia, assim, uma preocupação com a unificação e com a descrição mais completa de materiais impressos, não impressos e digitais. A integração de formatos atingiu alguns objetivos, como descrever diversos formatos físicos; expressar a serialidade de materiais não impressos; e controlar o arquivamento de materiais não impressos (GRAHAM; JOHNSTON, 1997).

Se o desafio já era grande tendo em vista apenas o domínio bibliográfico, a multiplicidade e a heterogeneidade dos recursos de informação existentes na atualidade adicionam uma camada de complexidade à integração desses recursos. Considerando a ideia geral de descrever materiais heterogêneos com um único instrumento, o conceito de integração de formatos se relaciona com as noções de integração e harmonização de modelos conceituais, e sua capacidade de descrever conteúdos de acervos heterogêneos.

Os modelos conceituais se propõem a facilitar a interoperabilidade entre acervos de diferentes unidades de informação, interligando os objetos digitais sem que haja perda de seu significado original (MARCONDES, 2015, 2017). Pode-se entender os modelos conceituais como

uma REPRESENTAÇÃO, sob a forma de uma descrição abstrata e genérica, de determinado DOMÍNIO da realidade, para fins de compreender esta realidade, raciocinar sobre ela e fixar um entendimento comum dessa realidade dentro de uma comunidade. Como toda representação de algo, um modelo conceitual é algo construído por alguém, com alguma finalidade prática (MARCONDES, 2021, *grifos do autor*).

O processo de harmonização de modelos conceituais – que tem sido usado como sinônimo para integração de modelos conceituais – tem por objetivo identificar correspondências entre diferentes modelos conceituais para ser viabilizada a integração de acervos heterogêneos. Esse processo possibilita a interligação dos dados e acesso amplo e significativo a conteúdos culturais (CARRASCO; VIDOTTI, 2019). E pode ser entendido como a integração dos princípios descritivos dos contextos da Arquivologia, Biblioteconomia e Museologia (SIMIONATO; SANTOS, 2017).



Para Carrasco (2019, p. 15), “a harmonização de modelos conceituais, entendida como um processo de consenso terminológico-conceitual, é um requisito importante para garantir a integração dos acervos heterogêneos do patrimônio cultural disponíveis em ambientes digitais”. Alguns exemplos de modelos conceituais são o Modelo de Requisitos Funcionais para Registros Bibliográficos (FRBR), utilizado em bibliotecas, uma clara evolução das regras de catalogação discutidas no tópico anterior; o Modelo CIDOC Conceptual Reference Model (CRM), em museus; o Modelo Records in Contexts conceptual model (RiC-CM), em arquivos; e o Europeana Data Model (EDM) utilizado para integrar a descrição dos recursos de informação das três unidades de informação mencionadas, além de outros centros de documentação.

A integração não é algo novo, mas, desde a informatização dos catálogos, tem sido considerada uma característica importante dos catálogos on-line, e continua atraindo interesse em vista das possibilidades de fornecer acesso a conteúdos diversificados no ambiente Web, como pode ser observado em iniciativas como a Europeana. A Europeana compreende uma “plataforma de pesquisa on-line e gratuita, desenvolvida no domínio das Humanidades Digitais, que oferece acesso global a conteúdos nato digitais e digitalizados do patrimônio cultural provenientes de bibliotecas, arquivos, museus e centros culturais europeus” (CARRASCO; VIDOTTI, 2020, p. 14). Assim, a integração auxilia no gerenciamento do conteúdo de coleções digitais e de recursos da Web, na indexação desses conteúdos, na unificação de banco de dados de texto completo e de outros recursos disponíveis em bibliotecas, arquivos e museus.

4 CONSIDERAÇÕES FINAIS

Conforme identificado na literatura recuperada na BRAPCI, base de dados com forte presença de publicações oriundas da Ciência da Informação, a integração é uma temática abordada, sendo mencionada nos tópicos que abrangem a organização do conhecimento em ambientes digitais e o uso das tecnologias da Web Semântica e dos Dados Interligados.

No entanto, nenhuma definição acerca do termo foi identificada na literatura consultada. Cabe destacar que a integração é abordada na literatura brasileira recuperada, mas não é um termo explicitamente definido. A integração é um objetivo a ser alcançado e alguns dos benefícios mencionados compreendem promover acesso unificado a um conjunto



de acervos que, outrora, estava disperso; e fornecer uma descrição padronizada para diferentes tipos de recursos de informação.

Para propor uma definição de “integração”, outros autores precisaram ser consultados. Svenonius (2000) define a integração como um princípio que envolve o uso de um conjunto de regras para descrever diferentes tipos de mídias, alegando que a integração demanda uniformidade na descrição. A definição de Svenonius (2000) no contexto da organização da informação. No Brasil, Serra et al. (2007) mencionam a definição de Svenonius acerca da integração de regras, mas a única definição formal de “integração” identificada está contida na obra “Dicionário de Biblioteconomia e Arquivologia” de Cunha e Cavalcanti. A partir dela, foi possível identificar na literatura internacional definições acerca da integração de dados, de formatos e de sistemas.

A contribuição deste estudo consistiu em levantar essa discussão e propor uma definição de integração. Nesse sentido, entende-se que a integração é a união entre dois ou mais elementos relacionados aos dados. Pode ocorrer no nível das fontes de informação, buscando integrá-las pela sua descrição e representação, a partir do momento que o mesmo conjunto de regras de catalogação, formatos e protocolos para descrição de recursos de informação são utilizados. Ocorre, ainda, no nível dos sistemas e das Unidades de informação, partindo da ideia que os sistemas devem se comunicar para que os acervos de bibliotecas, arquivos e museus possam ser localizados conjuntamente com maior facilidade. De forma geral, a integração está relacionada com uma visão integrada de um conjunto de recursos de informação, isto é, com a possibilidade de localizar e acessar esses recursos em um único sistema.

Assim, a integração ocorre por dois motivos: 1) para gerar uma visão integrada de sistemas dispersos através de um único sistema, de forma que o acesso, o uso e o reúso da informação sejam facilitados; e 2) para combinar os dados de sistemas de informação diferentes, mas complementares, de tal forma que se possa obter uma base de dados mais abrangente, capaz de melhor satisfazer as necessidades informacionais dos usuários.

Entender em que consiste a integração foi possível a partir da evolução das regras de catalogação descritas neste estudo, que evidencia o incremento na maneira de pensar a organização dos catálogos e a recuperação da informação para atender às necessidades de informação dos usuários. Como trabalho futuro, pretende-se investigar a integração de



recursos de informação, buscando identificar os requisitos necessários para que ela seja possível e como o assunto tem sido abordado na literatura da Ciência da Informação.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

REFERÊNCIAS

- ALVES, A. S.; OLIVEIRA, B. C. E. P. D.; SACRAMENTO, C.; GHIRARDELLI, C.; QUEIROZ, C. F.; NOGUEIRA, I.; SÁ, I. P.; PRAXEDES, K. V.; TERRA, M.; PINHEIRO, M. J. A.; FERRARI, R. Descrição de metadados mínimos para integração e preservação digital dos acervos científicos e culturais da Fundação Oswaldo Cruz: um estudo de caso. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 11, 2017. DOI: 10.29397/reciis.v11i0.1389. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1389>. Acesso em: 13 maio 2022.
- ALMEIDA, M. B. A necessidade de integração de fontes heterogêneas de dados em projetos de gestão de conhecimento. **Perspectivas em Ciência da Informação**, v. 7, n. 1, 2002. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/36527>. Acesso em: 28 jan. 2022.
- ALVES, M. D. R.; SOUZA, M. I. F. Estudo de correspondência de elementos metadados: Dublin Core e Marc 21. **RDBCi: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 5, n. 1, p. 20-38, 2007. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/2019>. Acesso em: 18 maio. 2022.
- ARAKAKI, F. A.; SIMIONATO, A. C.; SANTOS, P. L. V. A. Integrando catálogos entre bibliotecas, arquivos, museus e galerias de arte: perspectiva da Europeana e da DPLA. **Revista Brasileira de Biblioteconomia e Documentação**, v. 13, p. 2250-2268, 2017. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/1016>. Acesso em: 18 maio. 2022.
- ATTI, A. **Abordagem do processo de ensino de Seymour Lubetzky aplicado à disciplina de Catalogação descritiva**. 2021. Dissertação (Mestrado em Cultura e Informação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2021. doi:10.11606/D.27.2021.tde-30032022-142651. Acesso em: 29 abr. 2022.
- CARRASCO, L. B. **Integração de conteúdos culturais heterogêneos em ambientes digitais do patrimônio cultural**: harmonização de modelos conceituais. Tese (Doutorado) – Universidade Estadual Paulista (UNESP), Marília, 2019. Disponível em: <https://repositorio.unesp.br/handle/11449/182113>. Acesso em: 10 mar. 2022.
- CARRASCO, L. B.; VIDOTTI, S. A. B. G. Bibliotecas, arquivos e museus: um panorama sobre a curadoria digital da Europeana. **Memória e Informação**, n. 1, v. 4, set. 2020, p. 37-55. Disponível em: <http://www.memoriaeinformacao.casaruibarbosa.gov.br/index.php/fcrb/article/view/120>. Acesso em: 06 maio. 2022.



CARRASCO, L. B.; VIDOTTI, S. A. B. G. Patrimônio cultural: integração de acervos heterogêneos de museus e bibliotecas. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 20., 2019. Florianópolis. **Anais [...]**. Florianópolis: ANCIB, 2019. p. 1–14. Disponível em: <https://www.brapci.inf.br/index.php/res/v/121792>. Acesso em: 27 jan. 2022.

CARRASCO, L. B.; VIDOTTI, S. A. B. G.; SANTARÉM SEGUNDO, J. E. Convergência e ampliação de dados culturais em ambientes semânticos. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017. Marília. **Anais [...]**. Marília: ANCIB, 2017. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/105291>. Acesso em: 03 fev. 2022.

CASTRO, F. F.; SANTOS, P. L. V. A. C. Representação e descrição de recursos informacionais: aspectos estruturantes no delineamento de ambientes informacionais digitais. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 3, n. 1, 2010. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/119342>. Acesso em: 01 fev. 2022.

CHEATHAM, M.; PESQUITA, C. Semantic data integration. In: ZOMAYA, A. Y.; SAKR, S. **Handbook of Big Data Technologies**. New York: Springer, 2017. p. 263–305.

CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de Biblioteconomia e Arquivologia**. Brasília: Briquet de Lemos Livros, 2008.

DENTON, W. FRBR and the History of Cataloging. In: TAYLOR, A. G (Ed.). **Understanding FRBR: what it is and how it will affect our retrieval**. Westport, Connecticut: Libraries Unlimited, 2007.

DOBRESKI, B. Descriptive Cataloging: The History and Practice of Describing Library Resources. **Cataloging & Classification Quarterly**, v. 59, n. 2-3, p. 225-241, 2020. DOI: 10.1080/01639374.2020.1864693.

DUBLIN Core. **DCMI Metadata Terms**. [S.l.]: ASIS&T, 2022c. Disponível em: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/2022-05-05>. Acesso em: 05 maio. 2022.

GARDNER, S. P. Ontologies and semantic data integration. **Drug discovery today**, Elsevier, v. 10, n. 14, p. 1001–1007, 2005. DOI: [https://doi.org/10.1016/S1359-6446\(05\)03504-X](https://doi.org/10.1016/S1359-6446(05)03504-X).

GRAHAM, C.; JOHNSTON, J. Format integration and serials cataloging. **The Serials Librarian**, Taylor & Francis, v. 31, n. 1-2, p. 279–287, 1997. DOI: https://doi.org/10.1300/J123v31n01_13.

HYVÖNEN, E. **Publishing and using cultural heritage linked data on the semantic web**. San Rafael, CA: Morgan & Claypool Publishers, 2012. 159 p.

KALITA, D.; DEKA, D. Searching the great metadata timeline: A review of library metadata standards from linear cataloguing rules to ontology inspired metadata standards. **Library Hi Tech**, 2020. DOI: <https://doi.org/10.1108/LHT-08-2019-0168>.

MARCONDES, C. H. O papel dos modelos conceituais para interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus. In: ENCUESTRO IBERICO EDICIC, 7., 2015. Madrid. **Anais [...]**. Madrid: EDICIC, 2015. p. 1–14. Disponível em: <https://eprints.ucm.es/id/eprint/34550/>. Acesso em: 08 abr. 2022.



MARCONDES, C. H. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. **Perspectivas em Ciência da Informação**, v. 21, p. 61–83, 2016. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2735>. Acesso em: 04 fev. 2022.

MARCONDES, C. H. Publicando e interligando acervos digitais na web através das tecnologias de dados abertos interligados. **Revista Brasileira de Biblioteconomia e Documentação**, v. 13, p. 2135–2163, 2017. Disponível em: <https://www.brapci.inf.br/index.php/res/v/3085>. Acesso em: 03 fev. 2022.

MARCONDES, C. H. **Dados abertos interligados**: publicação, recuperação e integração de acervos de arquivos, bibliotecas e museus na web. Marília: Editora Oficina Universitária, 2021.

RIBEIRO, C. J. S.; BORGES, M. M.; SOUSA, A. M. C. d. Proposta de integração de recursos do patrimônio cultural científico no modelo EDM: uma perspectiva de representação de um domínio temático. **BIBLOS: Revista do Instituto de Ciências Humanas e da Informação**, v. 34, n. 2, p. 247–264, 2020. Disponível em: <https://periodicos.furg.br/index.php/biblos/article/view/11864>. Acesso em: 02 jan. 2022.

SERRA, L. G.; SANTARÉM SEGUNDO, J. E. S.; SANTOS, P. L. V. A. C.; ZAFALON, Z. R. Os princípios da descrição e os formatos MARC 21 e onix. **Ciência da Informação**, v. 46, n. 2, 2017. Disponível em: <http://revista.ibict.br/ciinf/article/view/2327>. Acesso em: 04 fev. 2022.

SIMIONATO, A. C.; ARAKAKI, F. A.; SANTOS, P. L. V. A. da C. Descrição em bibliotecas, arquivos, museus e galerias de arte: linkando recursos e comunidades. **Informação & Informação**, v. 22, n. 2, p. 449–466, 2017. ISSN 1981-8920. Disponível em: <https://www.uel.br/revistas/uel/index.php/informacao/article/view/31464>. Acesso em: 25 mar. 2022.

SIMIONATO, A. C.; SANTOS, P. L. V. A. C. Modelo conceitual DILAM: integração entre arquivos, bibliotecas e museus. **Informação & Sociedade: Estudos**, v. 27, n. 2, 2017. DOI: 10.22478/ufpb.1809-4783.2017v27n2.30902. Acesso em: 28 jan. 2022.

SOUSA, A. M. C. d.; RIBEIRO, C. J. S. Relações biográficas do patrimônio cultural no modelo edm: uma proposta de integração de acervos sobre Oswaldo Cruz. **Revista Ibero-Americana de Ciência da Informação**, Brasília, v. 14, n. 2, p. 367–381, 2021. Disponível em: <https://www.arca.fiocruz.br/handle/icict/50782>. Acesso em: 21 fev. 2022.

STUDWELL, W. E. Format integration: the mostly invisible new major change in cataloging. **Behavioral & Social Sciences Librarian**, Taylor & Francis, v. 14, n. 1, p. 63–65, 1996. DOI: https://doi.org/10.1300/J103v14n01_06.

SVENONIUS, E. **The intellectual foundation of information organization**. Cambridge: MIT press, 2000.

TEEL, K. Everything you always wanted to know about format integration, but were afraid to ask. **The Serials Librarian**, Taylor & Francis, v. 24, n. 3-4, p. 253–255, 1994. DOI: https://doi.org/10.1300/J123v24n03_38.



ENANCIB 2022

PORTO ALEGRE | UFRGS | PPGCIN

XXII Encontro Nacional de Pesquisa em Ciência da Informação • ENANCIB
Porto Alegre • 07 a 11 de novembro de 2022

ZIEGLER, P.; DITTRICH, K. R. Data integration: problems, approaches, and perspectives.

Conceptual modelling in information systems engineering. [S. l.]: Springer, 2007. p. 39–58.

DOI 10.1007/978-3-540-72677-7_3.