

# **Bibliotecário e Big Data: a intersecção da atuação profissional**

*Librarian and Big Data: the intersection of professional performance*

**Mariana Marroqui Rossi**

Discente da Faculdade de Biblioteconomia  
Pontifícia Universidade Católica de Campinas  
marianarossi13@icloud.com

## **Resumo**

A Quarta Revolução Industrial tem como um de seus principais impulsionadores tecnológicos, o Big Data. Essa tecnologia tem demonstrado grande usabilidade e aplicação em variados setores industriais e áreas do conhecimento. O presente estudo tem como objetivo compreender a atuação do profissional bibliotecário no contexto do Big Data, uma vez que dados e informação representam algumas das principais matérias-primas da profissão. Para sua elaboração, foi realizada uma pesquisa de caráter exploratório, com método misto, sendo uma análise de conteúdo para o tratamento qualitativo e estudos métricos da informação, mais especificamente, a bibliometria e cientometria para a realização do tratamento quantitativo, de artigos indexados na base LISA. Assim, estudos mostram que embora tímida, atuação do bibliotecário com o Big Data existe e se concentra na compreensão desta tecnologia para que a formação desse profissional possa ser atualizada e possa, dessa forma, desenvolver as habilidades e competências necessárias para lidar com esse fenômeno.

**Palavras-chave:** Big Data; Bibliotecário; Dados; Quarta Revolução Industrial.

## **Abstract**

*The Fourth Industrial Revolution has Big Data as one of its main technological drivers. This technology has demonstrated great usability in various industrial sectors and areas of knowledge. This study aims to understand the role of the professional librarian in the context of Big Data,*

*since data and information represent some of the main raw materials of this occupation. For its elaboration, an exploratory research was carried out, with mixed method, a content analysis for the qualitative treatment and metric studies of information, more specifically, bibliometrics and scientometry for the quantitative treatment, of articles indexed in LISA database. Therefore, studies show that although reticent, the librarian's role with Big Data exists and focuses on the comprehension of this technology so that the formation and training can be updated and, thus, develop the skills and competencies necessary to deal with this phenomenon.*

**Keywords:** Big Data; Librarian; Data; Fourth Industrial Revolution.

## Introdução

Os avanços científicos e tecnológicos ocorridos durante a 3ª revolução industrial, levaram a criação do conceito de Sociedade da Informação (CASTELLS, 2000), que tem entre suas características a flexibilidade, alta penetrabilidade e convergência de novas tecnologias e a informação como matéria-prima.

A Sociedade da Informação conta com as TICs (Tecnologias da Informação e Comunicação), instrumentos essenciais que perpassam todas as esferas da vida humana. O uso dessas TICs deixa um rastro digital que é formado de dados dinâmicos registrados, que podem ou não ser transformados em informação. Os dados são divididos em três categorias: estruturados, semiestruturados e não estruturados.

Essas categorias tratam dos modelos em quais os dados podem aparecer e sua organização. Dados estruturados, que possuem grande valor para o mercado e a ciência, são organizados e seguem modelos estritamente definidos, como por exemplo, bancos de dados relacionais ou planilhas, esses representam apenas 5% de todos os dados produzidos no mundo. Já os dados semiestruturados contém certo nível de organização em sua estrutura, porém não seguem um modelo padronizado. E por fim, os dados não-estruturados, não seguem um modelo pré-definido e nem estão organizados, representando 95% de toda a produção mundial e o maior desafio da área (SCHÖNBERGER; CUKIER, 2013).

Essa categorização de dados, de natureza digital e tradicional, recebe o nome de Big Data e são comumente caracterizados por 3 Vs: volume, velocidade e variedade. Contudo, essa caracterização tem crescido, e estudos já apontam 10 Vs (PATGIRI; AHMED, 2016).

As possibilidades proporcionadas pelo uso inteligente desses dados parecem ser infinitas. O Big Data trouxe consigo uma aplicabilidade versátil e flexível, se tornando um insumo econômico para todos os setores, sejam eles particulares ou públicos, dando “uma dimensão quantitativa [a vida] que nunca teve antes” (SCHÖNBERGER; CUKIER, 2013).

O surgimento desse conceito levou a novas considerações sobre tratamento e mensuração dos dados e da informação, já que os processos conhecidos se mostraram inadequados e obsoletos para lidar com suas particularidades.

Discutir Big Data e suas repercussões se mostra relevante visto o impacto desse conceito em diversas áreas e indústrias, como a saúde, o marketing, a previsão de desastres e outros.

Schönberger e Cukier (2013) discutem sua relevância dentro da área da saúde. A IBM, junto da Coriell Life Sciences, também tem explorado o atendimento personalizado, através Watson Health. O San Diego Supercomputer Center trabalha com a brigada de incêndio no estado da Califórnia para auxiliar na prevenção de incêndios florestais. O Big Data é um fenômeno universal, e como aponta Bernard Marr (2016), empresas de variados segmentos como mídias sociais, moda, telecomunicação, engenharia e varejo já o adotaram.

Ainda no âmbito das áreas e campos do conhecimento, também se observam reflexões e aplicações nos contextos de desenvolvimento científico, tecnológico e/ou reconfiguração das competências profissionais. No campo da Biblioteconomia e da Ciência da Informação, o Big Data já se faz presente.

O assunto é uma presença constante no campo, apesar de ter começado a ser discutido tardiamente (HOY, 2014). Existem diversos estudos sobre o Big Data dentro da Biblioteconomia e da Ciência da Informação, como por exemplo: curadoria digital em ambientes de Big Data e como esses ambientes podem ser usados como fontes de recursos informacionais para a curadoria digital (DUTRA, 2016), o uso das técnicas de Big Data, como mineração de dados, machine learning e análise de sentimento para análise de textos em grande escala em pesquisas científicas e jornalísticas para geração de conhecimento (ARCILA-CALDERÓN; BARBOSA-CARO; CABEZUELO-LORENZO, 2016), entre outros.

Sobretudo, prescreve-se a presença do Big Data na área da Ciência da Informação, avaliando a sua eficácia na análise de um conjunto de dados e a sua transformação. Além do mais, verifica-se o uso de Big Data na perspectiva de sua aplicação vinculada a parâmetros de organização de bases de dados e sua transformação em informação estratégica para variados contextos, inclusive o de negócios. Assim, diversas áreas do

conhecimento como a Ciência da Informação e a Biblioteconomia, têm aplicado o Big Data como sendo um meio fundamental para a evolução teórica e pragmática, em variados contextos.

Embora presente e inserido no contexto do Big Data, pouco se compreende sobre a atuação do profissional bibliotecário. Na perspectiva da atuação profissional, verifica-se que o Big Data tem ampliado as possibilidades de atuação do profissional bibliotecário, destacando-se dentre essas o bibliotecário de dados.

A falta de compreensão dessa atuação inviabiliza o estabelecimento estratégico de competências e habilidades que possam ser desenvolvidas a contento dos interesses, quer seja de negócios ou outras aplicações em que Big Data e o profissional bibliotecário sejam atuantes. Fundamentalmente, é preciso um olhar sobre o bibliotecário e sua atuação no mundo contextualizado dos grandes dados.

Nesse sentido, o objetivo do presente estudo é a compreensão das competências exigidas, os campos de atuação e relação com a própria Ciência da Informação quando condicionados a Big Data, a compreensão das atividades e também as limitações impostas aos profissionais bibliotecários que lidam com esse universo.

Para além disso, estudos como esse possibilitam um olhar sobre a formação atenta do bibliotecário para o contexto contemporâneo, exigido pela era do Big Data, visto que a adoção e a utilização dessa tecnologia têm se popularizado cada vez mais, fazendo do Big Data um dos impulsionadores tecnológicos da Quarta Revolução Industrial.

É preciso compreender quais as competências exigidas, os campos de atuação, qual a extensão das atividades e também suas limitações para que assim, seja possível apontar novos campos para os bibliotecários nesse cenário.

## **Big Data**

Não há um consenso quanto a definição do termo Big Data (SCHÖNBERGER; CUKIER, 2013). DeMauro e Grimaldi (2016) afirmam que o grande número de definições para o Big Data denotam seu "caótico estado da arte", por ser um assunto tão "jovem" e ainda afirmam que o termo abarca um "fenômeno social, recursos da informação, data sets, tecnologias de armazenamento, tecnologias de análise, processos e infraestrutura".

Zhan e Widén (2017) coletam 35 significações independentes do conceito em sua revisão de literatura. Entre as coletadas, existe um consenso das características citadas: são dados gerados em grande

velocidade, volume e variedade e geram desafios quanto a seu armazenamento, tratamento e processamento.

Por outro lado, Taurion (2013) destaca que o Big Data não é limitado a um produto, software ou hardware, mas sim, as tecnologias e práticas de análise de dados que permitem acesso a informações privilegiadas para a tomada de decisão estratégica e também para a automação de processos.

Schönberger e Cukier (2013), abordam a origem desse conceito, e relacionam sua origem com os campos da astronomia e da genômica, onde a quantidade de dados produzidos chegou a tal tamanho e variedade de formatos, que computadores normais não eram mais capazes de examiná-los.

Os contínuos estudos sobre Big Data significam evolução constante em sua conceituação. Inicialmente, Laney (2001), menciona 3 novas características em relação a gestão da explosão de dados: Volume, Velocidade e Variedade.

Gandomi e Haider (2014) sintetizam o significado dos 3 V's: o Volume abrange a magnitude dos dados. O aumento na produção de dados foi tamanho que as unidades de medidas e estratégias de armazenamento tiveram que ser revistas para suportar os números extraordinários; a Velocidade refere-se ao ritmo no qual os dados são gerados e finalmente, a Variedade ou os formatos nos quais os dados são produzidos: estruturados, semiestruturados e não estruturados.

O Big Data não para de crescer e seus números não demonstram nenhum sinal de desaceleração, assim como a população mundial na Internet e as dimensões da quantidade de bytes existentes chegaram a números exorbitantes: no ano de 2020, haverá 40x mais bytes de dados do que estrelas no universo observável (DOMO, 2019).

Os desdobramentos contínuos no estudo do Big Data agregaram mais características a sua conceituação. Após a consolidação dos 3Vs, foi sugerida a adição de mais 6Vs e 1C: Validade, Variabilidade/Volatilidade, Virtual, Visualização/ Visibilidade e Complexidade (TAURION, 2013; PATGIRI; AHMED, 2016).

Veracidade trata-se da precisão e verificação dos dados, um desafio do Big Data em relação a seu volume; Valor que é atribuído aos dados através de sua extração e análise; Validade dos dados dependendo do momento de sua produção ou entrada; Variabilidade/Volatilidade, embora os autores esclareçam que o termo Variabilidade é mais apropriado para descrição do Big Data, refere-se as alterações que os dados podem sofrer em curtos períodos de tempo; Virtual que refere-se aos processos de gestão de dados de maneira eficiente; Visualização/Visibilidade alude a capacidade de poder ver os dados, a visualização trata dos dados ocultos no Big Data,

ao tempo que a visibilidade discute o ponto de referência pelo qual o dado pode ser visto; e por fim, a Complexidade, associada a manipulação do Big Data, que por ter um número alto de dados de diversas naturezas, é considerada extremamente complexa (PATGIRI; AHMED, 2016).

Ainda é discutido o futuro das características do Big Data com a provável adição de mais Vs. Três deles estão associados ao Volume, devido ao constante aumento na proporção de dados gerados, sendo eles: Voluminosidade, Vácuo, Vitalidade. Outras adições são o "Vendee", que se refere "ao tamanho do cliente associado ao Big Data" e o "Vase", relacionado de maneira metafórica ao paradigma de armazenamento e as condições físicas necessárias para existência do Big Data (PATGIRI; AHMED, 2016).

O fenômeno do Big Data proporcionou aplicações nos mais variados campos. São governos, indústrias e áreas do conhecimento que a exploram constantemente, transformando suas áreas por meio técnicas de gestão de dados e análise, ou "analytics".

Gandomi e Haider (2015) apresentam algumas técnicas do Big Data Analytics, para dados estruturados e não estruturados: a análise de textos, que abrange a sumarização de textos, sistemas de resposta como a Siri da Apple e a mineração de opiniões; a análise de áudio e fala, que pode envolver sistemas LVCSR (transcritores de áudio) e sistemas baseados em fonética; a análise de vídeos; a análise de mídias sociais e a análise preditiva.

Para elucidar o uso dessas técnicas, pode-se resgatar um modelo rudimentar de análise preditiva em tempo real, combinado a modelo matemático, que permitiu a previsão de uma epidemia de gripe nos Estados Unidos, através do mapeamento dos termos de busca realizados no Google, por seus usuários. A eficácia da análise foi tamanha, que os números adquiridos por meio dessa técnica eram mais rápidos que aqueles fornecidos pelos Centro de Controle de Doenças. Esse sistema, desenvolvido por engenheiros do Google, foi de extrema importância para o lançamento de estatísticas na crise do H1N1, superando órgãos do governo americano (SCHÖNBERGER; CUKIER, 2013).

As diversas aplicações do Big Data foram popularizadas e consolidadas ao longo dos anos. Na área da saúde, seu uso tem se tornado cada vez mais frequente e oferecido resultados promissores, como o avanço na identificação e desenvolvimento de tratamentos personalizados de doenças, através do emprego dessa tecnologia. Pode-se apontar o exemplo da IBM, na identificação de seis novas potenciais proteínas para a pesquisa de câncer, realizada em apenas 30 dias, como uma aplicação satisfatória do Big Data, bem como a criação de tratamento personalizado para o

câncer, através da análise do histórico médico de pacientes. O IBM Watson Health também tem explorado soluções na oncologia, genômica e pesquisa clínica (LATTS, 2016).

A utilização do Big Data tem se tornado regular em outras áreas, entre elas o varejo. É possível listar mais de 40 grandes empresas, dos mais variados segmentos, que baseiam suas estratégias de venda, atendimento ao consumidor, marketing e modelos de negócios no Big Data, entre elas: Uber, Amazon, Netflix, Apple, Twitter, BBC, Google, John Deere, Rolls-Royce, Ralph-Lauren e outras (MARR, 2016).

Do mesmo modo, a aplicabilidade versátil do Big Data para expansão e melhoria de negócios, transformou-se em uma tendência entre as indústrias: a manufatura o utiliza para a contínua automação de fábricas e a aplicação no controle de qualidade; bancos, corretoras de seguro e serviços financeiros estão personalizando o tratamento aos clientes com o Big Data (esse recebe o nome de "personology"), detectando fraudes com mais facilidade. Na educação, novas estratégias de ensino e análise de comportamento garantem um novo panorama aos educadores, e também o emprego das tecnologias para evitar o plágio e "cola"; a agricultura, com o compartilhamento de dados sobre o clima, qualidade do solo e comportamento de mudas; o setor público tem investido na combinação de Big Data, Analytics e Internet das Coisas para a criação de Cidades Inteligentes, que podem apresentar melhorias no transporte, na segurança e em uma variedade de serviços oferecidos pela cidade. Outros segmentos que se destacam pelo uso do Big Data são os esportes, os serviços de hospitalidade, serviços profissionais, o setor de energia e a logística (MARR, 2016).

Semelhante à sua adoção nos setores previamente citados, a Biblioteconomia e Ciência da Informação aderiram ao Big Data e passaram a explorar as funcionalidades que poderiam ser desenvolvidas para o benefício de suas áreas.

## **Big Data, Biblioteconomia e Ciência da Informação**

A Biblioteconomia e a Ciência da Informação (C.I.) são áreas do conhecimento correlacionadas que têm em seu cerne e como objeto, a informação. A Biblioteconomia é comumente definida na literatura como uma prática de organização e gestão de bibliotecas e unidades de informação. Todavia, sua conceituação e escopo tornaram-se limitados com os aumentos na produção da informação, no decorrer da história. Em um primeiro momento, o boom informacional causado pela Revolução Industrial, leva a criação da Documentação, que é considerada a

predecessora da Ciência da Informação, quanto a seu surgimento (LECOADIC, 1994; TARGINO, 2006; SAMBAQUI, 1996).

O volume da produção informacional não desacelerou, e após a Segunda Guerra Mundial, foi ainda mais impulsionado. Dessa maneira, a Ciência da Informação tem como um de seus atos introdutórios, a publicação de "As We May Think", na qual Vannevar Bush discute a explosão informacional e os problemas de armazenamento, processamento e tratamento informacionais causados pelo aumento exponencial da produção da informação com o surgimento de novas tecnologias. Desse modo, a C. I. tem como objetivo a "gênese a comunicação da informação" e a interdisciplinaridade (FONSECA, 2007).

Essas duas áreas também se apropriaram do Big Data e suas possibilidades ao aplicá-las em diversas unidades de informação em atividades como: estudos métricos da informação, o compartilhamento de dados e sua comunicação, a curadoria de dados, criação e preservação de datasets, treinamento de usuário, gestão de dados de pesquisa, compartilhamento e comunicação entre bases de bibliotecas públicas, visualização de dados de usuários, geração de metadados de bibliotecas universitárias, criação de acervo de novas fontes, como a rede social Twitter (HOY, 2014; GOLUB; HANSSON, 2017; BALL, 2019).

A tecnologia e sua democratização têm reedificado o campo da Biblioteconomia e da Ciência da Informação e sua atuação sucessivamente, bem como, a era digital e incorporação da internet no cotidiano da sociedade, e agora o fenômeno do Big Data (KHAN; DU, 2018; NONTHACUMJANE, 2011).

Naturalmente, essas circunstâncias afetaram a atuação do bibliotecário, quanto as competências e habilidades necessárias. O volume imoderado no meio acadêmico, especialmente nas Ciências Sociais, levou a criação do Bibliotecário de Dados. A definição dessa função (Bibliotecário de Dados) aparece, em sua maioria, atrelada as atividades de curadoria, preservação e arquivamento de dados de pesquisa, ou informações científicas, em bibliotecas acadêmicas, porém, não estão originalmente voltados ao Big Data (SWAN; BROWN, 2008; TARTAROTTI, DAL'EVEDOVE; FUJITA, 2019; SEMELER, PINTO; ROZADOS, 2019; KHAN; DU, 2018; NONTHACUMJANE, 2011).

Todavia, por mais aplicações e verificações do uso do Big Data no contexto da Biblioteconomia e Ciência da Informação, ainda permanece imprecisa a compreensão das áreas de atuação do profissional bibliotecário a partir do Big Data, o que justifica o presente estudo.

## Metodologia



Para a realização da pesquisa foi utilizada uma abordagem mista, usando técnicas quantitativas e qualitativas. É uma pesquisa bibliográfica de natureza exploratória, no qual foi realizada uma análise de conteúdo e estudos métricos de artigos anexados na base de dados Library and Information Science Abstracts, ou LISA.

O caráter exploratório tem como um de seus principais objetivos o esclarecimento de um conceito e a aproximação do problema. É principalmente aplicada para oferecer uma visão global de determinada ideia, a apresentando por diversas perspectivas, através de técnicas como o levantamento bibliográfico, que será aplicado na pesquisa. As fontes utilizadas para realização da pesquisa foram livros de leitura corrente e impressos diversos para a elaboração da fundamentação teórica, enquanto o corpus da análise foi formado por artigos científicos (GIL, 2002).

A abordagem mista permite a integração dos dados coletados nas análises quantitativas e qualitativas e também é conhecida por "triangulação das fontes de dados". O método misto emprega procedimentos sequenciais, concomitantes e transformadores e pode fornecer melhores resultados para análise. O uso simultâneo dos dois métodos contribui para que o resultado da análise seja minucioso, preenchendo lacunas que poderiam irromper com a aplicação de apenas uma técnica (CRESWELL, 2007; PARANHOS et al. 2016).

O tratamento qualitativo foi discutido através de uma análise de conteúdo. Essa é uma técnica dentro da abordagem exploratória que descreve de forma sistemática aquilo que está sendo comunicado. Quando praticada sob a função heurística, a análise de conteúdo enriquece a pesquisa e "aumenta a propensão para a descoberta". É dividida em 3 etapas: a pré-análise, a exploração do material e o tratamento dos resultados, a inferência e a interpretação (BARDIN, 1995).

A análise quantitativa foi realizada por meio de Estudos Métricos da Informação, técnicas matemáticas e estatísticas de medição para elaborar indicadores sobre a produção científica. Esses estudos são compostos do exame de insumos tais como artigos científicos e produtos gerados, permitindo avaliações em pequenas e grandes escalas. Os resultados desses exames não somente quantificam os dados, mas também podem oferecer insights sobre o estado da arte e suas tendências. Especificamente, as técnicas utilizadas para o estudo realizado foram a bibliométrica e cientométrica, populares por concentrarem-se em áreas bem delimitadas, possuírem "grande potencial de aplicabilidade" e sugerirem objetos de estudo complementares (VITULLO; SANTOS, 2017; NORONHA, 2008; SANTOS; KOBASHI, 2009; PRICE; NORONHA; MARICATO, 2008; MACIAS-CHAPULA, 1998; ARAÚJO, 2015; SILVA; BIANCHI, 2001).

A base de dados utilizada para a coleta de dados foi a Library & Information Science Abstracts (LISA), dada sua relevância e renome dentro da área da Biblioteconomia e da Ciência da Informação. Essa é uma base de dados e ferramenta de indexação internacional dessas áreas.

A LISA resume periódicos de diversos países em mais de 20 idiomas diferentes. Sua cobertura de assunto cobre publicações desde 1969 nos seguintes assuntos: gerenciamento de informações, tecnologia da informação, tecnologias da internet, gerenciamento de conhecimento, biblioteconomia, bibliotecas e arquivos, gerenciamento de biblioteca, tecnologias de bibliotecas, uso e usuários da biblioteca e recuperação de informação online. A base é atualizada mensalmente e utiliza critérios como revisão de pares, diversidade internacional de dados de autoria e citação, tendências atuais na área para escolha de artigos, e também demonstra interesse em publicações de orientação tópica e publicações menores, que lidam com áreas especializadas e de interesse da comunidade (PROQUEST, 2020).

Visto o reconhecimento da base dentro das áreas de Biblioteconomia e Ciência da Informação, pela sua disponibilização de materiais de alta qualidade e autoridade, a LISA foi indicada como melhor opção para a coleta de dados para o cumprir a proposta do presente estudo.

Os artigos considerados para a etapa de análise foram aqueles recuperados através da pesquisa na base, utilizando as palavras-chaves, tanto em português como inglês: "big data", "librarian" e "bibliotecário", em duas combinações com o operador booleano "AND".

Foram considerados apenas artigos científicos indexados na base, com recorte de idioma, selecionando apenas resultados em português, inglês e espanhol. Não houve restrição de anos de publicação, a não ser o ano corrente no qual o estudo é realizado (2020). Com exceção das duplicatas, todos os itens recuperados, foram analisados.

Após a coleta, limpeza e mineração dos dados obtidos na LISA, as análises quantitativas, bibliométrica e cientométrica, foram realizadas através de ferramentas gratuitas para o tratamento de dados, como o Microsoft Excel e o Ucinet, enquanto as análises qualitativas dos periódicos foram obtidas através do Journal Citation Report (CLARIVATE, 2020).

Dessas ferramentas, o Microsoft Excel possibilita a organização e padronização dos dados através de planilhas. O processamento e agrupamento dos dados, realizados conforme os fins das análises, permitem a realização de cálculos e a visualização otimizada de indicadores eleitos para as análises.

A visualização dos indicadores ocorreu por meio de quadros e gráficos criados por meio do Excel, e do Ucinet, um pacote de software capaz de

efetuar análise de dados e que possui uma ferramenta de visualização de NetDraw (BORGATTI; EVERETT; FREEMAN, 2002).

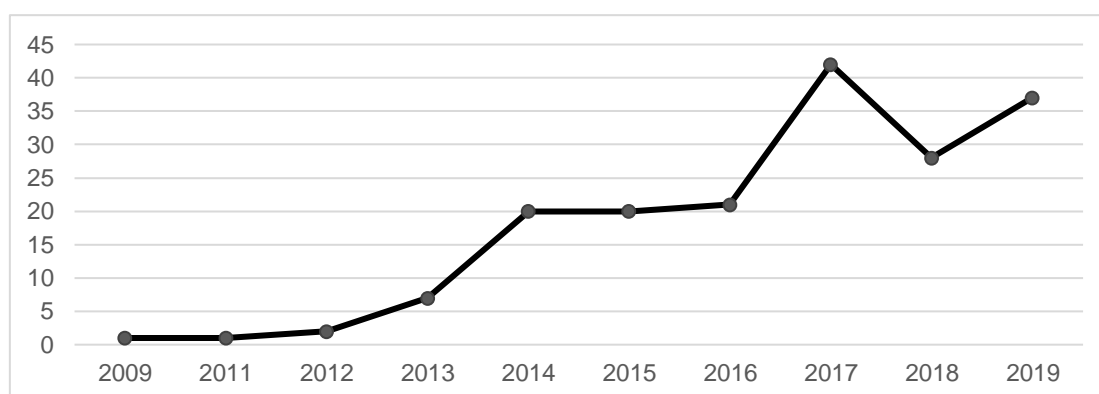
O JCR seleciona, coleta, estrutura e avalia coleções de periódicos que estão na cobertura do Web of Science, oferecendo métricas para a visualização de dados, entre elas o Fator de Impacto, utilizado para mensurar o impacto de periódicos (CLARIVATE, 2020).

## Resultado e Discussão

O corpus da análise é de 184 artigos que obedeceram aos critérios estipulados, na base de dados LISA. Após verificação, identificaram-se 5 artigos duplicados que foram retirados da coletânea. Assim, deu-se início ao exame de 179 artigos recuperados.

Por meio de análise, pode-se perceber que o crescimento temporal dos artigos é gradativo, tendo sua primeira ocorrência em 2009, com apenas um artigo. Os primeiros anos das publicações não possuem muitas intercorrências: o ano de 2010 não registra nenhum artigo, e os dois anos seguintes, registram apenas 3 artigos, um em 2011 e dois em 2012. Já em 2013, percebe-se um aumento significativo, com 7 ocorrências (Gráfico 1).

**Gráfico 1** - Crescimento temporal de artigos publicados sobre Big Data na Ciência da Informação na base LISA (2009 – 2019).

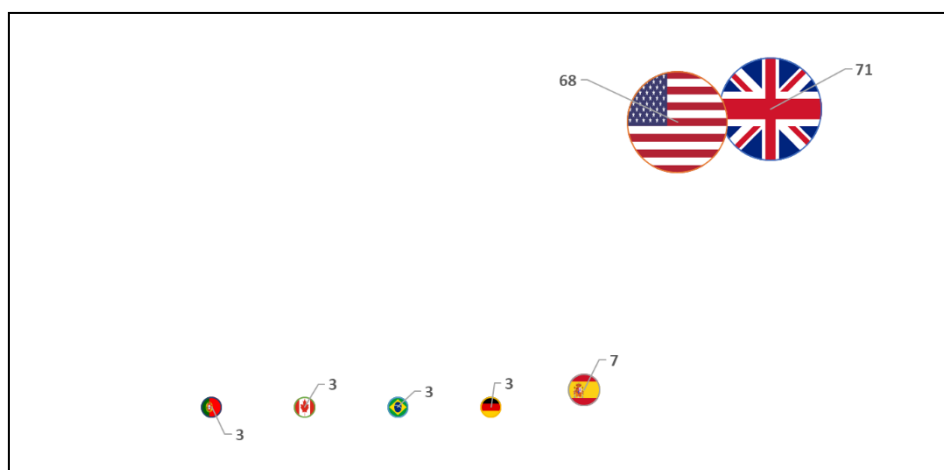


Fonte: Autoria Própria

O número cresce drasticamente em 2013, com 20 publicações e tem seu maior número de artigos indexados na base em 2017, com 42 publicações. Após 2013, embora oscilando, o número de artigos sempre se mantém acima de 20 publicações por ano. Fundamentalmente, a análise temporal permite compreender que os estudos sobre bibliotecários e Big Data continuam a crescer, indicando ainda, grande exploração quantitativa para os próximos anos.

A análise geográfica permitiu identificar 16 países, dentre os quais se destacam dois que detêm os maiores números de publicações: a Inglaterra, com 71 artigos e os Estados Unidos, com 68, representando juntos, 87,97% do total de artigos recuperados (Gráfico 2).

**Gráfico 2** - Distribuição de artigos publicados por país sobre Big Data na Ciência da Informação na base LISA (2009 – 2019)



Fonte: Autoria Própria

Os outros 40 artigos estão divididos entre a Europa, Ásia, América do Sul e 7 não tiveram seu país de origem identificado. Desta maneira, percebe-se que a concentração da produção científica sobre a temática concentra-se entre os Estados Unidos, com 43,04% e a Inglaterra, com 44,94% e o restante 12,03% é distribuído entre países da Europa, América do Sul e Ásia (Tabela 1).

**Tabela 1** - Países com Menor Número de Publicação sobre Big Data na Ciência da Informação na base LISA (2009 – 2019).

País	Quantidade
Espanha	7
Alemanha	3
Canadá	3
Brasil	3
Portugal	3
Cingapura	2
Escócia	2

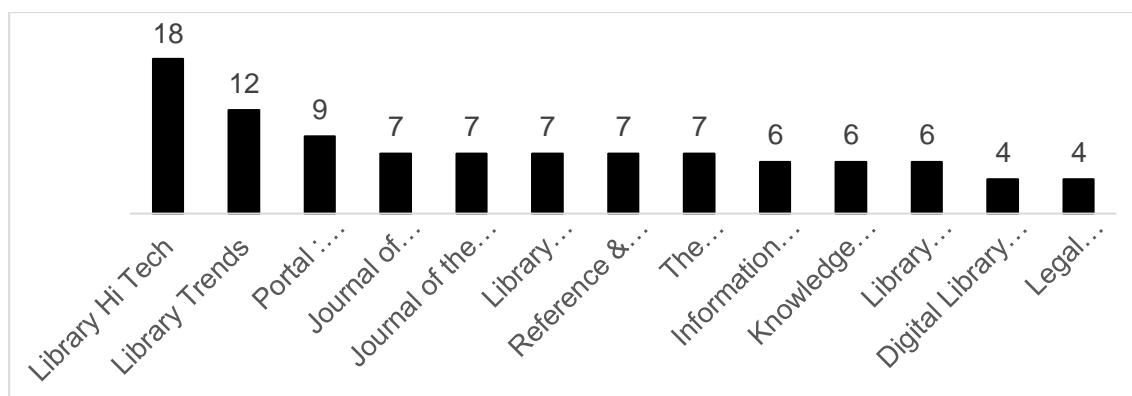
<b>Holanda</b>	2
<b>Índia</b>	2
<b>Itália</b>	2
<b>Taiwan</b>	1
<b>Paquistão</b>	1
<b>Colômbia</b>	1
<b>Japão</b>	1

Fonte: Autoria Própria

Os 179 artigos coletados estão distribuídos entre 66 periódicos, dos quais 50% possuem mais de uma publicação. Evidenciam-se 3 periódicos que possuem o maior número de publicações: *Library Hi Tech*, com 18 estudos publicados; *Library Trends*, com 12 publicações e *Portal: Libraries and The Academy*, com 9.

Ademais, pode-se pontuar outras 10 publicações que se distinguem das demais pelo número de publicação: o *Journal of Documentation*, *Journal of the Medical Library Association*, *Library Management*, *Reference & User Services Quarterly*, *The Electronic Library* com 7 publicações cada; o *Information Technology and Libraries (Online)*, *Knowledge Quest*, *Library Philosophy and Practice* com 6 estudos publicados cada; e por fim, *Digital Library Perspectives* e *Legal Information Management* com 4 publicações (Gráfico 3).

**Gráfico 3** - Títulos por Periódico



Fonte: Autoria Própria

No que se refere a relevância qualitativa dos periódicos avaliados, apenas 24% dos 66 apresentaram o Fator de Impacto. É possível perceber a relação quantitativa e qualitativa, ao comparar a quantidade de publicações e o fator de impacto dos periódicos (Tabela 2).

**Tabela 2** - Periódicos e Fator de Impacto Indexados na Base JCR — 2020

<b>Periódico</b>	<b>Quantidade de Publicações</b>	<b>Fator de Impacto</b>
<b>Library Hi Tech</b>	18	1.256
<b>Library Trends</b>	12	0.627
<b>Portal : Libraries and the Academy</b>	9	1.037
<b>Journal of Documentation</b>	7	1.573
<b>Journal of the Medical Library Association</b>	7	2.420
<b>Reference &amp; User Services Quarterly</b>	7	0.444
<b>Information Technology and Libraries (Online)</b>	6	0.800
<b>Informação &amp; Sociedade</b>	3	0.173
<b>Reference Services Review</b>	3	1.250
<b>Revista Espanola de Documentacion Cientifica</b>	3	0.985
<b>Aslib Journal of Information Management</b>	2	1.702
<b>Program</b>	2	0.868
<b>AI Magazine</b>	1	1.316
<b>Journal of Academic Librarianship</b>	1	1.608
<b>Journal of Librarianship and Information Science</b>	1	1.203
<b>Online Information Review</b>	1	1.928

Fonte: Autoria Própria

Também pode-se discorrer sobre a área desses periódicos: 15 entre 16 estão classificados na área "*Information and Library Science*", pela

*Journal Citation Report*, enquanto apenas 1 dos periódicos (*AI Magazine*), está classificado sob “*Computer Science, Artificial Intelligence*”. A área “*Computer Science, Information Systems*” também aparece como área secundária em 3 periódicos: *Revista Espanola de Documentacion Cientifica*, *Program* e *Online Information Review*, representando 4.5% do total.

Quanto a autoria, foram encontrados 375 autores entre 179 artigos recuperados. Apenas 7 desses autores (cerca de 5%) realizaram mais de uma publicação na área no período analisado, o que demonstra que não há núcleo de autores sobre a temática (Tabela 3).

**Tabela 3** - Autores com Maior Número de Publicações sobre Big Data na Ciência da Informação na base LISA (2009 – 2019).

<b>Autores</b>	<b>Quantidade Publicada</b>
<b>Koltay, Tibor</b>	3
<b>Cox, Andrew M</b>	2
<b>Joo, Soohyung</b>	2
<b>Noh, Younghee</b>	2
<b>Oliphant, Tami</b>	2
<b>Reinhalter, Lauren</b>	2
<b>Schopfel, Joachim</b>	2

Fonte: Autoria Própria

O autor de maior destaque é “Koltay, Tibor” que agrega 3 publicações. Também se sobressaem os autores “Cox, Andrew M”, “Joo, Soohyung”, “Noh, Younghee”, “Oliphant, Tami”, “Reinhalter, Lauren” e “Schopfel, Joachim”, todos com 2 publicações até o ano de 2019.

Em relação as instituições as quais os autores estão afiliados, não há concentração, apenas a Universidade do Sul da Austrália possui 2 trabalhos com representação sobre a temática, cabendo as demais 60 instituições, a distribuição de um trabalho por organização. Percebe-se, portanto, um interesse global pela temática e uma tentativa global de compreender a temática através dos estudos.

É interessante destacar as companhias nas quais esses profissionais, autores dos trabalhos, desempenham suas atividades com a preocupação da atuação do bibliotecário no universo do Big Data. Destacam-se: Altmetric, o Google, o Facebook, Khan Academy, National Center for Education Statistics (NCES), Twitter, Tumblr, Library of Congress, American Library Association (ALA), dentre outras universidades, espalhadas





e temas como "Research Data", "Privacy", "Information Management", "Knowledge Management", "Ethical", "Digitalization" entre outros apresentados na Tabela 4.

**Tabela 4** - Assuntos com maior número de ocorrências sobre Big Data na Ciência da Informação na base LISA (2009 – 2019)

<b>Assuntos</b>	<b>Ocorrências</b>
<i>Education for Librarianship</i>	10
<i>Academic libraries</i>	7
<i>Librarians</i>	4
<i>Big Data</i>	4
<i>Research Data</i>	3
<i>Metadata</i>	3
<i>Privacy</i>	2
<i>Library services</i>	2
<i>Knowledge Management</i>	2
<i>Information management</i>	2
<i>Information management</i>	2
<i>Ethical</i>	2
<i>Digitalization</i>	2
<i>Data-intensive research</i>	2
<i>Curriculum</i>	2
<i>Competencies</i>	2
<i>Artificial intelligence</i>	2

Fonte: Autoria Própria

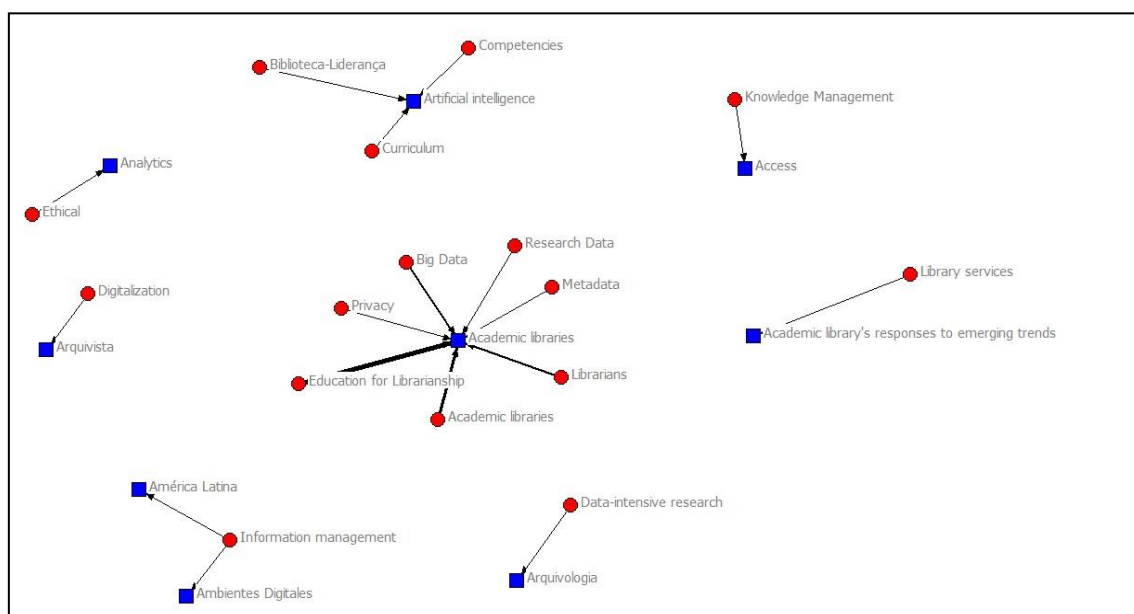
Essas publicações, em sua maioria, discutem a educação e a formação dos bibliotecários no contexto do Big Data, levando em consideração "Dados de Pesquisa", "Metadados", "Privacidade", "Serviços de Biblioteca". Também pode-se destacar os assuntos: "Ética", "Digitalização", "Gestão do Conhecimento" e "Gestão da Informação".

A concentração de aproximadamente 30% do total de assuntos relacionados a educação demonstra a preocupação da área com as competências necessárias para esse novo momento do profissional

bibliotecário e também, o interesse com a infraestrutura tecnológica a serem utilizadas por esses profissionais ou pelas unidades nas quais eles atuam.

Além do mais, deve-se destacar a relação de similaridade/proximidade entre esses assuntos (Gráfico 4).

**Gráfico 4** - Relações entre Assuntos sobre Big Data na Ciência da Informação na base LISA (2009 – 2019)



Fonte: Autoria Própria

O assunto “*Academic Libraries*” concentra o maior número de assuntos relacionados, associando-se com outros 7 termos, entre eles, o cerne da pesquisa, “*Big Data*” e também “*Privacy*”, “*Metadata*” e “*Research Data*”, assim como o agrupamento “*Academic library’s responses to emerging trends*” atrelado a “*Library Services*” o que demonstra o protagonismo das bibliotecas acadêmicas quanto a atuação de bibliotecários com o advento do Big Data.

Também pode-se salientar o agrupamento “*Artificial Intelligence*” representando 1,00% do total, cercado por “*Biblioteca-Liderança*” (0,50%), “*Competencies*” (1,00%) e “*Curriculum*” (1,00%), que ressaltam o foco dos estudos nas competências e habilidades necessárias dos bibliotecários para lidarem com o Big Data. O agrupamento “*Analytics*” (0,50%) vem acompanhando do assunto “*Ethical*” (1,00%) se mostrando um tópico em ascensão na elaboração e uso de tecnologias disruptivas.

Em outros clusters, encontram-se agrupamentos entre "Digitalization" (1,00% do total) e "Arquivista" (0,50%); "Information Management" (2,00%) e "Ambientes Digitales" (0,50%); "Access" (0,50%) e "Knowledge Management" (1,00%); e "Arquivologia" (0,50%) e "Data-Intensive Research" (1,00%), que podem demonstrar possíveis áreas de atuação para bibliotecários, associadas ao Big Data.

A temática da atuação do bibliotecário dentro do contexto do Big Data foi analisada a partir de artigos publicados em revistas indexadas na base LISA e evidencia que está em ascensão, distribuída principalmente entre Estados Unidos, Canadá e Inglaterra. Quanto as temáticas, observou-se foco em relação à formação do bibliotecário e atualização de seu currículo, envolvendo principalmente a biblioteca acadêmica, gestão do conhecimento e da informação, digitalização e discussões sobre a privacidade e ética dentro desse contexto.

## Conclusão

Após a presente análise, verifica-se que os objetivos propostos para o vigente estudo, foram alcançados. Por meio de análises quali-quantitativas, foi possível compreender o estado da produção científica em relação a atuação do profissional bibliotecário em face ao advento do Big Data, fenômeno da Indústria 4.0.

Mediante aos procedimentos metodológicos escolhidos, os 184 artigos recuperados oferecem insights interessantes sobre a temática, a começar por sua progressão temporal: nota-se a presença do tema dentro das áreas da Biblioteconomia e da Ciência da Informação desde 2009, com apenas 1 ocorrência representando 0,61% do total, porém é apenas entre 2012-2013 que o número dessas contribuições aumenta de maneira considerável, em 2012 com aumento de 1,23% e 2013, com 4,29%, e mantém-se em uma crescente constante, indicando o aumento do interesse na temática pela área, ainda que tardio quando comparado com outros campos e indústrias, chegando a picos de 25,77% no ano de 2017 e 22,70% no ano de 2019.

Quanto ao comportamento geográfico, percebe-se que os trabalhos estão majoritariamente distribuídos entre dois países: Inglaterra, com 71 publicações, que representam 44,94% do total de artigos e Estados Unidos, com 68 ocorrências ou 43,04%, formando juntos 87,97% do total recuperado. Essas publicações distribuem-se entre instituições acadêmicas e da área, e também empresas de diversos segmentos, entre eles: mídias sociais, educação e altimetria. Outras incidências de publicações ocorrem esporadicamente na América Latina e do Norte, além da Europa e Ásia.

Como um todo, nos indicadores escolhidos para realização das análises quantitativas, percebe-se que a dispersão é praticamente unânime e os núcleos ainda estão em formação, como demonstram os números relacionados aos autores: apenas 5% contém mais de uma publicação no período avaliado.

É incontestável o impacto do Big Data e sua aplicabilidade nos mais diversos setores industriais e áreas do conhecimento, é importante pontuar o notório “atraso” da Biblioteconomia e da Ciência da Informação em incorporar essa tecnologia e considerar suas repercussões para as áreas e principalmente, os efeitos na atuação dos bibliotecários. A ideia de grandes volumes de dados não é novidade para esses profissionais, todavia, não é possível apontar com exatidão as competências e as atividades práticas, assim, são necessárias novas abordagens investigativas sobre as competências e atividades que devem ser desenvolvidas.

Ainda assim, a análise qualitativa dos dados coletados mostra que mesmo que tardio, a Biblioteconomia e a Ciência da Informação se apropriam dos benefícios do Big Data, porém estudos que abordam a atuação do bibliotecário nessa área ainda se encontram em dispersão. Observa-se um núcleo de apenas 26% da concentração de estudos e esses vão de encontro com a ideia de necessidade de atualização da formação de um profissional antenado nesse cenário, o que indica preocupação com as competências e as habilidades necessárias visando sua atuação dentro dessa temática.

Percebe-se que o bibliotecário não está alheio a esse tipo de tecnologia disruptiva e pode atuar nesse contexto, porém, visto que a temática é recente em termos científicos, a preocupação inicial é compreender o Big Data dentro dos limites da Biblioteconomia, para assim, atualizar a formação do bibliotecário e capacitá-lo para trabalhar de maneira eficiente com essa inovação e sua interdisciplinaridade. Essa atualização tem que ser muito bem concatenada entre a área porque do contrário, reflexões começam a ser levadas ainda não solidificadas dentro da área para outros contextos e nichos de atuação.

Embora a dispersão em relação a produção científica não seja positiva, ela é esperada, visto o quão recente a temática tem sido abordada pela área. Todavia, também deve-se levar em conta que a dispersão em períodos iniciais é preferível a manutenção da dispersão, uma vez que a temática tenha se consolidado dentro desse espaço.

Outrossim, também foram identificados potenciais áreas nas quais as atividades dos bibliotecários podem ser desenvolvidas, como questões envolvendo privacidade, inteligência artificial e ética, que tem sido cada vez

mais discutidas considerando a popularização do uso de dados e que poderiam desfrutar de inputs feitos por bibliotecários.

Publicações acerca da digitalização e gestão do conhecimento e da informação sinalizam subcampos dentro da Biblioteconomia, que tem saído a frente da investigação científica quanto a atuação do bibliotecário. Também deve-se assinalar que a incidência de estudos relacionados a bibliotecas acadêmicas é a de maior proeminência, demonstrando o protagonismo dessas instituições em relação a temática.

Constata-se, através da pesquisa realizada, que a atuação de bibliotecários no contexto do Big Data está acontecendo, ainda que de maneira modesta, causada principalmente pela incorporação tardia da tecnologia nessa área do conhecimento. Verificou-se que os estudos, em um período de 10 anos, concentram-se principalmente em compreender o Big Data, para que a formação de bibliotecários possa ser atualizada conforme as habilidades e competências requeridas para a manipulação adequada desse fenômeno e mostrando-se, assim, como uma área com muitas oportunidades em potencial a serem exploradas por bibliotecários.

## Referências

ACKOFF, R.L. From data to wisdom. **Journal of applied systems analysis**, [s. l.], v. 16, p. 3-9, 1989.

ARAÚJO, R.F. Estudos métricos da informação na web e o papel dos profissionais da informação. **Bibl. Univ.**, Belo Horizonte, v. 2, p. 45-64, fev. 2015.

ARCILA-CALDERÓN, C.; BARBOSA-CARO, E.; CABEZUELO-LORENZO, F. Técnicas big data: análisis de textos a gran escala para la investigación científica y periodística. **El Profesional de la Información**, Barcelona, v. 25, n. 4, p. 119-159, 2016.

BALL, R. Big Data and their impact on libraries. **American Journal of Information Science and Technology**, [s. l.], v. 3, n. 1, p. 1-9, fev. 2019.

BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, 1995.

BORGATTI, S.P.; EVERETT, M.G.; FREEMAN, L.C. 2002. **Ucinet for Windows**: Software for social network analysis. Harvard, MA: Analytic Technologies.

CASTELLS, M. **A sociedade em rede**. 6. ed. rev. São Paulo: Paz e Terra, 2002. 698 p.

CLARIVATE Analytics. About Journal Citation Reports. *In: Journal Citation Reports Help*. [s. l.], 2020. Disponível em: <http://jcr.help-clarivate.ez128.periodicos.capes.gov.br/Content/home.htm>. Acesso em: 18 jun. 2020.

CRESWELL, J. W. **Projeto de pesquisa: Métodos Qualitativo, Quantitativo e Misto**. 2. ed. Porto Alegre: Artmed, 2007. 248 p.

DAVENPORT, T. H. **Ecologia da informação**. Brasil: Siciliano, 1998. 316 p.

DE MAURO, A.; GRECO, M.; GRIMALDI, M. "A formal definition of Big Data based on its essential features", **Library Review**, v. 65 n. 3, p. 122-135, 2016.

DUTRA, M.L.; MACEDO, D.; DYLLON J. Curadoria Digital: Proposta de um modelo para curadoria digital em ambientes big data baseado numa abordagem semi-automática para a seleção de objetos. **Informação & informação**, Londrina, v. 21, n. 2, p. 143, 2016.

FONSECA, E.N. **Introdução à Biblioteconomia**. 2. ed. Brasília: Briquet de Lemos, 2007. 152 p.

GANDOMI, A.; HAIDER, M. Beyond the hype: big data concepts, methods, and analytics. **International Journal of Information Management**, v. 35, n. 2, p. 137-144, 2015.

GARFINKEL, S.; COX, D. Finding and archiving the internet footprint. *In: First digital lives research conference: personal digital archives for the 21st Century*, Londres, 2009.

GIL, A.C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2002.

GOLUB, K. HANSSON, J. (Big) Data in Library and information science: a brief overview of some important problem areas. **Journal of Universal Computer Science**, [s. l.], v. 23, n. 11, p. 1099-1108, 2017.

HOY, M.B. Big Data: An introduction for librarians. **Medical reference services quarterly**, v. 33, n. 3, p. 320-326, 2014.

International Telecommunication Union. **Measuring the Information Society Report**. Geneva; International Telecommunication Union, 2018.

KHAN, H. R.; DU, Y. What is a data librarian?: a content analysis of job advertisements for data librarians in the United States Academic Libraries. In: IFLA WLIC, 2018, Kuala Lumpur. **84<sup>th</sup> IFLA General Conference and Assembly**. Kuala Lumpur, 2018.

LANEY, D. 3D data management: controlling data volume, velocity and variety. **Application Delivery Strategies**, Stamford, 2001.

LATTS, L. **The age of big data and the power of watson**. 2016. 20 slides.

LE COADIC, Y. **A Ciência da Informação**. Brasília: Briquet de Lemos, 1994. 120 p.

LI, G.; HOU, Y.; WU, A. Fourth industrial revolution: technological drivers, impacts and coping methods. **Chinese Geographical Science**, [S. l.], v. 27, p. 626-637, 18 jun. 2020.

LIBRARY AND INFORMATION SCIENCE ABSTRACTS: About. [S. l.], mai. 2020. Disponível em: <http://proquest.libguides.com/lisa>. Acesso em: 18 jun. 2020.

MACIAS-CHAPULA, C. A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, Brasília, v. 27, ed. 2, p. 134-140, maio/ago 1998.

MARR, B. **Beyond The Big Data Buzz**: How data is disrupting business in every industry in the world. [S. l.]: Kogan Page, 19 p.

MARR, B. **Big Data in Practice**: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results. Grã-Bretanha: Wiley, 2016. 308 p.

MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big Data**: A revolution that will transform how we live, work and think. New York: Houghton Mifflin Harcourt, 2013. 103 p.

MCCANDLESS, D. et al. **Information is Beautiful Studio for BBC Future**. Disponível em: <<https://www.bbc.com/future/article/20130621-byte-sized-guide-to-data-storage>>. Acesso em: 14 jan. 2020.

MING, Z.; WIDÉN, G. Understanding big data in librarianship. **Journal of Librarianship and Information Science**, London, v. 51, n. 2, p. 561-576, 2019.

MINISTERIO DA INDUSTRIA COMERCIO E SERVIÇO. **Agenda brasileira para a Indústria 4.0**. [S. l.], 18 jun. 2020. Disponível em: <http://www.industria40.gov.br/>. Acesso em: 20 jan. 2020.

NETO, J. A. R. **Big data para executivos e profissionais de mercado**. 2. ed. [s. l.], 2019. 552 p.

NONTHACUMJANE, P. Key skills and competencies of a new generation of LIS professionals. **IFLA Journal**, The Hague, v. 37, n. 4, p. 280-288, 2011.

NORONHA, D. P.; MARICATO, J. M. Estudos métricos da informação: primeiras aproximações. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, p. 116-128, abr. 2008. ISSN 1518-2924. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2008v13nesp1p116>. Acesso em: 18 jun. 2020.

PARANHOS, R. et al. Uma introdução aos métodos mistos. **Sociologias**, Porto Alegre, v. 18, n. 42, p. 384-411, ago. 2016. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1517-45222016000200384&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1517-45222016000200384&lng=en&nrm=iso). Acesso em: 27 mai. 2020.

PATGIRI, R.; AHMED, A. Big Data: the v's of the game changer paradigm. **The Institute of Electrical and Electronics Engineers, Inc.(IEEE) Conference Proceedings**, Piscataway, p. 17-24, 2016.

PINHEIRO, P. *et al.* Industry 4.0 and Industrial Revolutions: an Assessment based on Complexity. **FME Transactions**, [s. l.], v. 47, p. 831-840, 18 jun. 2020.

PRICE, D. J. de S. **Little science, big science... and beyond**. New York: Columbia University Press, v. 301, 1986.

RUSSO, M. **Fundamentos de biblioteconomia e Ciência da Informação**. Rio de Janeiro: e-papers, 2010. 177 p.

SAMBAQUI, L. de Q. Da Biblioteconomia à Informática. **Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun.1996.

SANTOS, R. F.; VITULLO, N. A. V. A importância dos estudos métricos da informação na construção de indicadores para a formação e



desenvolvimento de coleções. **Revista Informação na Sociedade Contemporânea**, v. 1, n. 2, p. 1-18, abr. 2017.

SANTOS, R. N. M.; KOBASHI, N. Y. **Bibliometria, cientometria, infometria: conceitos e aplicações. tendências da pesquisa brasileira em ciência da informação**, 2009.

SCHWAB, K. **The Fourth Industrial Revolution**. Geneva, Suíça: World Economic Forum, 2016. 172 p.

SILVA, J. A.; BIANCHI, M. L. P. Cientometria: a métrica da ciência. **Paidéia (Ribeirão Preto)**, Ribeirão Preto, v. 11, n. 21, p. 5-10, 2001. Disponível em:

[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-863X2001000200002&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-863X2001000200002&lng=en&nrm=iso). Acessado em: 6 abr. 2020.

SOGETI, L. **The fourth industrial revolution: things to tighten the link between it and ot sogeti labs**, 2014.