



UNIVERSIDADE DE BRASÍLIA (UNB)
FACULDADE DE CIÊNCIA DA INFORMAÇÃO (FCI)
CURSO DE GRADUAÇÃO EM BIBLIOTECONOMIA

THAIS KETLYN DA SILVA MEDEIROS

**ANÁLISE DE DADOS ABERTOS PARA SUBSIDIAR DECISÕES UTILIZANDO
DADOS DO ENEM: EVIDÊNCIA DO PROFISSIONAL DE BIBLIOTECONOMIA NA
CIÊNCIA DE DADOS**

BRASÍLIA
2023

THAIS KETLYN DA SILVA MEDEIROS

ANÁLISE DE DADOS ABERTOS PARA SUBSIDIAR DECISÕES UTILIZANDO
DADOS DO ENEM: EVIDÊNCIA DO PROFISSIONAL DE BIBLIOTECONOMIA NA
CIÊNCIA DE DADOS

Monografia submetida ao Curso de Graduação em
Biblioteconomia da Universidade de Brasília, como
requisito parcial para obtenção do Título de
Bacharel em Biblioteconomia.

Orientador: Prof.Dr. Márcio de Carvalho Victorino

BRASÍLIA
2023

Dados Internacionais de Catalogação na Publicação (CIP)

M488a Medeiros, Thais Ketlyn da Silva

Análise de dados abertos para subsidiar decisões utilizando dados do Enem: evidência do profissional de Biblioteconomia na ciência de dados / Thais Ketlyn da Silva Medeiros; orientador Márcio de Carvalho Victorino. – Brasília, 2023.

83 f. : il. color.

Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) – Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2023.

1. Dados abertos. 2. Análise de dados. 3. Ciência da informação. 4. Organização da informação. 5. Enem. I. Victorino, Márcio de Carvalho, orient. II. Título.

FOLHA DE APROVAÇÃO

Título: ANÁLISE DE DADOS ABERTOS PARA SUBSIDIAR DECISÕES UTILIZANDO DADOS DO ENEM: EVIDÊNCIA DO PROFISSIONAL DE BIBLIOTECONOMIA NA CIÊNCIA DE DADOS.

Autor(a): Thais Ketlyn da Silva Medeiros

Monografia apresentada em **20 de Novembro de 2023** à Faculdade de Ciência da Informação da Universidade de Brasília, como parte dos requisitos para obtenção do grau de Bacharel em Biblioteconomia.

Orientador(a) (FCI/UnB): Dr. Márcio de Carvalho Victorino

Membro Interno (FCI/UnB): Dra. Fernanda Farinelli

Membro Interno (STI/UnB): Me. Rodrigo da Fonseca Silveira



Documento assinado eletronicamente por **Marcio de Carvalho Victorino, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 04/12/2023, às 12:39, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Rodrigo da Fonseca Silveira, Analista de Tecnologia da Informação da Secretaria de Tecnologia da Informação**, em 04/12/2023, às 16:00, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Fernanda Farinelli, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 05/12/2023, às 12:04, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



A autenticidade deste documento pode ser conferida no site

http://sei.unb.br/sei/controlador_externo.php?

[acao=documento_conferir&id_orgao_acesso_externo=0](http://sei.unb.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **10634439** e o código CRC **BC5B9F2E**.

DEDICATÓRIA

Dedico este trabalho à minha melhor amiga Juliana Siqueira (em memória) que infelizmente faleceu durante o final da minha graduação.

Juju, a vida não é tão bonita sem você aqui, as aulas não tem mais a mesma leveza sem a sua presença e os trabalhos em grupo nunca mais serão os mesmos. Os dias passam e eu ainda sinto a sua falta, e eu vou sentir ainda pro resto da minha vida. Quando eu comecei o curso, eu já tinha certeza de tudo o que eu queria fazer, mas foi com você e com nosso quarteto que eu encontrei o meu lugar. Eu sou grata pelo privilégio de ter dividido uma parte da graduação com você, pelos lugares que trabalhamos juntas, e pela vida que dividimos. Você se foi cedo demais, e quanto a isso eu não posso fazer nada, mas dentro de mim sempre haverá um pedacinho seu. No dia da minha formatura, eu sei que você estará lá e você também se formará comigo. Com você eu aprendi que as pessoas e os relacionamentos são como livros. Conhecemos, aprendemos a entender, aprendemos a amar e os mantemos por anos na nossa estante. Vez ou outra releemos, sorrindo ou chorando, lembrando das sensações como se fosse a primeira vez. E após algum tempo, ele precisa ir. Seja para conhecer novos leitores, ou para um lugar destinado ao seu fim. Eu entendi que não se pode deixar um livro na nossa estante para sempre. Em algum momento o ciclo chega ao fim. Assim como os papéis, as pessoas não vivem para sempre.

Nesse momento, eu sou a última do nosso quarteto a se formar. Nós quatro nos unimos muito mais graças a forma como você nos ensinou a amar. Daqui desse lugar onde você nos deixou, só ficou a saudade e as lembranças, Ranganathan tinha razão ao dizer que para cada livro havia o seu leitor, o livro da nossa vida não terá novas linhas com o seu nome, mas nós vamos continuar escrevendo e deixando nele o amor que você deixou.

Nenhum espaço de tempo seria suficiente para eu dizer o quanto eu te amo e sinto sua falta, nesse momento o “pra sempre” cabe melhor do que eu gostaria.

Eu vou te amar pra sempre. E pra sempre eu sentirei a sua falta.

AGRADECIMENTOS

Agradeço primeiramente à minha mãe, que sempre confiou no meu potencial, me ensinou e me guiou pra eu chegar onde cheguei e apoiou todas as minhas decisões, mesmo quando ela tinha certeza que eram loucuras, eu sou totalmente cada pedacinho seu. Eu jamais estaria na vida acadêmica e conquistando tanta coisa se não fosse o seu apoio e toda a sua luta para nos colocar no caminho da sabedoria e da luta. Você é a mulher mais inspiradora do mundo, e mesmo que seu caminho até agora não tenha sido fácil, quero que sempre esteja orgulhosa de tudo o que construiu. Quero que se lembre que o mundo é seu e que tudo o que quiser, será seu, pois foi assim que nos ensinou. Ao meu companheiro Rafael, que sempre foi o meu pilar, que me apoiou das formas que lhe cabiam, sonhou comigo os meus sonhos mais malucos e por sempre acreditar em mim, mesmo quando eu não acreditava. Obrigada por me mostrar como é bom amar e ser amada com leveza e sinceridade, eu te amo além dos seus defeitos e não apesar deles. Mesmo que você brigue comigo por eu colocar 7 matérias e umas 3 pesquisas na minha grade, eu sei que eu não estarei nem perto de onde estou hoje se não fosse pelo seu apoio e por me puxar para a realidade quando necessário. Com você eu aprendi que os caminhos mais bonitos são os que fazemos de mãos dadas. Amo você.

Às minhas amigas Marilete e Graziela que acolheram meus surtos durante as madrugadas e foram meu apoio acadêmico em muitos sentidos. Vocês me inspiram.

Agradeço imensamente às minhas melhores amigas que fiz na graduação Camila, Letícia e Carem por serem sempre minhas companheiras nos melhores e nos piores momentos, eu amo vocês de uma forma que eu não sei descrever, todos os nossos momentos são especiais. Eu amo amar vocês.

Agradeço a Alessandra, minha “TiaDrasta”, que me ajudou muito a crescer academicamente, eu aprendi muito com você e você se tornou uma inspiração para mim enquanto mulher pesquisadora.

Também não posso deixar de destacar o apoio da minha sogra Patrícia, que mesmo com nossas diferenças, eu sou muito grata pelo seu cuidado como sogra e como avó. Seu apoio tem sido fundamental para que eu possa trabalhar e estudar tendo a certeza de que meu trabalho mais precioso, que é o meu filho, está sendo cuidado como mais ninguém faria.

Agradeço especialmente a Josenir, que sem dúvida é uma mulher admirável e muito forte, sem nenhuma romantização da sua luta, você me inspira muito e quando eu crescer quero ser como você. Vocês são de verdade uma segunda família e não tenho como agradecer por todo o suporte durante todos os anos da graduação, principalmente sendo a minha rede de apoio enquanto mãe estudante, sem vocês eu não teria chegado até aqui.

Às minhas melhores amigas Wedlaine, Mikaela, Bia e a minha irmã Thayane que sempre acreditaram que eu seria o que eu quisesse ser. A vida adulta separa nossos corpos, mas o meu coração sempre será de vocês.

Agradeço à minha ex-Chefe Patrícia Coelho do Senado Federal que me fez entender o meu lugar na Biblioteconomia.

Este trabalho eu dedico à minha melhor amiga Juliana, *in memoriam*, começamos juntas essa graduação, mas não vamos terminá-la juntas. A saudade que sinto de você não cabe dentro de mim, obrigada por todos os trabalhos que fizemos juntas, pelas risadas e pela maneira leve (ou não) que levamos a nossa graduação. Nosso quarteto sempre foi o meu suporte e minha base nesse mundo que é a UnB. As bibliotecas perdem muito por não ter você por lá e eu também. Sempre vou te amar.

Leona Rezende e Geovanna Melone, obrigada pelo apoio nas minhas empreitadas como *BiblioMei*, nessa vida eu sou muito grata pelo apoio e carinho de vocês duas. Vocês são as melhores. À minha amiga Lais Cirilo por ser alguém tão especial, obrigada por ser minha alma gêmea acadêmica.

Agradeço ao meu orientador Márcio Victorino por me incentivar e me fazer crescer com este trabalho, e todas as amigades que fiz durante a graduação, me apoiaram de alguma maneira e eu sou grata a todos.

Agradeço imensamente aquela que se tornou minha mentora, Shirley Carvalhêdo Franco, que mesmo sabendo que eu sou 60% *biblio*, tem me apoiado de uma forma como eu jamais esperaria de alguém que conheci como aluna. Meu carinho por você ultrapassa todos os muros da universidade, você é alguém que quero manter na minha vida pra sempre. Você me inspira e espero que os meus 40% *aquivo* te traga o orgulho e as premiações que nós merecemos. Muito obrigada.

Às minhas ex diretoras da minha escola de ensino médio, Ghisa e Leisa, eu sou alguém que jamais vai esquecer o que fizeram por mim. O caminho que eu

trilho agora, foi conduzido pelos pequenos cascalhos que vocês usaram para guiar o meu caminho. Todas as palavras são poucas para definir o meu carinho e a minha gratidão, vocês foram luz.

Por último, mas extremamente importante, eu dedico esse trabalho também ao meu filho, Oliver. Muitas das decisões que fiz na vida foram cheias de tropeços, erros e acertos. A escolha de ter você, mesmo no início da minha graduação, foi uma decisão da qual eu nunca me arrependi. Com você eu entendi que todo mundo tem o seu tempo, e mesmo que eu tenha demorado mais tempo que outras pessoas para chegar até aqui, o meu caminho foi único e especial. Ao mesmo tempo que quero me desculpar pela ausência em certos momentos, eu sei que não só pelo seu futuro financeiro, mas que o caminho que estou trilhando agora, será de extrema importância na sua vida quando for a sua vez de tomar decisões e trilhar o seu caminho. Não é sobre seguir os passos de alguém, mas sobre saber que na trilha em que pisará, os meus pés já afagaram o seu caminho. Nesse mundo, meu desejo é que um dia você tenha alguém que te ame como eu te amo.

“A infância não vai do nascimento até certa idade. A certa altura, a criança está crescida e descarta as coisas infantis. Infância é o reino onde ninguém morre.”
- Bella Swan.

RESUMO

Este trabalho de pesquisa faz uma análise dos microdados do Enem através da coleta e tratamento de microdados, por meio de eliminação de dados irrelevantes e preservação dos dados escolhidos como relevantes dentro do escopo de idade, gênero, localização e faixa de renda. Após análise da literatura dos autores da área para uma melhor compreensão dos termos expostos, este trabalho tem como proposta apresentar ao usuário diferentes interpretações de informação a partir da análise dos dados disponibilizados pelo Portal da Transparência. Nesse sentido tem como proposta contextualizar Dados, Microdados entre suas diferenças e semelhanças; assim como demonstrar todo o caminho percorrido pela Ciência da Informação até chegar ao que ela é hoje; e demonstrar a importância do profissional da Ciência da Informação formado em Biblioteconomia. Os dados foram tratados em planilha Excel e analisados através do *Software Power BI* para a criação de gráficos e tabelas dinâmicas por meio de análises estatísticas. Em conclusão, foi possível verificar que a partir da análise de dados o bibliotecário possui conhecimento necessário para embasar e justificar estudos em diversas áreas de pesquisa.

Palavras-chave: Dados abertos; Análise de dados; Ciência da informação; Organização da informação, Enem

ABSTRACT

This research does a micro data analysis of ENEM (Exame Nacional do Ensino Médio) through collection and management of micro data, by eliminating not necessary data and preservation of data picked as relevant in an age, gender, geolocation and income bracket ambit. After analyzing the literature of authors in this area for a better understanding of the exposed terms, this work aims to present to the user with different interpretations of information based on the analysis of the data made available by the Transparency Portal. In this regard, aims to contextualize Data, Micro data, through their differences and similarities; as much as to demonstrate all the way made by Information Science until it becomes what it is today; and show the importance of an Information Science professional, graduated in Librarianship. The data was processed in an Excel sheet and analyzed through the Power BI Software to create dynamic sheets and graphics, by a statistical analysis. In conclusion, it was possible to verify that, from data analysis, the librarian has the required knowledge to support and justify study in different research fields.

Key words: Open Data; Data Analysis; Information Science; Information Organization; Enem

LISTA DE FIGURAS

Figura 1 - Sistema de Recuperação da Informação.....	21
Figura 2 - Cadeia Documentária - visão sistêmica.....	22
Figura 3 - Etapas de uma SRI	22
Figura 4 - O ciclo do dado.....	26
Figura 5 - Hierarquia das estrelas.....	31
Figura 6 - Diagrama de Venn.....	35
Figura 7 - Pergunta 1.....	41
Figura 8 - Pergunta 2.....	41
Figura 9 - Pergunta 5.....	42
Figura 10 - Pergunta 6.....	42
Figura 11 - Pergunta 24.....	43
Figura 12 - Pergunta 25.....	43
Figura 13 - Coluna A de inscritos.....	46
Figura 14 - Coluna B de faixa etária.....	46
Figura 15 - Coluna C de gênero.....	47
Figura 16 - Coluna D de estado civil.....	47
Figura 17 - Coluna E de declaração racial.....	48
Figura 18 - Coluna F de nacionalidade	48
Figura 19 - Coluna G de situação escolar.....	49
Figura 20 - Coluna H de ano de conclusão.....	49
Figura 21 - Coluna I de tipo de escola.....	50
Figura 22 - Coluna J de município de residência.....	50
Figura 23 - Coluna K de Unidade Federativa.....	51
Figura 24 - Coluna L presentes dia 1.....	51
Figura 25 - Coluna M presentes dia 2.....	51

Figura 26 - Coluna N de alunos presentes na prova subjetiva.....	52
Figura 27 - Coluna O de alunos eliminados.....	52

LISTA DE GRÁFICOS

Gráfico 1 - Dificuldade de infraestrutura.....	44
Gráfico 2 - Tipo de dificuldade de infraestrutura.....	45
Gráfico 3 - Mapa de inscritos por região.....	54
Gráfico 4 - Mapa de homens e mulheres.....	57

LISTA DE TABELAS

Tabela 1 - Inscritos no Norte.....	54
Tabela 2 - Inscritos no Nordeste.....	55
Tabela 3 - Inscritos no Sudeste.....	55
Tabela 4 - Inscritos no Centro-Oeste.....	56
Tabela 5 - Inscritos no Sul.....	56
Tabela 6 - Tabela de declaração de raça por região.....	58
Tabela 7 - Tabela de declaração de uso de nome social.....	59

LISTA DE SIGLAS

ADI - *American Documentation Institute*
ASIS - *American Society for Information Science*
ASIST - *American Society for Information Science Technology*
BI - *Business Intelligence*
CC - Creative Commons
CDU - Classificação Decimal Universal
CI - Ciência da Informação
CGU - Controladoria Geral da União
CSV - *Comma Separated Values* ou Valores separados por vírgula
Enade - Exame Nacional de Desempenho dos Estudantes
ENEM - Exame Nacional do Ensino Médio
e-SIC - Sistema de Informação ao Cidadão
IBBD - Brasileiro de Bibliografia e Documentação
IBGE - Instituto Brasileiro de Geografia e Estatística
IBICT - Instituto Brasileiro de Informação em Ciência e Tecnologia
IIB - Instituto Internacional de Bibliografia
IID - Instituto Internacional de Documentação
INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
ISKO - *International Society for Knowledge Organization*
FID - Federação Internacional de Documentação
LAI - Lei de Acesso à Informação
MEC - Ministério da Educação
NASA - *National Aeronautics and Space Administration*
OC - Organização do Conhecimento
OI - Organização da Informação
OGP - *Open Government Partnership*
QR - *Quick Response* ou Resposta Rápida
SISU - Sistema de Seleção Unificada
SRI's - Sistema de Recuperação da Informação
TCU - Tribunal de Contas da União
TI - Tecnologia da Informação
UF - Unidade Federativa

URI - *Uniform Resource Identifier*, ou Identificador Uniforme de Recursos

URL - Uniform Resource Locator ou Localizador Uniforme de Recursos

W3C - *World Wide Web Consortium*

XLS - *Extensible Style Language*

SUMÁRIO

1 INTRODUÇÃO.....	1
2 PROBLEMA DE PESQUISA.....	3
2.1 Justificativa.....	3
3 OBJETIVOS.....	4
3.1 Objetivo Geral.....	4
3.2 Objetivos específicos.....	4
4 REVISÃO DE LITERATURA.....	5
4.1 A Ciência da Informação.....	5
4.2 Organização da Informação.....	13
4.3 Análise da Informação.....	24
4.4 Ciência de Dados e Dados Abertos.....	29
5 METODOLOGIA.....	38
6 ESTUDO DE CASO.....	40
7 CONSIDERAÇÕES FINAIS.....	61
REFERÊNCIAS.....	64

1 INTRODUÇÃO

Em 2011, foi promulgada a Lei de Acesso à Informação (LAI) sob o número 12.527/2011. Ela cumpre o princípio da Constituição Federal de 1988, que assegura aos cidadãos o direito de ter acesso às informações que sejam de seu interesse, referentes aos órgãos públicos, por meio do Art. 5., inciso XXXIII da Constituição Federal. Via de regra, esse acesso se dá por meio dos atuais portais da transparência dos órgãos públicos. Nos termos da LAI, desde que estes dados não sejam imprescindíveis à segurança da sociedade e do Estado (Brasil, 1988a), são necessariamente dados abertos.

Ainda em 2011, foi criada a *Open Government Partnership* (OGP), uma parceria entre diversos países e entidades com a proposta de Governo Aberto. A recomendação dessa iniciativa foi de que todos os países participantes fizessem compromisso com a transparência em relação aos dados. O Brasil é um dos países fundadores desta iniciativa e as ações são coordenadas pela Controladoria Geral da União (CGU). Na maioria dos casos, as informações devem ser solicitadas pelo cidadão através do site do órgão, seguindo a aba “Portal da transparência” e solicitar por escrito e de forma mais clara e objetiva possível, qual a informação desejada e esperar por alguns dias até que o órgão forneça uma resposta. Em casos em que as informações buscadas são dados ou microdados mais detalhados, é possível obtê-los em uma aba específica através da plataforma **Gov.br**¹, onde todo cidadão pode, de forma gratuita, fazer um cadastro e navegar pelos órgãos. A iniciativa **Gov.br**, foi uma maneira que o governo brasileiro encontrou de manter compilados todos acessos dos usuários aos sistemas e informações relacionadas ao governo.

Segundo o CENTRO REGIONAL DE ESTUDOS PARA O DESENVOLVIMENTO DA SOCIEDADE DA INFORMAÇÃO (2023), microdados são:

[...] a menor fração de um dado coletado em uma pesquisa. Retrata, por meio de códigos numéricos, a resposta individual dos informantes ao questionário aplicado. Por meio do agrupamento dos microdados em certas unidades [...] são gerados novos dados agregados, permitindo a construção de dados sobre unidades mais amplas.”

No contexto deste trabalho, os microdados podem ser entendidos como a menor fração de um dado ou um conjunto de dados detalhados e pode estar

¹ Disponível em: <https://www.gov.br/pt-br>.

relacionado a uma avaliação, como no caso das provas do Exame Nacional do Ensino Médio (Enem), ou a uma pesquisa. Normalmente, os microdados são gerados para serem compreendidos por plataformas de *softwares* específicos que fazem o tratamento e o cálculo estatístico quando for o caso. No Brasil, o Instituto Nacional de Pesquisas Educacionais Anísio Teixeira (INEP) é o maior responsável pela coleta, tratamento e produção de microdados relativos à educação. Além dele, o Instituto Brasileiro de Geografia e Estatística (IBGE) também é um grande produtor de microdados no que tange a informações da população brasileira.

Sobre tratamento de dados, observa-se que os profissionais da área de Tecnologia da Informação (TI) são quem dominam, em tese, esse trabalho. Mas é importante observar que, no que tange à análise da informação, atualmente, o profissional da Ciência da Informação tem se qualificado cada vez mais, visto que “dado” pode ser entendido como uma informação como qualquer outra, porém apresentada numa estrutura diferente das comumente conhecidas. Com base nisso, o profissional da Ciência da Informação é capaz de utilizar técnicas como *Storytelling* (em livre tradução, narrativa) para auxiliar o usuário na compreensão desses dados. *Storytelling*² nada mais é do que a habilidade de construir uma narrativa que apresente ao usuário as informações solicitadas de forma que seja facilmente compreendida por ele.

Neste cenário, o presente trabalho tem como propósito apresentar a importância do profissional da Ciência da Informação no tratamento de dados e na análise da informação, com o intuito de demonstrar ser possível apresentar ao usuário informações que inicialmente são tratadas em *softwares*, de forma simples e objetiva. Além disso, incentivar o uso de dados abertos como forma de subsidiar decisões.

Para alcançar essa meta foram escolhidos os microdados do Enem realizado em 2022, coletados no site do INEP.

² Nesse sentido, existe um termo chamado *Data Storytelling*, que diz respeito à habilidade de criar uma narrativa envolvente e que combine análise de dados e a visualização destes, seja por meio de gráficos, imagens, tabelas combinando recursos da contação de história.

2 PROBLEMA DE PESQUISA

Como transformar os dados abertos em informações úteis para subsidiar decisões?

2.1 Justificativa

Com o crescente volume de dados que o avanço da tecnologia, principalmente da internet tem trazido para a atualidade, faz-se necessário uma análise aprofundada de como tratar, recuperar e armazenar todas essas informações. O volume de dados produzidos por vezes ficam perdidos, ou não conseguem chegar de forma correta para os usuários, ora por não saber como procurar, ora por não conseguir compreender a informação da forma como ela chega. A proposta dos dados abertos surge de uma busca de flexibilização do acesso e da divulgação das informações científicas e governamentais, incentivando a contribuição para além da transparência, mas também para a participação, a inovação e até mesmo a geração de pesquisas e como consequência a geração de valor econômico.

Analisando a ciência de dados e trazendo para o contexto da Biblioteconomia, é possível perceber a importância da análise de dados e sua utilidade para facilitar o acesso à informação aos cidadãos, podendo ser útil na tomada de decisão de quem busca informações específicas, como no caso dos dados do Enem.

A proposta deste trabalho é democratizar a informação por meio da análise e do tratamento de dados abertos, para transformá-los em informação propriamente dita, colocando em evidência o profissional da Ciência da Informação. Além de dar mais clareza a dados, como os do Enem. O detalhamento proveniente desta pesquisa, gerados a partir de técnicas da ciência de dados, com gráficos estatísticos para demonstrar com mais facilidade a importância de análises estatísticas, tornando assim mais acessível a compreensão das informações.

3 OBJETIVOS

3.1 Objetivo Geral

O objetivo desta pesquisa, é apresentar a importância do profissional da Ciência da Informação no tratamento de dados e na análise da informação, com o intuito de demonstrar ser possível apresentar ao usuário informações que inicialmente são tratadas em softwares, de forma simples e objetiva.

3.2 Objetivos específicos

- Tratar os microdados coletados do site do Inep relativos ao Enem 2022 para manter apenas os dados relevantes que são o foco da pesquisa;
- Analisar estes dados de forma quantitativa e qualitativa;
- Mostrar a importância do profissional de Ciência da Informação no tratamento de dados como ferramenta humana na democratização da informação.

4 REVISÃO DE LITERATURA

4.1 A Ciência da Informação

O termo “informação” é uma palavra cheia de significados ambíguos em diferentes áreas do conhecimento, mesmo que parecidas, por ser uma palavra complexa, a cada vez que aparece pode trazer um conteúdo novo. Buckland (1991) aponta que nem sempre as definições encontradas para a informação são passáveis ou completas, ou que qualquer significado determinado como correto seja o mais satisfatório para definir o termo, mas existem três características complementares que são mais satisfatórias. A primeira seria considerar a informação como o “processo” de alguma coisa. Se uma pessoa passa pelo processo de ouvir uma informação sobre qualquer assunto que seja, ela sofre uma modificação. Nesse contexto, a informação passa a ser ação. O ato de informar alguém sobre algo se torna uma ocorrência que causou uma mudança tanto nele (ou nesse indivíduo), quanto no ambiente em que habita. O segundo seria a informação como conhecimento que acaba sendo um próximo estágio, como sequência de ser uma ação, ela se transforma em conhecimento logo em seguida, por se referir a um fato específico, trazendo para o indivíduo a sensação de saber, diminuindo o desconforto com um fato desconhecido. Como terceiro processo, e talvez o mais complexo dele, é entender a informação como coisa, uma vez que atribuímos, de forma quase que automática, a um documento a função de informação, tornando assim a informação um objeto, tendo ele o papel de informar, comunicar etc.

Porém, mesmo se se considerar de alguma forma a informação como objeto, deve-se lembrar que uma das características mais importante dela, enquanto conhecimento, é que ela é intangível, mas é preciso que esteja representada em algum suporte físico para se tornar comunicação. Taylor (2004, p.1) ressalta que a humanidade tem como característica fundamental a organização e diz ainda que “o aprendizado humano é baseado na habilidade de analisar e organizar dados, informação e conhecimento” e que “nós organizamos porque nós precisamos recuperar”, assim, usamos a informação concretizada em coisa (dado) para recuperar o que ficou no passado

Criado em torno de 1960 (HELPRIN, 1989, *apud* PINHEIRO; LOUREIRO, 1995), o termo “Ciência da Informação”- CI surgiu a partir de um estudo que relacionava a produção, o uso e o processamento da informação com atividades consideradas predominantemente dos seres humanos. No entanto, um estudo em torno de terminologias realizado por Wellish (1987, *apud* PINHEIRO; LOUREIRO, 1995) garante que o termo já havia sido utilizado antes, em 1959, para denominar um conhecimento registrado e sua transferência, de uma forma mais geral. Já Miranda (2002) assegura que o acontecimento da “explosão de informações” (expressão relacionada aos acontecimentos pós-2º Guerra Mundial) teria dado origem ao termo, e que o tratamento da documentação é uma sequência intrínseca a esse processo.

Le Coadic (1994) aponta em sua obra que as pessoas³ iniciam a busca por uma informação a partir da existência de um problema que precisam solucionar, ou de um objetivo que querem alcançar e a partir da constatação de um estado de ausência de certo conhecimento ou da insuficiência dele. Em paralelo a isso, Kuhlthau (1991) valida que o conhecimento do indivíduo que está em busca de informação é dinâmico, ou seja, é modificado conforme suas buscas demonstram avançar no processo.

Há quem defina a Ciência da Informação como um processo, e há quem, como Borko (1968), defenda que ela é muito mais que um processo. Para ele seria uma disciplina que trata de alguns processos para definir a informação, como coleta, tratamento, disseminação, armazenamento e recuperação. Borko visualiza assim a ciência aplicada diretamente ao desenvolvimento do uso da informação.

Barreto (2002, *apud* CAPURRO, 2003) resume os objetivos da Ciência da Informação da seguinte forma:

O objetivo do trabalho com a informação é promover o desenvolvimento do indivíduo, do seu grupo e da sociedade. Esse desenvolvimento deve ser entendido de uma forma ampla, como um acréscimo de bem-estar, um novo estágio de qualidade de convivência, alcançado através da informação. A ação social maior é fazer a luz brilhar para cada ser humano através da informação como mediadora do conhecimento.

Finalmente pode-se estabelecer, não de forma definitiva, mas concordando com Urdaneta (1992), que a informação pode ter quatro escopos diferentes: a

³ Organizações são pessoas. De acordo com BOOG et al (1994), a estratégia da Gestão de Pessoas é a alma, o coração e o cérebro da estratégia global da organização

informação como dado, informação como conhecimento, informação como a própria informação e informação como inteligência. Mas para entender melhor, de acordo com esta pesquisa, quais as vertentes da informação é importante, é prudente que seja escolhida uma definição para endossar melhor o objetivo que está sendo traçado. Por tanto, a ideia de informação que McGee e Prusak (1994) trazem, é a que melhor define o uso para este trabalho.

[...] na verdade informação são dados coletados, organizados, ordenados, aos quais são atribuídos significados e contexto, ou seja, para que os dados se tornem úteis como informação a uma pessoa encarregada do processo decisório é preciso que sejam apresentadas de tal forma que essa pessoa possa relacioná-los e atuar sobre eles.

Qualquer tentativa de contextualizar a Ciência da Informação, deve considerar que historicamente a ação humana frente à produção de registros materiais, independente de seu suporte, fez necessária a criação de instituições com o objetivo de salvaguardar todos esses documentos (entendidos aqui em seu sentido amplo), que deu origem ao que conhecemos hoje como instituições informacionais, dedicadas a acolher a informação cada um em seu suporte, como por exemplo as bibliotecas, os arquivos e os museus. Barreto defende que:

[...] informação sintoniza o mundo, pois referencia (sic.) o homem ao seu passado histórico, às suas cognições prévias e ao seu espaço de convivência, colocando-o em um ponto do presente, com uma memória do passado e uma perspectiva de futuro; o indivíduo do conhecimento se localiza no presente contínuo que é o espaço de apropriação da informação. (BARRETO, 2007, p. 23)

Com a invenção da imprensa, iniciou-se um aumento de produções bibliográficas e de onde surgiram os primeiros modelos de bibliografias diferente da finalidade que conhecemos hoje, viriam pela necessidade de se inventariar a produção. Essas primeiras bibliografias foram produzidas em grande parte por bibliófilos. Apesar de seu objetivo primário não ter nenhuma relação com a criação de uma coleção e menos ainda uma preocupação com a custódia, esta tarefa acabou criando uma atividade discricionária das obras.

Foi então que, em 1985, surgiu a I Conferência Internacional de Bibliografia, criada por Henri La Fontaine⁴ e Paul Otlet⁵ com um ambicioso objetivo de fundar um movimento cooperativo que fosse capaz de reunir todo tipo de produção humana em apenas um lugar, uma espécie de inventário. Para que isso fosse possível, dedicaram-se a criar uma padronização que pudesse ser utilizada como modelo de tratamento técnico dos registros, surgindo as primeiras ideias de fichas catalográficas e a Classificação Decimal Universal (CDU) em apoio à ideia.

A partir dessa e de outras movimentações dos cientistas, surge então o fenômeno da profissão, os cientistas da informação. Houve, entre os anos de 1920 e 1940, uma grande movimentação de cientistas de diversas áreas, que começaram a se dedicar à criação de resumos, índices e outros tipos de materiais que colaboraram com a disseminação das informações. Essa moção foi se transformando com a prática e acabou sendo um movimento muito importante para a institucionalização da área, e após isso, diversos movimentos documentais começaram a surgir em diversos países por meio de conferências e institutos.

Sendo assim, podemos considerar alguns marcos que deram início à cinesia da Ciência da Informação:

1. A criação do Instituto Internacional de Bibliografia (IIB) em 1895, idealizado por Paul Otlet e Henri La Fontaine, contextualizado através da ideia da obra "*Livro universal do conhecimento*" (MATTELART, 2005), seu objetivo não era criar um grande acervo bibliográfico, mas sim, uma grande biblioteca de referência (OLIVEIRA, 2005) para democratizar a informação.
2. A mudança do IIB para Instituto Internacional de Documentação (IID), em 1931, ideia que surgiu durante a X Conferência Internacional de Bibliografia.
3. A publicação de duas grandes obras de Paul Otlet em 1935, o *Traité de Documentation: le Livre sur le Livre: Théorie et Pratique* e a obra *Monde:*

⁴ Henri La Fontaine (1854-1943) formou-se em direito pela Universidade Livre de Bruxelas, ele e sua irmã foram os primeiros defensores dos direitos e do sufrágio das mulheres, fundador da Liga Belga pelos Direitos das Mulheres em 1890. Na Biblioteconomia também obteve papel revolucionário, sendo o criador, juntamente com Paul Otlet, da Classificação Decimal Universal em 1895. Além disso, também foi um dos fundadores da Biblioteca Mundial.

⁵ Paul Marie Gislain Otlet (1868-1944), advogado e ativista da paz, é considerado um "pai" da Biblioteconomia e Ciência da Informação, área essa que ele denominava "documentação". Visionário, foi o responsável pela criação da Classificação Decimal Universal.

Essai d'universalisme, que implementaram conceitos de união dos elementos da documentação, já evidenciando concepções relacionadas à Bibliometria.

4. A fundação do *American Documentation Institute* (ADI), em 1937, que mais tarde se tornou *American Society for Information Science - ASIS*, e depois, nos anos 2000 se tornou *American Society for Information Science Technology – ASIST*.
5. Em 1938, houve a mudança do IID para Federação Internacional de Documentação (FID). Essa instituição foi a responsável pela base do desenvolvimento de pesquisas científicas da CI.

Além disso, de acordo com Pinheiro (2022), a CI nasceu principalmente

- a) da explosão informacional, ocasionada pelo avanço científico e tecnológico da demanda da Segunda Guerra Mundial (RUSSO, 2010),
- b) da necessidade da recuperação, do registro (controle bibliográfico) e transmissão de informação e conhecimento (serviços informacionais dispostos para atender, de acordo com Miranda (2002), à pesquisa e desenvolvimento);
- c) do advento de novas tecnologias, especialmente o computador (que começou a ser usado no processamento de informações bibliográficas já na década de sessenta). Todavia, Barreto (2007) observa que de fato na CI, o uso de computadores passou a ser recorrente apenas na década de 80, do século XX.

Para Miranda (2002, p.10),

[o] surgimento da Ciência da Informação estaria [...] relacionado com a atividade subsequente ao controle da produção científica e à regularidade do fenômeno relativo à sua dispersão e uso, obsolescência, epidemiologia de sua propagação e outros aspectos detectados no processo de manipulação e análise da literatura.

Os autores chegaram então a um consenso, no que tange ao princípio da CI que é o da recuperação da informação, necessidade que surge em 1940, mas que só é considerada consolidada em 1990, com as discussões de natureza interdisciplinar que envolvem a área.

Segundo Russo (2010), a CI foi introduzida no Brasil na década de setenta, com a criação do curso de Mestrado em Ciência da Informação, criado pelo Instituto Brasileiro de Bibliografia e Documentação (IBBD) que hoje se tornou o Instituto

Brasileiro de Informação em Ciência e Tecnologia - IBICT. Para além de instituições, o periódico *Ciência da Informação*, criado em 1972, foi um importante aliado nessa jornada.

Mesmo após a sua consolidação em 1960, o conceito de CI foi muito questionado sobre o seu propósito, e diversas dúvidas sobre que tipo de ciência a CI era ou deveria ser, fez com que algumas propostas acerca disso fossem pautadas.

A primeira, diz respeito a sua interdisciplinaridade, Borko (1968) trazia a questão propondo que a natureza da CI era totalmente “natural”, com o que Saracevic (1970) concordava. Entretanto, nesse momento da história, praticamente não houve consequências epistemológicas relevantes para a CI. Esse entendimento se consolidou para justificar a grande atuação de cientistas de diversas áreas do conhecimento que ajudaram a construir o que temos hoje como disciplina.

Por outro lado, havia uma argumentação de que ela era interdisciplinar por ser totalmente moldável a qualquer área que necessitava de serviços de informação. Essa ideia de interdisciplinaridade, por décadas, não havia sido um consenso entre as ciências em si. Muitos entendiam o termo “natural” como uma forma de uma ciência nova, que usava de métodos e conceitos de outras áreas já consolidadas como uma forma de apropriação, sem esclarecer de onde vieram. No entanto, atualmente entende-se a Ciência da informação como interdisciplinar por ser capaz de mediar disciplinas diferentes, por conseguir fazer duas ou mais ciências chegarem a um entendimento comum em determinada situação. Essa especificidade tão plural e tão singular ao mesmo tempo, foi se moldando ao longo dos anos até se tornar o que conhecemos hoje como Ciência da Informação.

Apesar disso, não podemos deixar de lado um segundo grande ponto acerca do debate do que é ou deveria ser a CI. Wersig (1993) enfatiza o fato de que “informação” não pode ser considerado um objeto específico da realidade, não é algo que já foi estudado por uma ciência, é um fenômeno muito específico e muito singular dentro da sua própria existência e por isso tem o potencial de ser considerada uma ciência pós-moderna, que será discutida mais adiante.

Os grandes autores do tema conversam entre si, dentro da ideia de que a CI não precisa assumir o papel de uma ciência que estuda um objeto do mundo, mas seria a ciência que estuda a visão desses outros objetos que as outras ciências apresentam. De forma prática, a CI mostra a sua forma de enxergar um objeto que

existe, sem dizer qual o nome desse objeto. Se usar uma analogia de um objeto real, pode se entender que uma ciência que estuda um objeto real, será capaz de informar o que e qual a finalidade daquele objeto. Mas a CI trará uma análise informacional sobre esse objeto, e vai explicar como e por que ele é relevante.

É percebido que a CI passou por constantes questionamentos desde que surgiu o termo pela primeira vez. Berger e Luckmann (1966) em seu trabalho, traz uma nova discussão para o entendimento do que é e o que faz a CI. No momento em que colocamos em pauta que algo “real” é algo construído dentro de um contexto social e não com uma existência em si mesma, abre-se um caminho para um entendimento de que a informação não é somente um dado, uma *coisa*, mas que na realidade a informação é um processo, e que ela ainda vai ser percebida e compreendida dentro do ambiente em que ela está inserida, ou seja, a informação será percebida como informação a depender de como os sujeitos estão em relação a ela, contrariando Borko (1968) que propõe que nesse fluxo o sujeito é excluído, tal qual Buckland (1991) que nos traz a informação como *coisa*.

Para arrematar a provocação criada em torno da CI, que até o momento informa sobre algumas tensões em volta desse tema, como já apresentada, a CI surge num contexto em que já se observam críticas nos modelos cientificistas da época. Pode-se perceber que, na evolução da CI, que ela se agarrava a uma tentativa de se consolidar como uma disciplina resultante do processo da ciência moderna. Além disso, ela também já se inclinava na tentativa de constituir-se nos moldes da “ciência pós-moderna”.

A nova “ciência pós-moderna” (Santos, 1996) é percebida como uma espécie de movimento que destaca uma superação do momento de crise do paradigma científico dominante desde o século XVII, indo além do modelo cartesiano de racionalidade, da separação entre sujeito e objeto, da busca pela ordem da separabilidade dos elementos constitutivos da realidade, movimento motivado pelas crises geradas pela evolução e aplicação do conhecimento científico, como as guerras, totalitarismo, regimes econômicos, poluição e desastres ecológicos, exclusão do acesso ao conhecimento, reforço das desigualdades socioeconômicas, sofisticação dos instrumentos de dominação (Adorno, Horkheimer, 1990). Nesse movimento, diversos autores tomam partido da construção dessa ciência, são eles: Santos (1996, p.23-35) as contribuições de Einstein (Teoria da Relatividade), da mecânica quântica (Heisenberg e Bohr), de

Godel (Teorema da Incompletude) e de Prigogine (ordem através das flutuações). Outras perspectivas que aqui se incluem Christóvão e Braga (1997, p.35-39) são a Teoria do Caos (Lorenz, Stewart e muitos outros) e a Teoria da Autopoiese (Maturana e Varela).

No caso da CI, ela pode ser envolvida nesse movimento da “ciência pós-moderna” pela tranquilidade de se encaixar e se aceitar como uma ciência de natureza interdisciplinar.

“Constitui-se assim sua interdisciplinaridade, característica cada vez mais presente como componente da Ciência na sociedade atual, em que a magnitude dos problemas enfrentados (ecológicos, étnicos, demográficos) está a exigir soluções inovativas e plurais. A ciência da informação vem se consolidando, então, a partir de elementos emprestados da matemática, da física, da biologia, da psicologia, da sociologia, da antropologia, da semiologia e da teoria da comunicação e de quantas ciências puderem contribuir para sua fundamentação e aplicabilidade” (Cardoso, 1996, p.74).

Ainda que ela se aproxime das ciências exatas e biológicas, se percebida em um contexto de não se empoderar de conceitos ou até mesmo analogias para a construção de sua estrutura interdisciplinar, mas usando a busca por novos conhecimentos em um movimento transdisciplinar (além de interdisciplinar) num complexo pensamento de “religação dos saberes” (Morin, 1987 e 2001). Como também afirmava Wersig,

“such a science would be established as a prototype of a new or postmodern science. Postmodern science is not like classical science, driven by the search for complete understanding of how the world works, but by the need to develop strategies to solve in particular those problems which have been caused by classical sciences and technologies” (Wersig, 1993, p.229).⁶

A partir do momento em que entendemos que a CI muda completamente a forma como se estuda o objeto, abre-se uma lacuna para a reflexão sobre a relação desta com as ciências sociais e com o que se tem de mais moderno da ciência num geral, onde passa a ser fundamental estreitar as relações criadas nas manifestações da CI entre o social e o “pós-moderno”.

⁶ Tal ciência seria estabelecida como uma ciência nova ou pós-moderna. A ciência pós-moderna não é como as ciências clássicas, orientadas pela procura de um completo entendimento de como o mundo deve funcionar, mas visando a necessidade de desenvolver estratégias para resoluções, em particular, aos problemas causados pela ciência clássica e suas tecnologias. (tradução da autora).

No período renascentista, com as produções humanas crescendo e em paralelo à necessidade de salvaguardar esses registros, surgem então os primeiros tratados e manuais sobre o que se entendia por acervo na época, e isso compreendia “acervo” de museus, bibliotecas e arquivos. E com as transições de épocas e mudanças de discursos sobre as ciências, também surgiam novos tratados e novos entendimentos sobre o que seriam museus, bibliotecas e arquivos. No século XIX, essas transições tomam uma força e avançam consideravelmente e nasce a necessidade de uma sistematização das regras de preservação do vasto conhecimento que aumentava o tempo todo. É a partir daí que se observa uma movimentação quase que orgânica do nascimento das disciplinas Arquivologia, Biblioteconomia e Museologia como ciências relacionadas a acervos. É nesse momento também, que elas são entendidas como instituições.

A partir dessa movimentação e levando em consideração todas as evoluções citadas anteriormente, é fundamental entender que nesse momento a CI já encontrou o seu lugar na ciência, e que a Biblioteconomia trilhou o seu caminho para ocupar o seu espaço institucional como a que deixou de ser espaço de salvaguardar conhecimento e produção, e passar a ocupar o lugar que preserva e recupera.

4.2 Organização da Informação

Após entender os conceitos de informação dentro da Ciência da Informação, é necessário compreender o papel da organização da informação, que está intrínseco no processo de transformação de dados em informação, que tanto é discutido dentro do escopo da Ciência da Informação.

[...] informação sintoniza o mundo, pois referencia (sic.) o homem ao seu passado histórico, às suas cognições prévias e ao seu espaço de convivência, colocando-o em um ponto do presente, com uma memória do passado e uma perspectiva de futuro; o indivíduo do conhecimento se localiza no presente contínuo que é o espaço de apropriação da informação. (Barreto, 2007, p. 23).

O conceito de organização da informação (OI), parafraseando Taylor (2004, p.1) “organizamos porque nós precisamos recuperar”, advém de uma necessidade

primária humana, de controlar a informação para posteriormente poder ser recuperada, e isso independe se essa informação é um documento ou um objeto. Cada ser humano tem uma certa habilidade de organizar. Imagine que toda a disposição de móveis de uma casa é colocada de forma estratégica para que a dinâmica do ambiente seja fluida e não atrapalhe, por exemplo, a circulação das pessoas no ambiente. E que, se você procurar pelo sofá, você saberá onde encontrar e isso não se deve, nesse contexto, ao seu tamanho, mas sim por inicialmente você saber onde procurar já que previamente fora colocado num ponto “x”. Portanto, seguimos a mesma lógica para a organização da informação, além de destacar que esse processo está intimamente ligado à memória.

Autores como Svenonius entendem que a organização da informação pode compreender uma ação particular da linguagem.

A vantagem a ser obtida por considerar o ato de organizar a informação como a aplicação de uma linguagem de propósito especial é que os constructos da linguística, tais como vocabulário, semântica e sintaxe, podem ser utilizados para generalizar entendimento e avaliar diferentes métodos de organização da informação. Outra vantagem é que esses constructos possibilitam a conceitualização que pode unificar métodos, antes díspares, de organização da informação – catalogação, classificação e indexação. (SVENONIUS, 2000, p.6)

A Organização do Conhecimento (OC) tem seu princípio enraizado na base teórica da classificação, que trata sobre a organização dos objetos e/ou assuntos por “classes”, sejam elas separadas pelas diferenças ou semelhanças, com um propósito específico, e é também uma forma de representar esse objeto. Nesse aspecto, Hagar Gomes (2009) aponta em seu trabalho “Tendências da pesquisa em organização do conhecimento” que:

Percebe-se, então, que a OC é uma área de múltiplas aplicações. As pesquisas especificamente na área apontam para, no mínimo, dois aspectos importantes: de um lado, a abordagem teórica-filosófica voltada para questões epistemológicas ao analisar um domínio juntamente com as aplicações na organização de sistemas/tabelas de classificação e, de outro lado, os métodos empregados na elaboração propriamente dita dos sistemas, atividades que alguns autores consideram suplementar, mas que tem suas bases teóricas já sedimentadas. (GOMES, 2009, p. 63).

Neste contexto, surgem também termos como Organização do Conhecimento e destaca-se que apesar de similares, carregam significados bem distintos.

Os termos organização do conhecimento e organização da informação têm sido utilizados em diferentes contextos para denominar instituições, grupos e linhas de pesquisa, disciplinas e cursos na área de Ciência da Informação. No entanto, a análise do emprego desses termos nesses contextos revela falta de clareza quanto à delimitação do conceito. Por vezes o termo organização do conhecimento é utilizado no sentido de organização da informação e vice-versa e, em determinadas situações, empregam-se os termos conjuntamente – organização da informação e do conhecimento. (BRASCHER; CAFÉ, 2008, p.2)

Por um outro lado, Capurro (1991) reforça que no contexto do paradigma da representação, os seres humanos são capazes de obter conhecimento por serem observadores, e que o processo de construir o conhecimento se daria por meio de assimilação das coisas com os meios em que elas interagem. Isso quer dizer que os seres humanos seriam processadores biológicos de informação, e assim, utilizam as informações, de forma geral, para necessidades específicas. Olhando por este lado, poderia ser aceitável que máquinas processassem o conhecimento armazenado nelas e teriam capacidade de “interpretar” qual seria a necessidade humana demandada naquele momento.

Barreto (1994, p.12) atenta ao fato de que a informação é “[...] um instrumento modificador da consciência e da sociedade como um todo [...] qualificada como um instrumento codificador da consciência do homem e de seu grupo [...] deixando de ser uma medida de organização para ser a organização em si...”, e reforça que o processamento da informação se dá por meio da assimilação do objeto percebido no ambiente em que ele se insere, assim como acredita Capurro (1991). Assim, pode se considerar que para o indivíduo gerar o conhecimento ele deve estabelecer vínculo de significação com o conhecimento.

Barreto (2002, p.72) valida essa visão quando considera que “destino da informação organizada em estruturas mentais, por meio das quais um sujeito assimila a coisa informação”. Levando em consideração esse entendimento, é possível compreender então que toda informação depende de uma construção de um ou mais significados e que é um acontecimento cultural. Lima (2003, p.81) também apoia que para processar uma informação, é necessário ter um conhecimento prévio do objeto de reconhecimento, nesse caso a informação,

portanto, está associada à forma como o indivíduo vê o mundo, sugerindo que para cada indivíduo, a informação se torna diferente.

É impossível entender a Organização da Informação, sem entender que ao longo dos anos a sociedade vem demonstrando uma necessidade maior de participação do coletivo e não mais do indivíduo de maneira singular, a busca por essa participação, tem levantado um questionamento sobre como a OI é essencial para a preservação do conhecimento humano e principalmente para o desenvolvimento de toda uma nação, isso só demonstra que a disponibilização descentralizada desses conhecimentos podem e vão se tornar um problema. O ambiente informacional precisa de uma estrutura e um cuidado para que os fluxos criados sejam capazes de atender as demandas da sociedade, e que sejam capazes de diminuir a exclusão provocada pela desigualdade social.

Barreto (1994, p.4) declara que:

Democratizar a informação não pode, assim, envolver somente programas para facilitar e aumentar o acesso à informação. É necessário que o indivíduo tenha condições de elaborar este insumo recebido, transformando-o em conhecimento esclarecedor e libertador, em benefício próprio e da sociedade onde vive.

Além disso, Smit (2002, p.27) expressa que:

[...] o sonho da farta distribuição de informações como condição necessária e suficiente para a melhoria da sociedade e da qualidade de vida do cidadão torna-se irreal, utópico, impossível [...] uma vez que esta é determinada por uma capacidade apropriativa da informação, capacidade esta, elaborada e fomentada, oficialmente pelo processo educacional no indivíduo, e não pelas instituições coletoras de cultura.

Nesse panorama, o estudo da Organização da Informação se mostra totalmente inserido no contexto da CI, já que melhorar o fluxo da informação faz parte do que se entende como sua própria ciência, como reconhecida por Saracevic (1996) que nos lembra que a CI está interligada ao estudo que relaciona o indivíduo com o seu contexto social, institucional e individual no contexto em que estabelece uma conexão significativa com o usuário, pois apenas ao reconhecer seus interesses e necessidades de informação é possível criar um vínculo entre os sistemas de informação e os usuários. Isso, por sua vez, melhora o desempenho da função comunicativa.

Abordando de forma mais teórica, destaca-se também a organização do conhecimento, que também é uma área de estudos da CI, García Marco (1995, p.219) considera a OC da seguinte forma:

[...] área dedicada ao estudo e desenvolvimento dos fundamentos e técnicas da planificação, construção, gestão, uso e avaliação de sistema de descrição, catalogação, ordenação, classificação, testemunhar, conservar e transmitir seu saber e seus atos, a partir de conteúdo, com o fim de garantir sua conversão em informação capaz de gerar novo conhecimento.

No contexto da Biblioteconomia, a OC diz respeito à organização dos registros bibliográficos, nessa mesma linha Hjørland (2003) acredita que não só a Biblioteconomia, mas que a comunidade inteira da CI pensa da mesma forma. Ademais, sendo ou não toda a comunidade, é fato notório que para estabelecer o melhor caminho para a CI, deve-se levar em consideração a otimização dos meios para se obter os registros no momento da recuperação da informação.

Sobre a recuperação da informação, Robredo aponta para um cenário onde a informação ultrapassa os limites da oralidade.

Registrada, duplicada, transmitida, armazenada, organizada, processada, recuperada. Sim, mas somente quando extraída da mente e codificada, pela linguagem natural (falada ou escrita), seguindo normas e padrões (gramática, sintaxe) próprios de cada língua, ou de outras linguagens criadas pelo homem (linguagens de programação, que também têm suas gramáticas e sintaxes). (ROBREDO, 2007, p.60).

Num cenário onde há uma necessidade de tratamento, advinda do considerável aumento de informação. No entanto, outros autores complementam os pensamentos de Robredo, mas que ao mesmo tempo traz uma atenção a outro cenário, Barreto e Smit (2002, p.21) consideram importante realçar que “[...] nem toda informação é estocada, nem toda informação é registrada...ao estocar a informação, esta passa a ter uma existência institucional e portanto, social.”. Isto porque como já destacado, qualquer que seja o tratamento, é necessário levar em consideração o contexto em que ela está inserida, para que a informação seja capaz de criar um vínculo com o usuário.

Ainda parafraseando Hjørland (2003), o conceito da OC, ainda é bastante amplo visto que se encontra envolto de muitas questões relacionadas às instituições e suas disciplinas, linguagens e conceitos ainda muito teóricos.

Com a Revolução Industrial, a necessidade de organizar uma espécie de “mapa do saber” a partir das criações humanas, se tornou muito necessário dado o momento e a situação das bibliotecas, que ansiava pela criação de um instrumento que fosse capaz de armazenar e recuperar de forma mais estruturada o conhecimento produzido ao longo dos anos.

A primeira vez que a expressão Organização do Conhecimento surgiu, foi com o bibliotecário Henry Bliss⁷, também autor de três obras relacionadas aos estudos da OC, são elas: *Organization of Knowledge (1927)*, *Organization of Knowledge in Libraries and subject approach to books (1933)* e *A System of Bibliographic Classification (1935)*.

Muitos foram os nomes que influenciaram as mudanças em torno da OC, Ranganathan⁸, foi um forte influenciador com suas ideias em torno da criação de tesouros e de um sistema de classificação especializado. O periódico “International Classification”, criou uma espécie de bibliografia voltada para o conceito de OC, desde 1974, e em 1993 mudou seu nome para “*Knowledge Organization*”. Em 1989, uma pesquisadora alemã chamada Ingetraut Dahlberg⁹ fundou a *International Society for Knowledge Organization (ISKO)*, uma espécie de sociedade com o objetivo de desenvolver estudos específicos na área, visando uma amplitude maior do conceito e buscando por meio de criação de bases de dados, bibliotecas e outros suportes, propostas que fossem capazes de alavancar o trabalho em torno da OC. Assim como a CI em si, essa sociedade investe na interdisciplinaridade que a OC precisa, contando portanto com profissionais de diversas áreas do conhecimento. Todo essa diligência em torno da OC, tem como justificativa, de acordo com Barite

[...] porque a humanidade só avança na medida em que se sistematiza o saber acumulado para cumprir determinados propósitos e só através do estabelecimento de consensos sobre este saber, se torna possível o intercâmbio, a comunicação, o debate a difusão e os modos de circulação do conhecimento especializado. (BARITE, 2001, p.39).

⁷ Henry Evelyn Bliss (1870-1955) foi autor e bibliotecário estadunidense renomado. É o criador do sistema de classificação *Bibliographic Classification*, obra de grande referência na Biblioteconomia até os dias atuais.

⁸ Shiyali Ramamrita Ranganathan (1892-1972), nasceu na Índia e é conhecido como pai da Biblioteconomia. Bibliotecário e professor, muito da moderna Teoria da Classificação é fundamentada em seus extensos estudos na área.

⁹ Ingetraut Dahlberg (1927-2017) foi pesquisadora, cientista da informação, editora e filósofa alemã que desenvolveu a Classificação Universal de Codificação de Informações. Atuava na área de classificação, tesouros, terminologia e organização do conhecimento.

Visando que as práticas ligadas a esse processo sejam justificativas para o tratamento e gestão da informação.

A Organização do Conhecimento e a Organização da informação, está intimamente interligada ao processo de recuperação. Como já discutido anteriormente neste trabalho, a necessidade de recuperar se tornou natural e necessário para os seres humanos. Borko (1968, p.3) defende que a CI desenvolve um papel de gerenciamento da informação, uma vez que a CI “está ligada ao corpo de conhecimentos relativos à origem, coleta, organização, estocagem, recuperação, interpretação, transmissão, transformação e uso da informação”, sendo assim capaz de desenvolver a partir de uma informação inicial um serviço e até mesmo um produto. Partindo desse ponto de vista, percebemos então que a CI possui características de uma ciência aplicada, sendo assim capaz de criar ou modelar um produto para além de informar, atender a uma demanda específica de um usuário ou um grupo.

Nessa mesma linha de pensamento, Siqueira (2003) afirma que para a CI é natural que seja necessário a criação de um produto ou material para atender as demandas dos usuários, já que como uma ciência social, e para sempre atingir esse objetivo faz-se necessário o uso de serviços diversos, clareando ainda mais a ideia de uma ciência interdisciplinar.

Nesse contexto Lima (2003, p.86) atesta que:

A Ciência da Informação tem consolidado seus pontos de vistas e valores desde que ela foi reconhecida como uma disciplina social e científica nos anos 60 e, mais recentemente como uma área interdisciplinar que reúne a biblioteconomia, a ciência cognitiva (CC), a ciência da computação e a comunicação, com íntima dependência dos processos da comunicação humanos e da tecnologia no seu contexto contemporâneo.

Fica claro então que para uma boa recuperação da informação, é necessário um bom gerenciamento e organização dos materiais coletados.

[...] compreende todos os tipos de instrumentos utilizados para organizar a informação e promover o gerenciamento do conhecimento, incluindo os esquemas de classificação — que organizam materiais em nível geral (como livros em uma estante); cabeçalhos de assunto — que provêem acesso mais detalhado; catálogos de autoridade — que controlam versões variantes de informação fundamental (como nomes geográficos e nomes pessoais) e outros instrumentos menos tradicionais - como redes semânticas e ontologias (MIRANDA, 2005, p.118).

Sem entrar no mérito da Classificação e da Indexação, partindo do pressuposto de que a presente pesquisa cumpriu com o seu objetivo de explicar que a OC e a OI são de extrema importância no ato de recuperar, e após entender os pensamentos dos autores acerca desse tema, conclui-se que para a recuperação da informação, a gestão é imprescindível, e com esse argumento cada unidade de informação deve tomar a decisão de utilizar de mecanismos para tal fim.

Nesse momento surge então uma preocupação com relação à explosão informacional. Para a CI, a solução de recuperar informações seria bem sucedida para esse momento em questão. Mas algumas questões deveriam ser levadas em consideração no sentido de o que, porque e como recuperar, além disso era necessário pensar em como informar de fato. Sobre esse tema, novos estudiosos começam a aparecer como Calvin Mooers¹⁰, trazendo consigo o termo Recuperação da Informação, embora supracitado algumas vezes ao longo desta pesquisa, é em 1950 que o termo é utilizado pela primeira vez na reunião da *Association for Computing Machinery da Universidade Rutgers*. Para o autor o termo “engloba os aspectos intelectuais da descrição da informação e de sua especificação para a busca, bem como qualquer sistema, técnica ou máquina que são utilizadas para realizar a operação” (MOOERS, 1951).

Ainda preocupado em como seria possível de fato descrever as informações de forma intelectual, além de quais mecanismos e técnicas deveriam ser utilizadas para tal objetivo, Mooers criou uma espécie de lei que leva o seu próprio nome, para responder a essas questões. A lei estabelece que: “Um sistema de recuperação de informações terá a tendência de não ser usado se é mais irritante e problemático para um usuário obter a informação do que não obtê-la”. (Mooers 1960).

Essa afirmação destaca um princípio fundamental na concepção de sistemas de recuperação de informações, especialmente em contextos digitais. Em termos simples, isso significa que os usuários são mais propensos a abandonar ou evitar o uso de um sistema de busca ou recuperação de informações se a experiência de usá-lo for frustrante, difícil ou irritante.

A usabilidade é uma consideração essencial no design de sistemas de informação e motores de busca. Se um sistema é complicado de usar, se os

¹⁰ Calvin Northrup Mooers (1919-1994) foi um cientista da computação estadunidense conhecido por seu trabalho em recuperação de informação e pela linguagem de programação TRAC. Em 1950, cunhou o termo "Recuperação de Informação" e obteve diversas patentes em recuperação e sinalização de informação.

resultados não são relevantes ou se a interface é confusa, os usuários podem desistir de usá-lo. Em vez de enfrentar desafios irritantes, os usuários podem optar por não procurar a informação de todo.

Em corolário à lei de Mooers, Garfield (1997) explica que “quanto mais informação relevante um sistema de recuperação de informação provê, mais usado este sistema será”. Autores como Austin (2001), se arriscaram a propor, como uma forma de extensão a Lei de Mooers que:

Primeira Lei de Mooers

“Em um ambiente no qual é mais irritante e problemático para um usuário ter a informação em mãos do que não tê-la, um sistema de informações terá a tendência de não ser usado”;

Segunda Lei de Mooers

“Em um ambiente no qual é absolutamente crítico para um usuário ter informação, um sistema de informação terá a tendência de ser utilizado, não importa o quanto deficiente o mesmo seja”;

Terceira Lei de Mooers

“Em um ambiente no qual o problema de ter informação versus o de não tê-la estão razoavelmente equilibrados, o fator decisivo se um sistema de informação será utilizado ou não será o design e a performance do mesmo”.

Com base nessas contribuições de Austin, seria possível também identificar comportamento do usuário quanto ao acesso às informações buscadas. Além disso, após entender os aspectos da Lei de Mooers e suas extensões, é possível perceber como a recuperação da informação tem a sua importância. E na atual conjuntura, os Sistemas de Recuperação da Informação - SRI's surgem para atender as necessidades informacionais dos usuários. Segundo Miranda elas possuem subsistemas com estruturas conforme apresentado na figura 1:

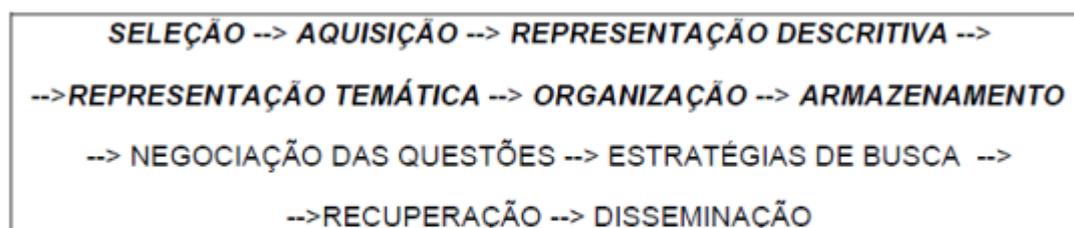


Figura 1: Sistema de Recuperação da Informação.
Fonte: Miranda (1997, p.43).

Nessa mesma linha, e sabendo que as SRI's podem ser diferentes para cada objetivo, Feitosa apresenta um outro subsistema conforme figura 2:

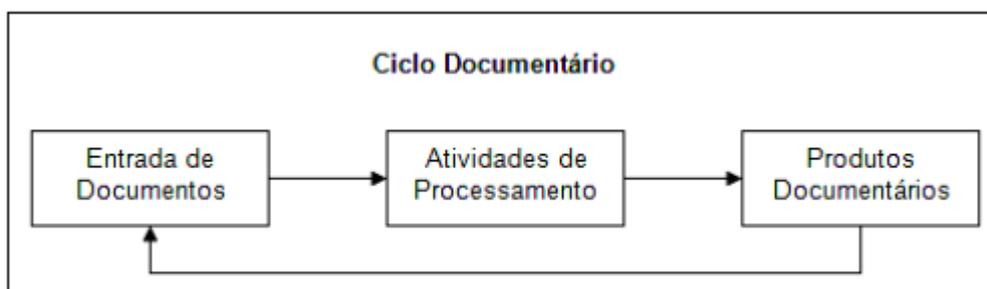


Figura 2: Cadeia Documentária - visão sistêmica.
Fonte: Feitosa (2006, p.19).

Apesar das figuras apresentadas se referirem a aquisição de documentos, a ideia central é demonstrar que sistemas estruturados são capazes de fornecer subsídio no momento da recuperação e transmissão da informação para o usuário que busca. Nesse sentido Rowley (1994, p.113) assevera que:

[...] profissionais da informação atribuem termos de indexação a um documento ou item de informação. Selecionam os tópicos a serem representados pelos termos de indexação com base num julgamento subjetivo, dos assuntos tratados nos documentos. Em seguida associam esses assuntos aos termos de indexação, que acreditam ser provavelmente procurados por um usuário

A autora defende os SRI's, independente se seu suporte é tradicional ou em ambiente virtual, e complementa que devem obedecer as etapas demonstradas na figura 3.

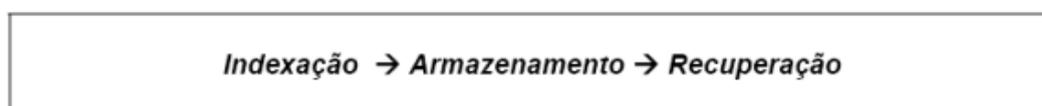


Figura 3: Etapas de uma SRI
Fonte: Rowley (1994, p.113)

Diante o exposto, chegamos à conclusão de que para recuperar, é necessário um gerenciamento das informações, esse gerenciamento/organização diz respeito às formas de tratamentos dos dados, antes dele se tornar uma informação. Os autores apresentados demonstram claramente a importância de tal prática para se

chegar a uma ciência de informação como a que conhecemos atualmente, porém, ainda levando em conta que as ciências estão em constantes desenvolvimentos e descobertas, é necessário que se entenda que outros aspectos atuais devem ser considerados quando se trata de informação, recuperação entre outros termos nessa linha. O desenvolvimento da *web*¹¹ é um dos principais pontos que devem ser considerados, em especial quando se fala em SRI's. Ao falar sobre OI, também deve-se levar em consideração a Representação da Informação.

Hessen (2003, p.59) abordada na teoria geral do conhecimento que o intelectualismo se passa por mediador entre o racionalismo e o empirismo, ressaltando que “[...] se para o racionalismo o pensamento é a fonte e o fundamento do conhecimento, para o empirismo essa fonte e fundamento é a experiência e o intelectualismos considera que ambas participam na formação do conhecimento”.

Muito embora se acredite que o conhecimento pode surgir unicamente da experiência, assim como Hessen traz essa abordagem, Russel (2001, p.385), cita Kant com o intuito de trazer à tona o pensamento de que a experiência não é responsável exclusivamente para que o conhecimento nasça. O conhecimento pode passar por dois estágios, o conhecimento que existe independente de qualquer experiência, e o conhecimento que é adquirido após uma experiência.

Na busca por uma construção sólida do significado de OC, Dahlberg (2006) acredita que a OC é uma ciência capaz de ordenar, estruturar e sistematizar conceitos, baseados em suas características. A partir desse entendimento, também leva-se em consideração as contribuições de Bräscher e Café (2008) que defendem que a OC é:

o processo de modelagem do conhecimento que visa a construção de representações do conhecimento. Esse processo tem por base a análise do conceito e de suas características para o estabelecimento da posição que cada conceito ocupa num determinado domínio, bem como das suas relações com os demais conceitos que compõem esse sistema nocional. (BRASCHER; CAFÉ, 2008, p.2).

As autoras ainda defendem que todo esse processo de OC, no intuito de recuperar informação, é também um processo cognitivo, visto que é um trabalho

¹¹ Desenvolvimento da Web é uma área da tecnologia voltada à criação, desenvolvimento, codificação e programação de sites e seus respectivos elementos e linguagens na Internet ou em uma intranet. Pode variar desde simples páginas estáticas a ricas aplicações, *e-commerce* ou redes sociais. Ele também envolve integração com sistemas e ferramentas, armazenamento de banco de dados, cibersegurança, entre outros, com o objetivo de estruturar cada detalhe de uma página da internet.

basicamente de perceber o objeto, ser capaz de descrever e até mesmo classificar, envolvendo assim memória e pensamento, trabalhando de forma a entender os conceitos envolvidos nesse processo de forma sistemática e semanticamente estruturada.

A representação construída não se restringe ao conhecimento expresso por um autor, ela é fruto de um processo de análise de domínio e procura refletir uma visão consensual sobre a realidade que se pretende representar [...] é feita por meio de diferentes tipos de sistemas de organização do conhecimento (SOC) que são sistemas conceituais que representam determinado domínio por meio da sistematização dos conceitos e das relações semânticas que se estabelecem entre eles (BRÁSCHER e CAFÉ, 2008 p.6 e 8).

Contudo, o objetivo da atual pesquisa é de esclarecer que dentro da ciência da informação, e em especial para o profissional da área, a organização da informação é processo crucial no tratamento de dados, indexação e disseminação da informação. Por fim, destaca-se que a organização da informação dentro do contexto apresentado até o momento, trata-se do processo de descrição dos objetos informacionais, visando trazer representações a partir das análises. E na busca pela transformação de dados em informação, esse processo é indiscutivelmente necessário, não somente pela recuperação da informação, mas pela disseminação dela. Nesse contexto, o papel do profissional de Biblioteconomia (área de estudo dentro da ciência da informação) se torna indispensável em especial pelos conhecimentos em análises de informações que serão tratados mais à frente neste trabalho.

4.3 Análise da Informação

A Análise da Informação - AI pode ser definida através de Lancaster (2004), como uma forma de identificar um conteúdo descritivo, temático e informativo, tornando-se um conjunto de procedimentos normalmente destinado a expressar documentos bibliográficos. Similarmente, a expressão de origem francesa “*analyse documentaire*” (análise documentária) originária do campo da Documentação “[...] é definida como um conjunto de procedimentos efetuados com o fim de expressar o conteúdo de documentos, sob formas destinadas a facilitar a recuperação da informação” (CUNHA, 1987, p.38). A CI entende que é um conjunto de ações e

operações intelectuais que trazem uma interpretação e seleção de conteúdos dos meios que estão sendo analisados. Essas representações podem ser feitas de diversas formas, utilizando linguagens naturais ou controladas (CESARINO, 1985; LANCASTER, 2004), já os instrumentos utilizados para a realização de uma análise são diversos, desde sistemas básicos não necessariamente criados para este fim, como sistemas automatizados e complexos, cada um com uma especialidade e/ou peculiaridade.

A Ciência da Informação, através de cursos como Biblioteconomia e Arquivologia, tem trabalhado na conduta de acessibilizar a informação por meio da descrição e representação. Entende-se que no momento em que um documento é analisado, interpretado, descrito e representado ele se torna acessível ao público alvo ou um público geral, a depender da forma de disponibilização dessa informação.

É possível identificar que o processo de análise, se estabelece em 3 níveis. O primeiro é o de identificação, que podemos chamar de nível intuitivo, que é quando percebe-se a interação do objeto de análise com o meio físico, a percepção do senso comum com o objeto e o ambiente. Um exemplo disso seria que em dias frios se deve utilizar um casaco. O casaco neste caso é o objeto de análise, o segundo nível podemos chamar de nível racional, é o momento em que o indivíduo que analisa, deve intermediar as situações com um certo nível de racionalidade. O terceiro nível é o que podemos chamar de profissional, é nesse momento que é necessário assumir uma postura mediadora entre os que produzem e os que consomem a informação com o objetivo de facilitar o acesso à informação.

De forma genérica, o objetivo da análise é apenas disponibilizar de forma democrática, por meio do uso de linguagem documentária previamente construída, a descrição, representação e organização da informação, independentemente de seu suporte para tornar disponível ao usuário, e quem mais a procura, utilizando algumas ações preliminares básicas para este processo, que se iniciam com ações de reconhecimento e descrição do documento, ações de representação por meio de indexação e processamento, e ações complementares a depender da tomada de decisão de quem analisa.

Contextualizando a linguagem documentária, é uma linguagem construída de forma artificial e controlada, que visa constituir sistemas de busca e armazenamento

de documentos por meio do vocabulário controlado, para facilitar a indexação e recuperação da informação.

Com a explosão informacional a partir da criação da web e o desenvolvimento da internet, criou-se a necessidade de ampliar algumas áreas do conhecimento para lidar com o imenso volume de dados produzidos diariamente. Tem sido desafiador desde então atender as demandas de tratamento destes dados e com isso surge então a Ciência de dados (*data science*), como uma nova ciência cujo propósito é “identificar os princípios, métodos e técnicas fundamentais para o gerenciamento e análise de grandes volumes de dados heterogêneos” (PORTO; ZIVIANI, 2014).

No entanto, deve-se destacar a diferença entre a ciência de dados e a análise de dados, sendo que a ciência de dados demonstra ser um termo mais abrangente que dispõe de diversos métodos para conseguir juntar informações. A ciência de dados trata, de uma maneira mais ampla, da coleta de dados para que por fim sejam úteis à instituição. E a análise de dados destaca-se por ser uma parte mais específica, buscando, de acordo com suas ferramentas, uma maneira de disponibilizar os dados em um suporte específico para cumprir o seu propósito. (CETAX, 2022).

De acordo com Amaral (2016) a ciência de dados define-se através de processos que transcorre através do seu ciclo de vida, podemos visualizar através da figura 4.

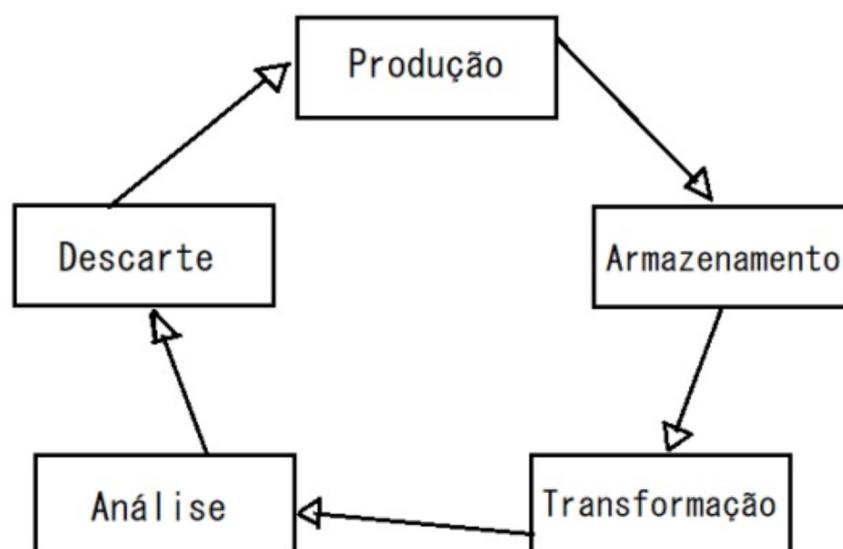


Figura 4 - O ciclo do dado
Fonte: Amaral (2016, p.6)

Diante do exposto nos parágrafos acima, elucida-se a necessidade de um profissional capaz de lidar com dados e informações com destreza. Observando como a ciência de dados, enquanto matéria, utiliza de ciências básicas em suas técnicas e teorias, é possível então defini-la como multidisciplinar, sendo assim o profissional da ciência da informação se torna tão somente útil e capacitado.

Em 2008, D. J. Patil e Jeff Hammerbacher, líderes das iniciativas de análise de dados no *LinkedIn* e *Facebook*, firmaram o termo “Cientista de Dados” para os profissionais que lidavam com as descobertas do mundo da *big data*.

A função do cientista de dados começou a se difundir nas organizações no final dos anos 2000, principalmente na região da Baía de São Francisco, onde as empresas se voltaram para a exploração de dados provenientes da internet e das mídias sociais [...]. As empresas acumulavam dados e mais dados, e começou-se a desenvolver um enorme variedade de novas tecnologias (muitas delas inventadas por cientistas de dados) para armazenar, processar e analisar esses dados. (DAVENPORT, 2014, p.85).

Com o entendimento do papel do cientista de dados e da análise da informação, como consequência surge a necessidade de fundamentar a análise de dados. Santos (2016) explica que para analisar um dado, é necessário que este tenha sofrido uma transformação para a resolução de um problema. Nesse contexto, destaca-se que a partir do conceito supracitado, existem 3 tipos de análises, a exploratória, a implícita e a explícita.

As análises implícitas não deixam a informação em evidência, neste caso o conjunto de dados precisará ser avaliado de uma forma mais profunda, qual seja com mecanismos mais complexos. Nas análises explícitas, Amaral (2016) coloca que as informações são apresentadas de forma mais clara ainda no conjunto de dados, neste caso o conhecimento fica em evidência de forma mais simples, sem necessidade de complexidade na hora de reproduzir a informação.

O termo análise exploratória surgiu com John Tukey em 1977, o estatístico define em sua obra *Exploratory Data Analysis*, que o procedimento para este tipo de análise, se dá por meio de técnica de interpretação de resultados a partir de um procedimento, de coleta de dados matemáticos. As definições de Tukey (1961) ainda são consideradas até os dias de hoje, como sendo uma definição básica para esse tipo de análise.

No método de análise explícita, a informação já foi percebida e o conhecimento já reconhecido no momento em que se analisa o dado. Isso quer dizer que a informação se torna clara de forma mais simples, sem necessidade de grandes interferências. Ou seja, nesse modelo, poucos esforços já geram bons resultados, pesquisas automatizadas por exemplo, já se mostram eficientes quanto aos resultados.

Num contexto em que a informação é valorizada, e muito utilizada para diversos fins e concordando com Baptista, Araújo Júnior e Carlan (2010, p.63), observa-se que, ainda que de um modo geral, as pessoas acabam se tornando uma espécie de analista da informação. Isto porque como já aludido anteriormente, esses processos passam pelos níveis de intuição, racionalização e profissional como já visto pela perspectiva dos autores.

Na perspectiva da Biblioteconomia, os autores elucidam que esse processo de análise da informação é “característico de algum tipo de mediação especializada que se torna necessária entre produtores e usuários da informação”, com a intenção de proporcionar acesso à informação. Guimarães (2003) é retilíneo ao salientar que apesar de fazer parte da natureza humana analisar coisas, é preciso se lembrar que não basta seguir o bom senso ou a intuição quando se trata de analisar uma informação que deverá ser disseminada. No contexto profissional, e nessa linha estamos colocando o profissional da Biblioteconomia como foco, deve-se seguir etapas analíticas e sintéticas para este objetivo.

Na etapa analítica, o profissional deverá aplicar técnicas de leituras para que seja possível identificar o que é relevante a depender do objetivo, e aplicar “[...] ao documento um conjunto de categorias conceituais, visando à construção de enunciados de assunto.” (GUIMARÃES, 2003, p.112), desse modo atendendo ao seu usuário de forma clara e objetiva.

A etapa sintética é o momento de aplicar o que Guimarães (2003) chama de classificar os assuntos por “principais, secundários e ordenados logicamente, tendo como parâmetros a estrutura, a função e os usos (tipo de busca informacional a que se presta o documento).” desse modo condensando para que seja possível representar. Essa representação, como bem aplicado pelo autor, é uma espécie de “tradução do conteúdo temático de documento em linguagem de indexação [...]” (GUIMARÃES, 2003, p.112-113).

Diante do exposto, percebe-se que a atuação deste profissional é de extrema importância como mediador entre a informação e o seu usuário, dada a complexidade do processo, é conveniente que quem entenda da ciência da informação seja capaz de traduzi-la de forma sucinta, objetiva e democrática.

4.4 Ciência de Dados e Dados Abertos

A palavra dados tem sido muito utilizada atualmente, na era da tecnologia é uma palavra que se transformou em uma coisa cotidiana. Uma pesquisa feita pela *Strategy Analytcs* em 2021, sobre smartphones, estima que pelo menos metade do mundo todo possui um smartphone. A cada momento em que alguém utiliza o seu smartphone, dados novos são criados, de todo lugar do mundo. A cada vez que algum produto é criado e identificado com um código QR ou de barras, mais dados são criados e armazenados. Isso cria uma enxurrada de dados que são passíveis de serem analisados, processados e transformados em informações. O uso de tecnologias novas para transformar esses dados em informações, tem se mostrado relevantes e necessárias no processo de democratização da informação, por meio da disponibilização de forma aberta e livre desse conteúdo, esses recursos são capazes de gerar capacitações e formar pessoas que podem a partir disso produzir novos negócios, produtos e soluções para diferentes áreas da sociedade.

Contextualizando dados, é necessário conhecer o movimento *Big Data* que nasceu em 1990 na *National Aeronautics and Space Administration (NASA)* e foi utilizado como uma forma de descrever grandes conjuntos de dados que à época, era uma dificuldade da rede de computadores.

Para Pires (2015) nem todos os dados podem ser considerados dados abertos.

Todo dado que é público deve ser aberto, mas nem todo dado é público. A legislação brasileira trata como exceção à abertura dados particulares, capazes de identificar indivíduos, ferir seu direito à privacidade ou sua honra, dados considerados sigilosos ou dados que possam comprometer a segurança nacional. (PIRES, 2015, p.11).

Tim Berners-Lee, o inventor da Web, propôs que os dados abertos podem ser classificados por uma escala baseada em estrelas (BERNERS-LEE, 2006). Com base nessa classificação, como mostra a figura 5. Quando um dado é publicado, em

qualquer formato e disponibilizado através de uma licença que permita sua utilização e reutilização sem restrição nenhuma, ela passaria a ser avaliada com *1 Estrela*. Embora já seja um grande avanço encontrar hoje em dia materiais disponíveis com *1 Estrela*, esses dados necessitam ainda de uma manipulação manual ou através de sistemas específicos para permitir o acesso aos dados.

Assim que os dados são disponibilizados num formato que permite o processamento deste, de forma automatizada por software, seja ele aplicativo de criação de planilhas, ao invés de estar em formato de imagens, ele passa a ser classificado com *2 Estrelas*, e podemos dizer que são dados em formato estruturado.

Podemos dizer que os dados passam a ter um formato estruturado e serem abertos, quando são publicados em um tipo de formato conhecido como não proprietário, que significa que não possui uma cobertura de *copyright* ou de patentes, como por exemplo no formato *Comma Separated Values* ou Valores separados por vírgula - CSV (diferente do formato do Excel que possui uma formatação, os arquivos em *Extensible Style Language* - XLS costumam ter campos separados por cédulas, ao invés de separação por vírgula como é no CSV). Nesse caso esse nível recebe *3 Estrelas*, porque assim esse tipo de formato não depende integralmente de outros programas ou softwares para serem lidos. Dados que estão em nível de *3 Estrelas* são os mais comuns na esfera governamental de publicação espontânea, pois os usuários podem baixar a qualquer momento e manipular da forma que desejar, sem muito esforço com relação a programas de computadores.

A partir da quarta estrela, passamos a considerar o conceito de dados conectados, que precisa, necessariamente, abranger todas as outras estrelas, mas, além de ser necessário ter disponível como documento, os dados precisam estar disponíveis na Web, utilizando URI's, que significa *Uniform Resource Identifier* em tradução livre significa Identificador Uniforme de Recurso. É uma sequência de caracteres que identificam recursos da internet. Um dos mais comuns é a URL (Localizador de Recurso Universal) que faz a comunicação entre um local de hospedagem e um determinado recurso. A vantagem de ter um dado com 4 estrelas, é a facilidade de poder manipular esses dados a partir de sistemas conectados, assim diferentes partes dos dados podem ser usados em combinação com outros dados também disponíveis em rede.

O próximo nível, e último, da hierarquia do sistema de estrelas, são as 5 estrelas, neste caso, os dados publicados precisam estar conectados a outras bases de dados e publicados nas mesmas condições, assim é possível explorar melhor o potencial dos dados publicados, agregando cada vez mais valor a eles.

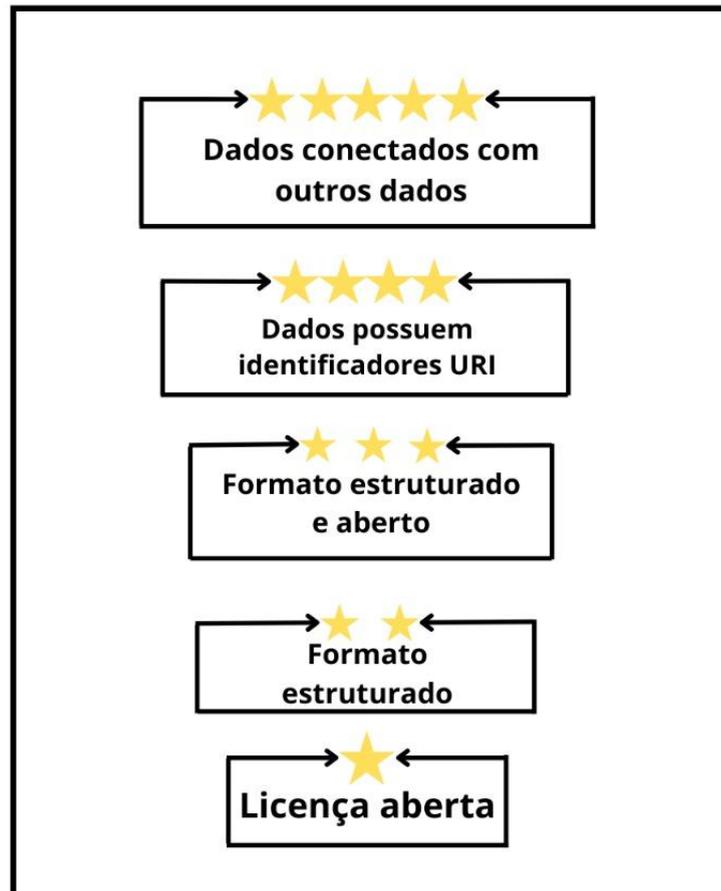


Figura 5 - Hierarquia das estrelas
Fonte: Elaborado pela autora.

De acordo com a *Open Knowledge Foundation* (atualmente conhecida como *Open Knowledge International*), dados são considerados abertos quando qualquer pessoa pode livremente acessar, modificar e utilizar independente da finalidade, sujeito aos termos das leis cabíveis como por exemplo, a de autoria. Basicamente eles precisam ser uma espécie de domínio público.

O ativista, também especialista em políticas públicas, David Eaves, em 2009 propôs as seguintes “leis” que regem os dados abertos que foram aceitas pela comunidade e pelo W3C¹²:

1. Se o dado não pode ser encontrado e indexado na Web, ele não existe;
2. Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado;
3. Se algum dispositivo legal não permitir sua replicação, ele é inútil.

Nesse quesito a W3C vem propondo uma forma de fazer com que a web atinja seu potencial, ao máximo, por meio da criação de diretrizes. Foi em 2007 que os primeiros indícios de novos caminhos para os dados abertos surgiram. Um grupo de ativistas em dados na web, um grupo de americanos decidiram criar um projeto que colocaria os dados abertos em ascensão utilizando a premissa das informações governamentais. O projeto era uma forma de fazer com que o governo pudesse de forma simples disponibilizar dados para a população, através de 8 princípios (TCU, 2015):

1- Completos: todos os dados públicos estão disponíveis. Entende-se por dado público o dado que não está sujeito a limitações válidas de privacidade, segurança ou controle de acesso.

2- Primários: os dados são apresentados tais como os coletados na fonte, com o maior nível de granularidade e sem agregação ou modificação.

3- Atuais: os dados são disponibilizados tão rapidamente quanto necessária à preservação do seu valor.

4- Acessíveis: os dados são disponibilizados para o maior alcance possível de usuários e para o maior conjunto possível de finalidades.

5- Compreensíveis por máquinas: os dados são razoavelmente estruturados de modo a possibilitar processamento automatizado.

6- Não discriminatórios: os dados são disponíveis para todos, sem exigência de requerimento ou cadastro.

7- Não proprietários: os dados são disponíveis em formato sobre o qual nenhuma entidade detenha controle exclusivo; e

8- Livres de licenças: os dados não estão sujeitos a nenhuma restrição de direito autoral, patente, propriedade intelectual ou segredo industrial. As restrições sensatas relacionadas à privacidade, segurança e privilégios de acesso devem ser permitidas.

(GOVERNO DIGITAL,

<https://www.gov.br/governodigital/pt-br/dados-abertos>)

¹² O **World Wide Web Consortium (W3C)** é a principal organização de padronização da World Wide Web. Consiste em um consórcio internacional com 450 membros, agrega empresas, órgãos governamentais e organizações independentes com a finalidade de estabelecer padrões para a criação e a interpretação de conteúdos para a Web.

O Tribunal de Contas da União (TCU), enquanto órgão de controle externo do Governo Federal, publicou em 2015, cinco motivos para demonstrar os benefícios da implementação da política de dados abertos governamentais no Brasil, são eles:

1. Porque a sociedade exige mais transparência na gestão pública;
2. Porque a própria sociedade pode contribuir com serviços inovadores ao cidadão;
3. Porque ajuda a aprimorar a qualidade dos dados governamentais;
4. Para viabilizar novos negócios; e
5. Porque é obrigatório por Lei (Lei n.12.527/2011). (TCU, 2015, [Dados Abertos — Governo Digital](#))

Assim, os dados abertos governamentais passaram a ser considerados uma metodologia para nortear a publicação de dados do governo no que podemos chamar de formato reutilizável, ou seja, está disponível para a população usar para pesquisas, consultas pessoais e projetos afins. Apesar disso, do ponto de vista de um profissional da informação, o formato que os dados são disponibilizados, não são acessíveis e nem democráticos, pois são publicados em forma de balanço e relatório, gerado nos sistemas da administração pública. Mesmo que seus objetivos sejam a transparência do governo para com o cidadão, inclusão da população nas ações governamentais e responsabilidade, em sua maioria fiscal, dos chefes de estado e outros, há ainda a necessidade de destacar que se os dados não são tratados e apenas exposto numa página da web, ele não é uma informação visto que consideramos dado qualquer elemento identificado em sua forma bruta, que sozinho não traz compreensão do fato ou situação determinada e, que Informação pode ser definido como a junção de um conjunto de dados e/ou conhecimentos dispostos de forma organizada, que possam compor certas referências sobre um determinado acontecimento, ocorrência ou fenômeno. Informação é também um resultado sobre os dados, após ser processado, tratado, organizado e etc. é um conhecimento que deve ser transmitido através de um suporte.

Para além das necessidades governamentais, os dados abertos são muito necessários dentro da ciência, para isso devemos considerar o conceito de acesso aberto, que quer dizer que os resultados de pesquisas científicas estarão disponibilizados de modo claro, simples e sem limitações. Eles podem ser disponibilizados de diversas formas e sobre diferentes publicações, como artigos científicos, documentos relacionados a conferências nacionais e internacionais,

teses e dissertações, capítulos de livros, entre outros, de forma gratuita ou não. É considerado gratuito quando o acesso é online, gratuito e livre, ou com alguns direitos adicionais de utilização através de licenças Creative Commons (CC)¹³. De uma maneira geral, essas questões estão ligadas ao que podemos chamar de ciência aberta, que tem como preocupação tornar a pesquisa e todas as atividades intrínsecas a ela uma atividade cada vez mais transparente. Os pressupostos dessas atividades devem ter como preocupação o uso de softwares e sistemas que podem inibir a interoperabilidade e de compartilhamento.

A LAI, foi sancionada em 2011 regulamentando constitucionalmente o direito, garantindo ao cidadão brasileiro o acesso às informações consideradas públicas aplicadas aos três poderes da União, dos estados e do Distrito Federal e dos municípios, em paralelo a isso, garante igualmente o dever dos órgãos e entidades públicas, de disponibilizar alguns dados dentro do mínimo determinado, de forma espontânea, além de atender às solicitações advindas do portal. A sanção desta lei representa para o povo um grande avanço na consolidação da democracia, fortalecendo a comunicação entre população e governo através da transparência. Uma das formas de obter esses dados é através do Portal da Transparência dos sites de onde se deseja obter informações, ou pelo Sistema de Informação ao Cidadão (e-SIC).

Após entender onde as tecnologias e a ciência se encontram, é necessário conhecer o *Data Science* (do inglês Ciência de Dados). Este nome anda lado a lado com o *Big Data*, que já fora apresentado anteriormente, mas conhecer as suas diferenças é importante para compreender o papel do bibliotecário nesse sentido. Conhecido por envolver um conjunto de conhecimentos e/ou habilidades que relacionam métodos e técnicas de Estatística e tecnologias (em especial a Ciência da Computação). Na atualidade, surge o termo *e-Science* que pode ser usado para se referir ao uso conjunto de tecnologia computacional nos trabalhos de pesquisas científicas. Essas tecnologias e habilidades se mostraram eficazes e importantes no processo de representação da informação no uso de dados para tratamento da informação (HEY; HEY, 2006).

¹³ São licenças públicas que permitem a distribuição gratuita de obras protegidas por direitos autorais. Esse direito refere-se ao poder de utilizar, compartilhar, construir algo sobre um trabalho criado por outro autor.

Ao relembrar os impactos do *Big Data*, Brietman (2014) afirma que apenas 1% dos dados gerados são realmente analisados, e nesse sentido a Organização e a Análise da Informação se tornam termos essenciais nessa discussão. Para um Bibliotecário esses estudos fazem parte do dia a dia e são completamente naturais. No momento em que fica claro que, o que muda na atuação de um bibliotecário no que concerne aos dados, são a forma como a informação é registrada, a transformação de dados em informação se torna natural. No caso dos dados, a informação é processada e armazenada no computador já que Dado é uma entidade matemática e portanto, puramente sintática (SETZER, 1999).

Partindo desse contexto, verifica-se que “os dados são informação em potencial, que somente são percebidos por um receptor se forem convertidos em informação [...]” (FERNÁNDEZ-MOLINA, 1994, p.328). Fica claro, portanto, que há uma interligação entre dados, informação e conhecimento. Isso significa que existe uma dinâmica entre esses conceitos que pode ou não ser evoluída para uma sequência cíclica. Além disso, ainda no contexto matemático, o diagrama de Venn¹⁴ apresentado na figura 6, pode ser capaz de explicar o relacionado entre a biblioteconomia e os estudos dos dados.

¹⁴ John Venn (1834-1923) responsável pelo desenvolvimento da lógica matemática de Boole, como forma de representar as intersecções e uniões de conjuntos matemáticos através de um diagrama que leva seu nome.

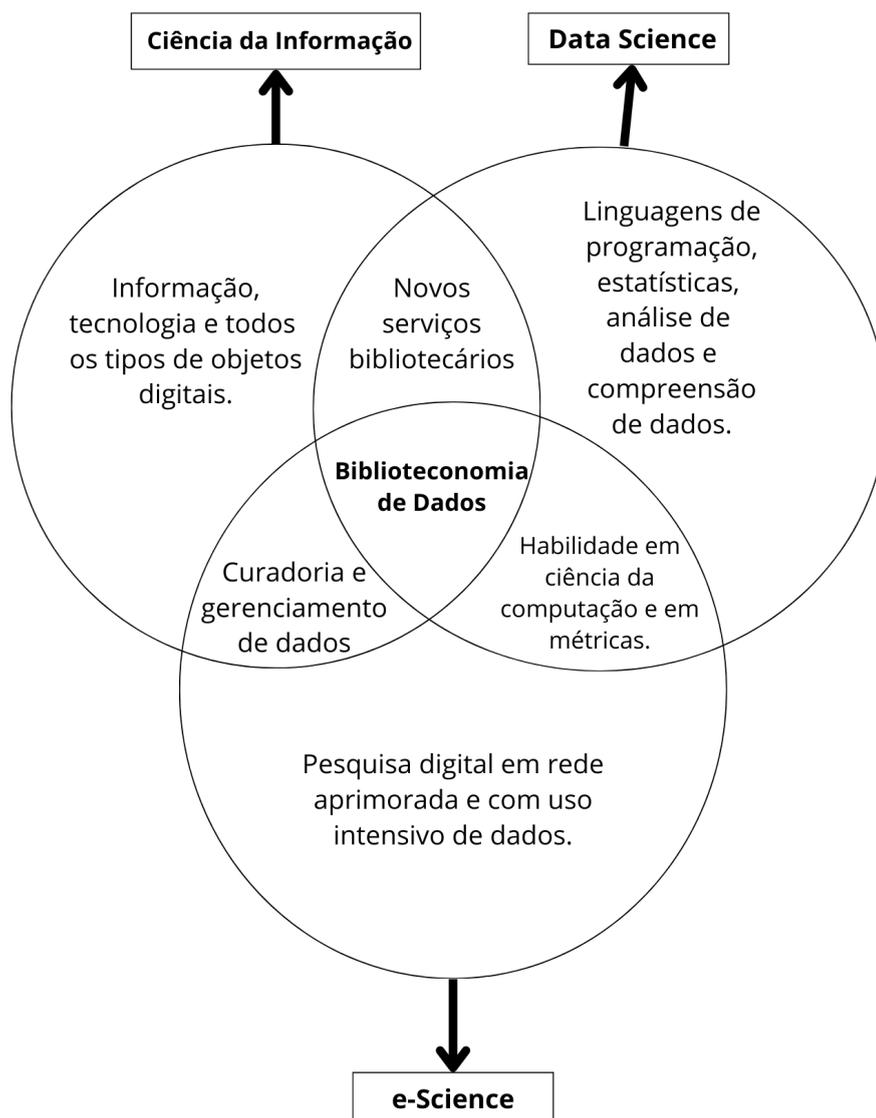


Figura 6 - Diagrama de Venn
Fonte: Adaptado de Semeler and Pinto (2019b).

Para essa pesquisa, não se faz necessário falar detalhadamente de todas as facetas apresentadas pela análise. A figura 6 ajuda a construir a relação entre os campos de estudo, e colabora na criação da visão de que o bibliotecário é capaz de atuar de forma interseccional entre os três campos. Diante do exposto, a partir da perspectiva de dados apresentada anteriormente e do que foi entendido como informação e conhecimento, surge então nesse contexto um profissional da informação que seja capaz de orientar os dados até que tornem um conhecimento, e sendo capaz de criar e gerenciar produtos destinados a lidar com essa questão, visto que lidar com fluxos de informação faz parte da sua capacidade nata enquanto

profissional e “a articulação dos planos de gestão de dados é simplesmente uma manifestação moderna destas funções” (CORRÊA, 2016, p.391). E por isso o bibliotecário de dados surge como detentor do conhecimento necessário para interpretar os dados e transmiti-los para a sociedade, como uma forma de democratização da informação com segurança de compartilhamento e garantido a disseminação da informação e exercendo o seu papel fundamental de um ser informacional.

5 METODOLOGIA

A metodologia é indispensável para o desenvolvimento de uma pesquisa científica, segundo Demo (1995) a metodologia significa o “estudo dos caminhos, dos instrumentos usados para se fazer ciência. [...] ao mesmo tempo que visa conhecer caminhos do processo científico, também problematiza criticamente, no sentido de indagar os limites da ciência, seja como referência à capacidade de conhecer, seja com referência à capacidade de intervir na realidade”. Desse modo, através de métodos e técnicas que permitem tornar os processos de pesquisa como planejamento, organização, análise entre outros, mais confiáveis por seguirem protocolos pré-estabelecidos para se seguir e compor um bom trabalho.

Podemos analisar Microdados e Dados por uma perspectiva de funil, o termo mais geral pro mais específico. Microdados se refere especificamente a conjuntos de dados que oferecem informações detalhadas sobre unidades individuais. Dados é um termo amplo que pode incluir informações em qualquer nível de detalhe. A escolha entre dados e microdados depende dos objetivos da análise e da profundidade de detalhe necessária para responder às perguntas de pesquisa.

Foi utilizado os microdados do Enem 2022 que são disponibilizados pelo o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), como objeto de análise dessa pesquisa, portanto será conduzida uma abordagem qualitativa no presente estudo. Os microdados são disponibilizados por meio de documento em formato de tabelas, contendo diferentes dados que são coletados pela entidade realizadora dos trâmites da criação e aplicação da prova. Esses dados serão analisados pela perspectiva da biblioteconomia, utilizando os recursos bibliográficos disponíveis. Nessa abordagem, a população escolhida é o próprio microdado.

A pesquisa exploratória-descritiva emerge como uma estratégia valiosa que combina elementos de exploração e descrição para proporcionar uma compreensão mais profunda e abrangente de determinado fenômeno, neste trabalho é explorado a ciência da informação em suas inúmeras facetas e como ela se aplica nos estudos dos dados. A busca da bibliografia apresentada, se dá a partir da longa caminhada no curso de graduação em biblioteconomia, além de utilização de palavras chaves relacionadas à Ciência da Informação e Ciência de Dados.

A análise inicial consiste em separar quais dados são relevantes para desenvolver um apanhado geral de informações para que também seja feita uma análise quantitativa dos dados escolhidos após o tratamento, para que possa nortear a criação de gráficos que serão apresentados num formato de *dashboard* que será desenvolvido através do sistema de análise *Power BI*. Essa forma de separação é onde os microdados passam a se tornar dados efetivamente.

A ferramenta *Microsoft Power BI* é um software de análise de dados e de negócios, criado pela empresa Microsoft em 2015, tem como objetivo facilitar a criação de interfaces, também chamadas de *dashboards*, que fornecem uma espécie de visualização interativa de gráficos, tabelas, mapas e outros recursos visuais, possibilitando também a criação de relatórios personalizados dos dados que estão disponíveis no recurso, por meio do que chamamos de Inteligência de Negócios (do inglês *Business Intelligence* - BI), termo utilizado para traduzir o processo de coleta, organização, análise, compartilhamento e monitoramento de informações que são utilizadas para dar suporte à tomada de decisões dentro de grandes empresas. Essa tecnologia ganhou seu espaço por ser capaz de suportar um volume de dados muito superior a outras plataformas e *softwares* de controles mais básicos, principalmente pela sua possibilidade de demonstração em tempo real das mudanças nos cenários baseado na forma como os dados são manipulados dentro da plataforma. É utilizado majoritariamente pelas áreas de finanças, tecnologias, engenharia e marketing, mas tem ganhado muita força em áreas como a da saúde. Apesar de seu maior sucesso ser entre os profissionais das áreas de finanças, essa ferramenta possui grande potencial na divulgação de dados de qualquer natureza, visto que a sua finalidade é a de analisar dados estruturados. Diferente de outras ferramentas que possuem recursos muito específicos, o *Microsoft Power BI* tem funcionalidades muito genéricas que podem ser utilizadas em qualquer área do conhecimento. A versão utilizada dessa ferramenta é a versão *desktop* com atualização de outubro de 2023 (2.122.442.0).

6 ESTUDO DE CASO

Neste capítulo, serão apresentados os dados que foram obtidos através da análise dos microdados do Enem. Os dados serão exibidos conforme a análise qualitativa e quantitativa com uma breve descrição com o intuito de desenvolver o tema e as informações adquiridas.

Buscando a melhor maneira de deixar os dados a serem compreendidos de formas simples e sucintas, criou-se uma tabela dinâmica a partir da tabela já disponível no site do Inep, separando dados que, de acordo com a tomada de decisão desta pesquisadora, seriam mais relevantes para demonstrar dados reais.

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) é uma autarquia Federal vinculada ao Ministério da Educação (MEC), criada em 1937 antes com o nome de **Instituto Nacional de Pedagogia**, que é responsável por promover estudos, pesquisas e avaliações periódicas relacionada ao sistema educacional brasileiro, subsidiando a implementação de políticas públicas na área da educação. Realiza também levantamentos estatísticos nas modalidades de ensino de uma forma geral, as avaliações mais conhecidas que são de responsabilidade dessa entidade são o Enem e Enade (Exame Nacional de Desempenho dos Estudantes), mas além disso realiza o censo escolar em todos os anos do ensino desde o básico até o superior, contemplando desde as instituições públicas às particulares.

O Exame Nacional do Ensino Médio (Enem) é uma prova facultativa, destinada aos estudantes que já concluíram o ensino médio, podendo ser realizada a qualquer momento a partir dessa conclusão, que foi aplicada pela primeira vez em 1997. As primeiras 10 edições do exame tinham a finalidade exclusiva de avaliar o nível de habilidade dos concluintes do ensino médio e avaliar o nível em que estavam as escolas do país. Somente a partir de 2009 é que a prova passou a ser utilizada como uma forma de incentivo para o estudante ingressar ao ensino superior através do Sistema de Seleção Unificada (SISU). O programa possibilitou inúmeros ingressos às universidades públicas, incentivando a mobilidade inclusive de estudantes oriundos de locais mais interioranos que se deslocaram para as grandes capitais para ocupar vagas em diversos cursos. Apesar de ser uma forma de avaliação do progresso das escolas da rede pública de ensino, não é uma avaliação gratuita. Também é possível ser utilizada para pleitear bolsas em

instituições particulares, a depender da instituição, já que é uma prática facultativa às entidades particulares.

É uma prova interdisciplinar mais uma redação, que ocorre todos os anos desde então. Até 2016, as provas eram realizadas em um único fim de semana, normalmente no domingo. A partir de 2017, o formato de aplicação foi modificado sendo realizado então em 2 domingos, com o intervalo de uma semana. Essa mudança se deu a partir de uma consulta pública realizada pelo Inep em busca de dar voz ao pedido dos jovens e também de todos os responsáveis pela realização do exame, como professores e avaliadores. O exame realizado no ano de 2022 apresentou no primeiro dia de prova 90 questões das áreas de Ciências Humanas, Linguagens e Códigos e uma redação, já no segundo dia as áreas avaliadas foram de Matemática e Ciências da Natureza que engloba física, química e biologia, também com 90 questões.

Um dos procedimentos para a realização da prova é a coleta de informações socioeconômicas dos participantes, que são coletados no momento da sua inscrição para a prova através de um questionário com 25 perguntas. E após a realização dela, novos dados são adicionados às planilhas. As perguntas variam de informações financeiras à escolaridade dos pais. A relevância desses dados depende dos objetivos traçados na avaliação deles,

Todo ano, desde 1998, o site do Inep disponibiliza em formato aberto os dados coletados ao longo do processo de inscrição até a realização das provas do Enem. O objeto de estudo desta pesquisa foram os dados coletados no ano de 2022. Os dados foram analisados e tratados. Separando os dados considerados relevantes para subsidiar tomadas de decisões e descartando os considerados irrelevantes. O Enem do ano de 2022, contou com um total de **3.476.105** inscritos.

Nas próximas figuras é possível ver as perguntas que são feitas no questionário socioeconômico aplicadas no ano de 2022.

Figuras 7 e 8 são aspectos escolares relacionados aos responsáveis pelo aluno.

Figura 7 - Pergunta 1.**1- Até que série/ano seu pai, ou o homem responsável por você, estudou?**

- A)** Nunca estudou.
- B)** Não completou a 4ª série/5º ano do ensino fundamental.
- C)** Completou a 4ª série/5º ano, mas não completou a 8ª série/9º ano do ensino fundamental.
- D)** Completou a 8ª série/9º ano do ensino fundamental, mas não completou o ensino médio.
- E)** Completou o ensino médio, mas não completou a faculdade.
- F)** Completou a faculdade, mas não completou a pós-graduação.
- G)** Completou a pós-graduação.
- H)** Não sei.

Fonte: Elaborado pela autora com base nos microdados

Figura 8 - Pergunta 2.**2- Até que série/ano sua mãe, ou a mulher responsável por você, estudou?**

- A)** Nunca estudou.
- B)** Não completou a 4ª série/5º ano do ensino fundamental.
- C)** Completou a 4ª série/5º ano, mas não completou a 8ª série/9º ano do ensino fundamental.
- D)** Completou a 8ª série/9º ano do ensino fundamental, mas não completou o ensino médio.
- E)** Completou o ensino médio, mas não completou a faculdade.
- F)** Completou a faculdade, mas não completou a pós-graduação.
- G)** Completou a pós-graduação.
- H)** Não sei.

Fonte: Elaborado pela autora com base nos microdados

É natural que ao mudar as gerações, o nível de escolaridade numa média vá se modificando. O ambiente familiar, não somente em relação ao desempenho escolar, está diretamente ligado ao desenvolvimento de crianças e adolescentes, e isso se reflete na vida adulta e profissional desses indivíduos. Barros et al (2006), classifica também que a renda familiar é outro importante fator nesse quesito.

Diante disso, apresenta-se as perguntas da segunda seção, que são informações relacionadas aos aspectos financeiros do estudante e de seu núcleo familiar, além de perguntas que envolvem detalhes físicos da residência, como a posse de certos tipos de eletrodomésticos e veículos. Para esta pesquisa, não se fazem necessárias avaliar e cruzar todos os dados socioeconômicos, portanto apresenta-se apenas as consideradas relevantes.

A figura 9, é uma pergunta relacionada à quantidade de pessoas que integram o núcleo familiar do estudante.

Figura 9 - Pergunta 5.

5- Incluindo você, quantas pessoas moram atualmente em sua residência?

- | | |
|------------------------------------|----------------------|
| A) 1, pois moro sozinho(a). | K) 11 pessoas |
| B) 2 pessoas | L) 12 pessoas |
| C) 3 pessoas | M) 13 pessoas |
| D) 4 pessoas | N) 14 pessoas |
| E) 5 pessoas | O) 15 pessoas |
| F) 6 pessoas | P) 16 pessoas |
| G) 7 pessoas | Q) 17 pessoas |
| H) 8 pessoas | R) 18 pessoas |
| I) 9 pessoas | S) 19 pessoas |
| J) 10 pessoas | T) 20 pessoas |

Fonte: Elaborado pela autora com base nos microdados

Já a figura 10 apresenta as opções de faixa de renda dos estudantes

Figura 10 - Pergunta 6.

6- Qual é a renda mensal de sua família? (Some a sua renda com a dos seus familiares.)

- | | |
|---|---|
| A) Nenhuma renda. | I) De R\$ 6.060,01 até R\$ 7.272,00. |
| B) Até R\$ 1.212,00 | J) De R\$ 7.272,01 até R\$ 8.848,00. |
| C) De R\$ 1.212,01 até R\$ 1.818,00. | K) De R\$ 8.848,01 até R\$ 9.696,00. |
| D) De R\$ 1.818,01 até R\$ 2.424,00. | L) De R\$ 9.696,01 até R\$ 10.908,00. |
| E) De R\$ 2.424,01 até R\$ 3.030,00. | M) De R\$ 10.908,01 até R\$ 12.120,00. |
| F) De R\$ 3.030,01 até R\$ 3.636,00. | N) De R\$ 12.120,01 até R\$ 14.544,00. |
| G) De R\$ 3.636,01 até R\$ 4.848,00. | O) De R\$ 14.544,01 até R\$ 18.180,00. |
| H) De R\$ 4.848,01 até R\$ 6.060,00. | P) De R\$ 18.180,01 até R\$ 24.240,00. |
| | Q) Acima de R\$ 24.240,00. |

Fonte: Elaborado pela autora com base nos microdados

A figura 11 apresenta-se a penúltima questão que é relacionada à posse e quantidade de computadores na residência do candidato.

Figura 11 - Pergunta 24.

24- Na sua residência tem computador?

- A) Não.
- B) Sim, um.
- C) Sim, dois.
- D) Sim, três.
- E) Sim, quatro ou mais.

Fonte: Elaborado pela autora com base nos microdados

A figura 12 apresenta uma questão sobre acesso à internet na residência do aluno.

Figura 12 - Pergunta 25.

25- Na sua residência tem acesso à Internet?

- A) Não.
- B) Sim.

Fonte: Elaborado pela autora com base nos microdados

O questionário socioeconômico se torna um instrumento de grande importância na hora de traçar o perfil dos alunos que são focos do exame. É um aliado na busca por melhores condições de aplicabilidade das provas, e é importante para identificar qual o aspecto em comum que a maioria realizadora do exame possui. Porém, por não ser uma método passível de fiscalização, já que as respostas são de responsabilidade dos estudantes e muitas respostas não são obrigatórias, não é possível verificar a autenticidade das informações, ou se os estudantes de fato respondem ao questionário com o auxílio dos responsáveis, aconselha-se que quaisquer resultados advindo dessas respostas, sejam amparados por dados que possam ser mais consistentes. Contudo, as perguntas escolhidas para compor a análise desse estudo, são relevantes para traçar um perfil

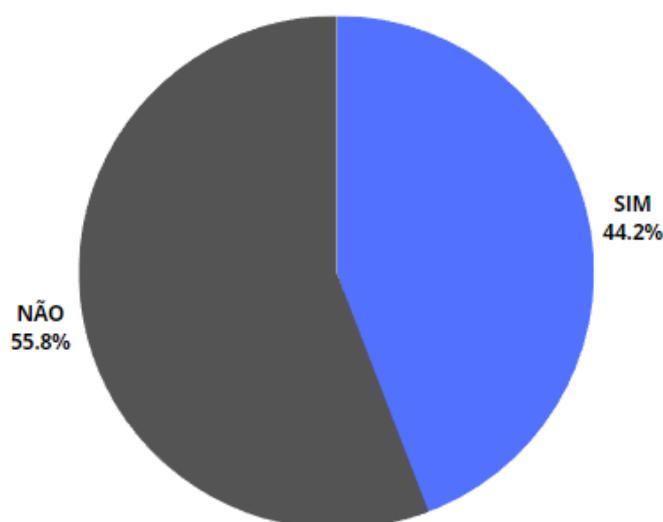
básico em busca de mapear o acesso à informação dos alunos que se candidataram ao Enem.

Além de perguntas relacionadas às questões socioeconômicas, também foram coletados dados relativos ao comportamento de estudos desses estudantes, uma vez que o país ainda estava sob o alerta de emergência para os riscos da pandemia do Covid-19¹⁵, analisar as dificuldades de preparo que estes estudantes tiveram ao longo do percurso até chegar os dias do exame, se fazem muito importantes e necessários, inclusive para mensurar os impactos que um isolamento social e a restrição para com a sala de aula, podem causar na qualidade do ensino em geral. Nesse sentido, algumas perguntas estratégicas foram feitas. O gráfico 1 demonstra a porcentagem de estudantes que afirmam ter tido dificuldades de infraestrutura para manter a rotina de estudos.

Gráfico 1: Dificuldade de infraestrutura

Dificuldade de infraestrutura para estudar

Total de respostas - 598.793



Fonte: Feito pela autora com base nos dados

¹⁵A Covid-19 é uma doença respiratória aguda, causada pelo coronavírus SARS-CoV-2. A pandemia que se iniciou em meados de 2019 inicialmente na China e se espalhou pelo mundo todo, foi um evento devastador que causou medo e insegurança em todo o mundo. Além de todas as perdas trágicas, fez com que o mundo inteiro tivesse que se adaptar a uma vida especialmente reclusa, com o isolamento social, e os impactos não somente sobre saúde vêm sendo observados e estudados até os dias atuais.

O gráfico 2, demonstra as principais dificuldades de infraestrutura que os alunos tiveram. Essa tabela pôde contar com 237.827¹⁶ respostas. É importante destacar que a pesquisa de hábitos de estudos igualmente à pesquisa socioeconômica, não era obrigatória e apesar do número de inscritos para a prova ter sido alta, destes apenas 928.564 estudantes responderam a pesquisa.

Gráfico 2: Tipo de dificuldade de infraestrutura

	54,27%
a . Equipamento (computador pessoal ou notebook) pouco disponível por ser compartilhado com outros familiares.	9,49%
b . Conexão ruim ou limitada com a internet.	26,02%
c . Equipamento (computador pessoal ou notebook) com configuração insuficiente para estudar ou acessar materiais.	2,81%
d . Material pedagógico insuficiente para estudar, faltando livros, apostilas, textos, vídeos.	5,11%
e . Local de estudos faltando boa iluminação, higiene, ventilação e silêncio.	1,58%
f . Local de estudos sem mobília para guardar e para dispor os materiais de estudo.	0,41%
g . Recursos insuficientes para alimentação.	0,30%

Fonte: Feito pela autora com base nos dados

A partir desse momento, serão apresentadas as colunas consideradas relevantes nesta pesquisa. As tabelas foram tratadas utilizando o programa *Excel*. Apesar da ferramenta *Power BI*, ser completa e totalmente capaz de tratar tabelas, para este trabalho a decisão de utilizar a ferramenta *Excel*, demonstra que é possível trabalhar os dados dentro da ferramenta em que ela é disponibilizada, não colocando como barreira a falta de outras ferramentas.

As colunas são as representações das informações, são separadas por códigos enquanto estão apresentadas como microdados, no tratamento desses dados, as colunas recebem nomes que facilitam a identificação. No caso do *Excel*, as colunas são representadas dentro do programa por letras do alfabeto, portanto os dados tratados possuem 15 colunas necessárias, de A a O.

A seguir serão apresentadas imagens parciais de cada coluna da planilha, de forma a identificar melhor suas transcrições. A figura 13, representa a coluna A que diz respeito à quantidade de inscritos. Nesse caso, cada linha representa 1 inscrito e por isso a numeração é por ordem crescente.

¹⁶ Todos os dados numéricos foram retirados dos microdados disponibilizados pelo site do INEP.

Figura 13: Coluna A de inscritos

A
Quantidade de inscritos
1
2
3
4
5
6
7
8
9
10

Fonte: da pesquisa

A figura 14, é a coluna B que representa a faixa etária de cada um.

Figura 14: Coluna B de faixa etária

B
Faixa Etária
Entre 41 e 45 anos
Entre 41 e 45 anos
20 anos
19 anos
20 anos
19 anos
21 anos
23 anos
19 anos
19 anos
17 anos
19 anos
21 anos
22 anos
23 anos
20 anos

Fonte: da pesquisa

A figura 15 representa o gênero marcado pelos estudantes, representados na coluna C.

Figura 15: Coluna C de gênero

C
Gênero
M
M
F
M
M
F
M
F
F
F
F
F
M

Fonte: da pesquisa

A figura 16, na sequência, é a coluna D que diz respeito ao estado civil.

Figura 16: Coluna D de estado civil

D
Estado Civil
Casado(a)/Mora com companheiro(a)
Casado(a)/Mora com companheiro(a)
Solteiro (a)
Solteiro (a)
Não Informado
Solteiro (a)

Fonte: da pesquisa

Figura 19, informa sobre a situação escolar do estudante, se concluído ou não o ensino médio.

Figura 19: Coluna G de situação escolar

G
Situação Do EM
Já concluí o ensino médio
Estou cursando e concluirei o Ensino Médio em 2022
Já concluí o ensino médio
Já concluí o ensino médio
Já concluí o ensino médio
Estou cursando e concluirei o Ensino Médio em 2022

Fonte: da pesquisa

A coluna H, representada na figura 20, ainda sobre o ensino médio, informa sobre o ano de conclusão.

Figura 20: Coluna H de ano de conclusão

H
Ano de Conclusão EM
2020
Antes de 2007
2020
2020
2021
2020
2020
Não Informado
Não Informado
2021
2020
Não Informado
2021
Não Informado
2019

Fonte: da pesquisa

A coluna I, representada na figura 21, informa sobre o tipo de escola em que cursa ou cursou o ensino médio.

Figura 21: Coluna I de tipo de escola

I
Tipo de Escola
Não respondeu
Privada
Não respondeu
Não respondeu
Não respondeu
Pública

Fonte: da pesquisa

A figura 22, apresenta a coluna J que informa sobre o município de residência do estudante.

Figura 22: Coluna J de município de residência

J
Município
Brasília
Brasília
Presidente Tancredo Neves
Cariacica
Salinópolis
Rio de Janeiro
Recife
São Gonçalo
Amargosa
Arcoverde
Patos
Nossa Senhora da Glória
Salvador

Fonte: da pesquisa

Em complemento, a coluna K representada na figura 23, informa sobre a Unidade Federativa de residência do estudante.

Figura 23: Coluna K de Unidade Federativa

K
Unidade Federativa
DF
DF
BA
ES
PA
RJ
PE
RJ
BA
PE
PB
SE
BA
SP
MT

Fonte: da pesquisa

As figuras 24 e 25, são responsáveis por informar a presença do estudante nos dias de prova 1 e 2, respectivamente.

Figura 24: Coluna L presentes dia 1

L
Presença Dia 1
Faltou à prova
Faltou à prova
Presente na prova
Presente na prova
Faltou à prova
Faltou à prova
Faltou à prova
Presente na prova
Faltou à prova
Presente na prova
Faltou à prova

Fonte: da pesquisa

Figura 25: Coluna M presentes dia 2

M
Presença Dia 2
Fatou à prova
Fatou à prova
Presente na prova
Presente na prova
Fatou à prova
Fatou à prova
Fatou à prova
Presente na prova
Fatou à prova
Presente na prova
Fatou à prova

Fonte: da pesquisa

A figura 26, representada pela coluna N, informa sobre os alunos que estiveram presentes no dia da prova subjetiva (redação) e suas notas.

Figura 26: Coluna N alunos presentes na prova subjetiva

N
Redação
Eliminado
Eliminado
760
320
Eliminado
Eliminado
Eliminado
440
Eliminado
360
Eliminado

Fonte: da pesquisa

Já a coluna O, representada na figura 27, mostra os alunos que tiveram sua participação “eliminada”, por não comparecer a nenhum dos dois dias de prova.

Figura 27: Coluna O alunos eliminados

O
ELIMINADOS
Eliminado
Eliminado
Eliminado
Eliminado
Eliminado
Eliminado
Eliminado

Fonte: da pesquisa

Vale lembrar que as imagens demonstram apenas imagens parciais dos dados que foram tratados. Outra informação relevante para compreensão desta análise, é que nas colunas que dizem respeito aos aspectos sociais dos estudantes, são em sua maioria de respostas não obrigatórias, portanto algumas linhas aparecem em branco e isso se dá pela decisão do estudante de não responder. Isso significa que, como qualquer análise estatística, há uma margem de erro nos resultados, porém, para esta pesquisa a renúncia não altera os resultados esperados.

Avaliações socioeconômicas são fundamentais para a aplicação e criação de políticas públicas para a sociedade, a fim de diminuir a desigualdade social entre os indivíduos. Fávero (2014) utiliza a terminologia Estudo Social, como uma forma de explicar que dentro da área de Serviço Social, esse estudo leva em consideração o contexto social e econômico em que o indivíduo está inserido, de forma geral, visa construir um perfil de uma pessoa, de forma crítica, as suas particularidades. Mioto (2009) ressalta que esse estudo é uma forma de reconhecer a realidade social do usuário, conhecendo a sua demanda de forma pessoal, mesmo que seja para a implementação de uma política pública, que vai beneficiar toda uma comunidade.

[...] Nessa sequência, o estudo socioeconômico deve/deveria/poderia iniciar a partir das necessidades individuais do/a estudante, relacionando-as com seu contexto familiar, territorial, cultural, de raça, de gênero, de orientação sexual, dentre outras, e com esse embasamento, que o discente fosse incluído nos auxílios e nos serviços ofertados. Hoje, a lógica é inversa, e os limites de atendimento estão dados, uma vez que o que se faz são comparações e classificações das situações socioeconômicas entre as famílias de estudantes requerentes (MIRANDA, 2021, p.120).

Vale reforçar que traçar um perfil socioeconômico é, historicamente relevante no momento em que se decide avaliar aspectos do ensino, porém do ponto de vista adotado para esta pesquisa não se fez necessário aprofundar nesses aspectos, ainda que possa refletir no desempenho. Porém, com base em uma breve análise de respostas, considerando escolaridade dos responsáveis, faixa de renda e acesso a computadores e internet com facilidade percebeu-se o perfil típico desses estudantes.

O cruzamento de dados desta pesquisa, levou em consideração tanto os aspectos socioeconômicos, quanto às localizações onde foram aplicadas as provas do Enem, com o intuito de diagnosticar quais localizações tiveram o maior e o

menor número de participantes presentes na realização do exame. Foi realizado mapeamento e criado um gráfico do mapa por regiões geográficas com a quantidade total de inscrito por cada uma, e pode ser verificado no gráfico x, abaixo.

Gráfico 3: Mapa de inscritos por região



Fonte: Feito pela autora com base nos dados

Também foi dividido quantidade de inscritos por Unidade Federativa - UF, ainda separados por regiões geográficas. As UF's que compõem a região norte estão dispostas na tabela 2.

Tabela 1: Inscritos no Norte

Unidade da Federação	399.922
Rondônia	32.107
Acre	22.248
Amazonas	80.976
Roraima	8.430
Pará	203.914
Amapá	22.388
Tocantins	29.859

Fonte: Feito pela autora com base nos dados

A região nordeste conta com uma quantidade significativa de inscritos no total, a tabela 2 demonstra a quantidade dividida por seus Estados.

Tabela 2: Inscritos no Nordeste

Unidade da Federação	1.212.182
Maranhão	133.981
Piauí	84.068
Ceará	227.145
Rio Grande do Norte	87.787
Paraíba	104.139
Pernambuco	186.454
Alagoas	65.225
Sergipe	56.541
Bahia	266.842

Fonte: Feito pela autora com base nos dados

Observa-se que mesmo não sendo a região com maiores Estados, é a região com maior quantidade de inscritos.

Já a região Sudeste, que conta com a segunda posição de mais inscritos com 1.169.311 sendo um número muito próximo da região nordeste, a tabela 3 demonstra que a concentração de inscritos está no estado de São Paulo. De acordo com o Censo de 2022¹⁷, a região sudeste é a mais populosa do Brasil.

Tabela 3: Inscritos no Sudeste

Unidade da Federação	1.169.311
Minas Gerais	312.772
Espírito Santo	63.855
Rio de Janeiro	250.272
São Paulo	542.412

Fonte: Feito pela autora com base nos dados

Em sequência, a tabela 4 apresenta os dados dos inscritos da região Centro-Oeste. Apesar de não ser o maior Estado da região, o Goiás concentra o maior número de inscritos no exame.

¹⁷ Podem ser verificados em: <https://censo2022.ibge.gov.br/panorama/>

Tabela 4: Inscritos no Centro-Oeste

Unidade da Federação	310.462
Mato Grosso do Sul	44.527
Mato Grosso	59.414
Goiás	142.742
Distrito Federal	63.779

Fonte: Feito pela autora com base nos dados

A região Sul do Brasil, é a menor em termos de tamanho geográfico, porém conta com uma boa quantidade de inscritos, além disso não está em último lugar no quesito quantidade de inscritos totais. A tabela 5 mostra que está ainda à frente do Centro-Oeste nessa questão.

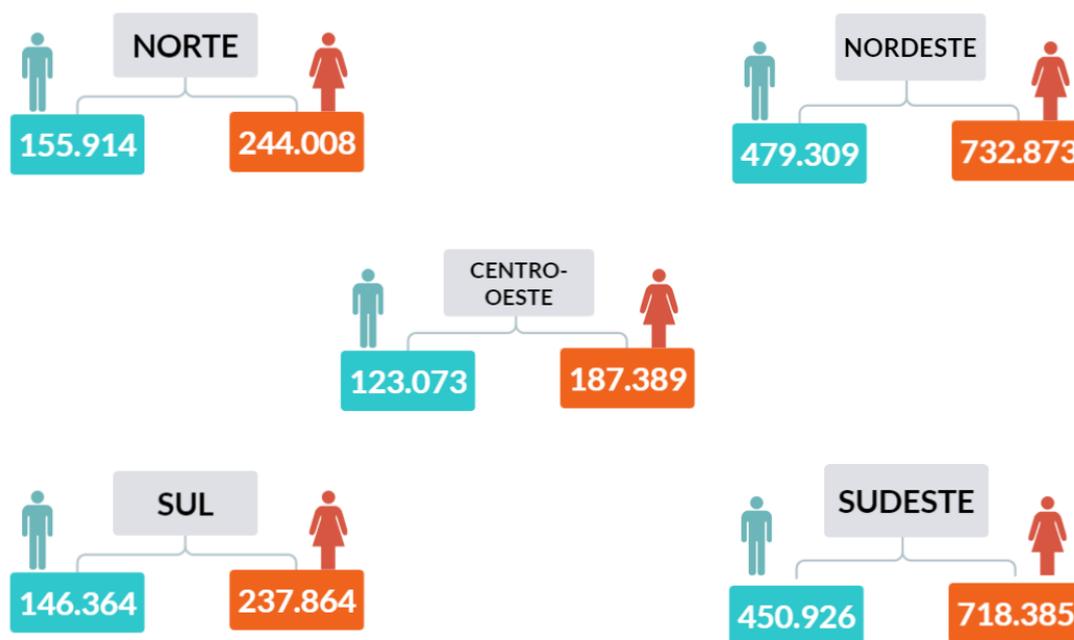
Tabela 5: Inscritos no Sul

Unidade da Federação	384.228
Paraná	149.840
Santa Catarina	81.942
Rio Grande do Sul	152.446

Fonte: Feito pela autora com base nos dados

Para uma questão de comparação, foi separado a quantidade de inscritos de acordo com gênero e região. O gráfico 2 ilustra esse comparativo.

Gráfico 4: Mapa de homens e mulheres



Fonte: Feito pela autora com base nos dados

Com base na análise da quantidade de inscritos, foi possível constatar que a maioria de inscritos em todas as regiões foram por pessoas que se declararam do gênero feminino. As mulheres representam mais de 60% das inscrições para o exame e essa realidade não é incomum. A primeira constatação desse fato se deu em 2004 pelo próprio Ministério da Educação¹⁸, e esse comportamento vem se repetindo desde então.

Outro aspecto de grande importância, são as declarações de racialidade dos estudantes. Desde o início dos anos 2000 vem ocorrendo no Brasil um fenômeno de mudança de identificação racial, como observado por Soares (2008).

A história é bastante clara e simples. Até o início dos anos 1990, a população negra vinha aumentando de modo relativamente lento e vegetativo devido a uma taxa de fecundidade um pouco mais alta [e também à miscigenação e outros fatores que envolvem a identificação racial ao nascer]. [Na segunda metade dos anos 1990, no entanto, iniciou-se] um processo de mudança em como as pessoas se veem. As pessoas passaram a ter menos vergonha de dizer que são negras; passaram a não precisar se branquear para se legitimarem socialmente (SOARES, 2008, p. 116).

¹⁸ Pode ser verificado em:

<http://portal.mec.gov.br/ultimas-noticias/201-266094987/2592-sp-1049840106>

Além disso, o autor conclui que “poucas vezes nas Ciências Sociais há um fenômeno com uma explicação tão clara quanto o escurecimento da população brasileira” Soares (2008, p.116). O autor é certo ao afirmar também que esse processo estava longe de terminar, e a partir da análise de apenas um aspecto educacional, foi possível notar que mesmo que em algumas regiões as pessoas que se declaram branca sejam uma boa maioria, a diferença entre as que se declaram negras não estão muito longe uma da outra. Verifica-se na tabela 6 abaixo.

Tabela 6: Tabela de declaração de raça por região

	total	Não Declarada	Branca	Preta	Parda	Amarela	Indígena
Brasil	3.476.105	64.867	1.428.081	408.422	1.488.844	66.051	19.840
Norte	399.922	7.982	84.013	41.762	252.627	8.205	5.333
Nordeste	1.212.182	24.869	328.486	169.674	656.531	24.481	8.141
Sudeste	1.169.311	18.073	612.792	141.823	372.456	20.939	3.228
Centro-Oeste	310.462	7.288	119.901	34.318	138.554	7.949	2.452

Fonte: Feito pela autora com base nos dados

É possível perceber ainda que, o número de pessoas que não se declaram dentro de nenhum desses aspectos, não é ideal. Chama atenção para verificar se o motivo para essa escolha seja apenas de não conhecimento de suas origens. A tabela mostra ainda uma visão geral das declarações dos inscritos do Brasil todo, e nota-se que a maioria deles se autodeclaram Pardos e que exceto na região Sul do Brasil, a maioria por região também se autodeclaram da mesma forma.

Ao analisar esse comportamento, esse tipo de pesquisa abre um espaço para que outras entidades possam iniciar estudos nesse campo, visto que a falta de declaração também pode significar uma falta de informação, pois isso reflete na forma como a pessoa se vê.

Um último aspecto que chama atenção nas análises, são as solicitações de pessoas que fazem uso de nome social devido a sua identidade de gênero. No que tange a este aspecto, informações como esta podem ser importantes para a criação e implementação de políticas públicas para esta minoria, assim como as minorias raciais que com grande luta vem conquistando seus espaços, que ainda são escassos e merecem atenção. A identificação de gênero segue a mesma linha da declaração racial, já que parte do ponto em que o indivíduo vê a si mesmo e se

reconhece como tal. Nesse sentido e levando em conta que a perspectiva política e social conversam entre si, Silva e Barboza (2009) diz que tão somente a imposição de se autodeclarar como ir em busca de uma “militância” “[...] lhes possibilita participação ativa em processos de formulação de políticas públicas a cada vez que pensam ações possíveis de governo [...]”, ainda que os números sejam baixos como mostra a tabela 7, qualquer luta começa aos poucos.

Tabela 7: Tabela de declaração de uso de nome social

Região Geográfica	Nome Social
Brasil	363
Norte	18
Nordeste	103
Sudeste	169
Sul	48
Centro-Oeste	25

Fonte: Feito pela autora com base nos dados

O Enem é um processo muito rico de informações e com inúmeras possibilidades de análises e interpretações. A CI entende que qualquer dado disponibilizado pelo governo é de interesse de todos os cidadãos. Mesmo com muitas formas de acesso e que hoje boa parte da população tenha acesso a pelo menos um computador e internet, uma parcela pequena é capaz de interpretar informações disponibilizadas em formatos não pesquisáveis. Igualmente uma boa parte da população não tem conhecimento científico necessário para utilizar dados como tomada de decisão. Esta pesquisa visou demonstrar que existem formas diferentes de interpretar os dados que estão disponibilizados sem o devido tratamento, além disso foi possível demonstrar que a partir do trabalho de um bibliotecário, outros profissionais podem se beneficiar de informações tratadas e desenvolver um trabalho multidisciplinar, assim como foi demonstrado que é a Ciência da Informação.

7 CONSIDERAÇÕES FINAIS

Do ponto de vista de Moreira, Beira e Oliveira (2020) considera dado, informação e conhecimento como parte necessária para a comunicação, identificando o dado como uma espécie de matéria prima da informação. A partir disso, o dado pode ser compreendido como uma informação bruta que precisa ser tratada (analisada) para ser considerada apta a ser transmitida. A grosso modo, informação é obtida através da habilidade de analisar, e a partir disso é gerado o conhecimento.

Neste trabalho verificou-se que o profissional da ciência da informação, em particular o profissional formado em biblioteconomia, tem total capacidade de lidar com o tratamento dos dados brutos e criando gráficos e tabelas a partir da ferramenta Power BI, gerando informações de maneira simples e intuitiva.

A partir das análises, o profissional pode tomar a iniciativa de criar infográficos que podem trazer uma nova perspectiva da informação trabalhada, por meio de linguagem mais simples e mais democráticas, trazendo imagens e uma análise menos técnica sobre os dados. Além disso, conforme Amaral (2016), criar uma visualização dos dados é capaz de tornar a informação mais clara de uma forma mais resumida.

Relembrando a interdisciplinaridade da área, conclui-se que através de estudos entre os autores, aliado ao conhecimento que o profissional da CI adquire ao longo de sua formação, demonstrou que um pesquisador se constrói de maneira multidisciplinar e portanto capaz de buscar, analisar e defender o conteúdo buscado.

Os dados abertos, no Brasil, ainda são incipientes, ainda que atualmente existam inúmeras políticas que envolvam essa temática, e que o estudo da área tenha aumentado consideravelmente, faz se necessário um incentivo por parte do Governo em parceria com as universidades públicas, já que é o espaço em que ocorre o desenvolvimento científico. Porém, o Governo Federal já possui uma iniciativa de criação e manutenção de um portal com o objetivo de concentrar os dados abertos para o acesso público, como uma parte de um conceito de governo aberto e transparente. A iniciativa já possui documentação que detalha os objetivos e os processos para alcançá-los. O portal se chama “dados.gov.br” e já pode ser consultado, porém ainda não está completo em todas as suas funcionalidades.

Principalmente por haver uma movimentação em busca da concretização desse projeto, percebe-se a partir dessa análise que o portal tem uma alimentação independente por partes dos órgãos, algo que dificulta uma atualização precisa dos resultados, sugerindo que é necessário profissionais alocados para essa função, por esse motivo, a análise de dados para além de análises estatísticas é um tema que deve continuar sendo pesquisado e desenvolvido deixando evidente a necessidade de ser desenvolvido como um trabalho multidisciplinar. Vale mencionar que houve uma tentativa de contato para obter mais informações sobre a formação profissional da equipe responsável pelo desenvolvimento a fim de constatar a falta do profissional em questão, entretanto devido ao órgãos serem setorizados, não houve uma confirmação precisa dessa informação.

É importante destacar que apesar do modelo gerencial horizontal adotado pelos desenvolvedores, a demonstração dos dados apenas por meio de gráficos, do ponto de vista do profissional da CI, ainda é uma linguagem pouco democrática para a maioria da população, o que demonstra a necessidade de uma participação ativa deste profissional em projetos como este.

Em linhas gerais, foi possível perceber que os bibliotecários estão (e deve ser assim) ocupando espaços informacionais para além do escopo das bibliotecas, visto que a partir do momento em que não se vê mais as bibliotecas como espaço de depósito de conhecimento, mas sim de espaço de democratização, se faz necessário que este profissional esteja trabalhando arduamente dentro de todos os espaços profissionais que lidam com a informação e com a disseminação desta. O estudo dos “Bibliotecários de Dados”, é uma pesquisa de muita relevância para a CI, já que os Dados são o futuro, e a CI continuará formando bibliotecários que irão cada vez mais buscar uma maneira de entender e recuperar informações. Se considerar que dado é o novo meio de informar, estima-se que sempre haverá um bibliotecário por trás.

Atualmente, o acesso à informação e ao conhecimento foram facilitados pela tecnologia, tirando o conhecimento do escopo elitista, nada mais justo que tenha um bibliotecário que é o profissional com a missão de resguardar, recuperar e disseminar, como agente democratizador da informação. Constatado a importância

dos dados, é cabível dizer que se para cada livro o seu leitor, então para cada Dado o seu bibliotecário¹⁹.

Por fim, esta pesquisa nunca teve a intenção de esgotar todas as fontes e trazer resultados irrefutáveis, a pesquisa sobre dados no Brasil ainda se projeta num futuro distante visto que a internet e as tecnologias não deixaram de avançar em nenhum momento. A relevância do profissional formado em biblioteconomia se provou por meio de novos autores discutindo o tema e do crescente desempenho desse profissional em áreas de tecnologias, não apenas pela escassez de local de atuação, em especial no serviço público devido a precarização das bibliotecas, mas especialmente por esse profissional ter se mostrado cada vez mais capaz de lidar com excelência no que tange a conexão entre informação e web.

¹⁹ Esta é uma interpretação desta autora, de forma bem subjetiva, baseada nas 5 leis de Ranganathan.

REFERÊNCIAS

AUSTIN, Brice. **Mooer's Law: In and Out of Context**. Journal of the American Society for Information Science, v 52(8):607-609. 2001.

AMARAL, Fernando. **Introdução à ciência de dados: mineração de dados e Big Data**. Rio de Janeiro: Alta Books, 2016.

Barros, R., Foguel, M., & Ulyseia, G. (2006). **Desigualdade de Renda no Brasil: Uma Análise da Queda Recente** IPEA, Rio de Janeiro.

BERNERS-LEE, T. et al. **The Semantic Web**. 2001. Scientific American. Disponível em: <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21> Acesso em: 2 jun 2023.

BRÄSCHER, M.; CAFÉ, L. Organização da Informação ou Organização do Conhecimento? In: **ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO**, 9, 2008, São Paulo, Anais. São Paulo: ANCIB, 2008. Disponível em: <http://www.enancib2008.com.br> Acesso em: 2 jun. 2023.

BRIETMAN, K. **Big Data Overview**. Palestra apresentada no 1o. EMC Summer School on Big Data. EMC/NCE/UFRJ. Rio de Janeiro. 2013.

BORKO, H. Information Science: what is it? **American Documentation**, p. 3-5, Jan. 1968.

BUCKLAND, M. K. **Information as thing**. Journal of the American Society for Information Science, v. 45, n. 5, p. 351-360, 1991.

BURRELL, Gibson; MORGAN, Gareth. **Sociological paradigms and organizational analysis: Elements of the sociology of corporate life**. Routledge, 2017.

CAPURRO, R. Epistemologia e Ciência da Informação. In: **V Encontro Nacional de Pesquisa em Ciência da Informação**, Belo Horizonte, 2003.

CARDOSO, Ana Maria Pereira. **Pós-Modernidade e informação: conceitos complementares? Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 63-79, jan./jul. 1996.

CESARINO, M. A. **Sistema de Recuperação da informação**. Revista da Escola de Biblioteconomia da UFMG, Belo Horizonte, v. 14, n. 2, p. 157-168, set. 1985. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/reb/>>. Acesso em: 28 out. 2023.

CETIC.BR - Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação. **Bases de microdados**. [s.l.]. [2023]. Disponível em: <https://www.cetic.br/microdados/> Acesso em: 29 nov. 2023.

CETAX. **A diferença entre ciência de dados e análise de dados**. [s.l.]. [2022]. Disponível em: <https://www.cetax.com.br/blog/ciencia-de-dados-e-analise-de-dados/> Acesso em: 20 jun. 2023.

CONTROLADORIA-GERAL DA UNIÃO (CGU). **Acesso à informação pública: uma introdução à Lei nº 12.527, de 18 de novembro de 2011**. Brasília: CGU, 2011. Disponível em: <https://www.gov.br/acessoainformacao/pt-br/central-de-conteudo/publicacoes/arquivos/cartilhaacessoainformacao-1.pdf> Acesso em: 2 jun. 2023.

CUNHA, I. M. R. F. Análise documentária. In: SMIT, J. W. (Coord.). **Análise documentária: a análise da síntese**. Brasília: IBICT, 1987. p. 38-60.

DAVENPORT, T. H. **Big data no trabalho: derrubando mitos e descobrindo oportunidades**. Rio de Janeiro: Elsevier, 2014. _____.; PATIL, D. J. Data scientists: the sexiest job of the 21st century. **Harvard Business Review**, [S.l.], v. 90, n. 10, p. 70-76, out. 2012.

FÁVERO, E. T. **O Estudo Social: fundamentos e particularidades de sua construção na Área Judiciária**. In: CFESS (org.). **O Estudo Social em Perícias, Laudos e Pareceres Técnicos: debates atuais no Judiciário, no Penitenciário e na Previdência Social**. 11. ed. São Paulo: Cortez, 2014.

FERNÁNDEZ-MOLINA, J. C. **Enfoques objetivo y subjetivo Del concepto de información**. Revista Española de Documentación Científica, [S. l.],v.17, n.3, p. 320-330, 1994.

FRIAS, A. A escrita científica e a divulgação do conhecimento científico. **Cogitare Enfermagem**, v. 20, n. 2, 28 jun. 2015. DOI 10.5380/ce.v20i2.41922. Disponível em: <http://revistas.ufpr.br/cogitare/article/view/41922>.

GARCÍA MARCO, E. J. **Los contenidos y a secuencia docente de la Organización y Representación del Conocimiento: una propuesta interdisciplinar**. **Organización del conocimiento en sistemas de información y documentación**. Zaragoza: Librería General, v. 1, p. 219-228, 1995.

GARFIELD. Eugene. A Tribute To Calvin N. Mooers, **A Pioneer Of Information Retrieval**. The Scientist, Vol: 11(4) March 17, 1997.

GOMES, H. E. **Tendências da pesquisa em organização do conhecimento. Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 2, n. 1. p.60-88. 2009. Disponível em: <https://brapci.inf.br/index.php/res/v/119303> Acesso em: 22 jun. 2023.

GUIMARÃES, J. A. C. A Análise Documentária no Âmbito do Tratamento da Informação: Elementos Históricos e Conceituais. In : MEDLEG, G.; LOPES, iL. (org.) **Organização e representação do conhecimento na perspectiva da Ciência da Informação**. Brasília: Thesaurus, 2003. (Estudos avançados em Ciência da Informação, 2).

KULHTHAU, Carol Collier.; TAMA, S. L. **Information search process of lawyers: a call for just for me 'information services'**. Journal of Documentation, New Jersey, v. 57, n. 1,p.25-43, Jan., 2001. Disponível em: <http://comminfo.rutgers.edu/~tefko/Courses/612/Articles/KuhlthauTama.pdf> Acesso em: 13 out. 2011.

LAKATOS, Eva Maria; MARCONI, Marina. **Sociologia geral**. São Paulo : Atlas, 1999.

LANCASTER, F. W. Indexação e resumo: teoria e prática. 2. ed. Brasília: Briquet de Lemos Livros, 2004. 452 p.

MCGEE, J.; PRUSAK, L. **Gerenciamento Estratégico da Informação**. 11. ed. Rio de Janeiro: Campus, 1994.

MUELLER, S. P. M.; PASSOS, E. J. L. **As questões da comunicação científica e a ciência da informação**. In: MUELLER, S. P. M.; PASSOS, E. J. L. (Orgs.). **Comunicação científica**. Brasília: Ciência da Informação, 2000. p. 13–22.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos: 1999.

MIOTO, R. C. T. **Estudos socioeconômicos. In: CONSELHO FEDERAL DE SERVIÇO SOCIAL; ASSOCIAÇÃO BRASILEIRA DE ENSINO E PESQUISA EM SERVIÇO SOCIAL. Serviço Social: direitos sociais e competências profissionais**. Brasília/DF: CFESS/ABEPSS, 2009.

MIRANDA, A. M. **O Estudo Socioeconômico nas Políticas de Assistência Estudantil dos Institutos Federais da Região Centro-Oeste: configurações, contradições e perspectivas**. 2021. Dissertação (Mestrado em Política Social) – Universidade de Brasília, Brasília/DF, 2021. Disponível em: <https://repositorio.unb.br/handle/10482/41764> Acesso em: 09 out. 2023. » <https://repositorio.unb.br/handle/10482/41764>

MORIN, E. **O método III: o conhecimento do conhecimento**. Lisboa : Publicação Europa-América, 1987.

_____. **O método II: a vida da vida**. Porto Alegre : Sulina, 2001.

MOOERS, Calvin N. **Zatocoding applied to mechanical organization of knowledge**. *American Documentation*. 2(1), p.20-32, 1951.

MOOERS, Calvin N. **Mooers' law or, Why Some Retrieval Systems Are Used and Others Are Not**. *American Documentation*. 11 (3): ii, July 1960.

OPEN KNOWLEDGE FOUNDATION. The Open Definition. Disponível em: <https://opendefinition.org/> Acesso em: 2 jun. 2023.

OPEN GOVERNMENT PARTNERSHIP - OGP. Senado federal. Brasília. Disponível em: <https://www25.senado.leg.br/web/atividade/sinfleg/relatorio-de-gestao/saiba-mais-23> Acesso em: 2 jun. 2023.

O que é iniciativa. Senado federal, Brasília, 11 dez. 2014. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/o-que-e-a-iniciativa> Acesso em: 2 jun. 2023.

PINHEIRO, L. V. R.; LOUREIRO, J. M. M. **Traçados e limites da ciência da informação**. *Ciência da Informação*, v. 24, n. 1, 1995.

PIRES, Marco Túlio. **Guia de dados abertos**. *In: Guia de Dados Abertos*. São Paulo, 2015. Disponível em: https://nic.br/media/docs/publicacoes/13/Guia_Dados_Abertos.pdf Acesso em: 2 jun. 2023. *informação. Ciência da Informação*, v. 24, n. 1, 1995.

PORTO, F., ZIVIANI, A. **Ciência de dados**. III Seminário de Grandes Desafios da Computação no Brasil, Rio de Janeiro, 2014.

ROLIM, M. V. **Análise do perfil do profissional da informação para a atuação como cientista de dados em ambientes de big data: uma perspectiva a partir das disciplinas do curso de Biblioteconomia da UnB**. 2018. 71 f., il. Trabalho de Conclusão de Curso (Bacharelado em Biblioteconomia)—Universidade de Brasília, Brasília, 2018. Disponível em: <https://bdm.unb.br/handle/10483/20898> Acesso em: 2 jun. 2023.

ROBREDO, J. **Filosofia da ciência da informação ou ciência da informação e filosofia? - Uma questão que merece ser pensada**. Brasília, 2007. Disponível em www.cinform.ufba.br/7cinform/soac/papers/369a2be3343ea1ed160564371174.pdf Acesso em: 10 set 2023.

SANTOS, Boaventura de Sousa. **Um discurso sobre as ciências**. 8. ed. Porto : Afrontamento, 1996.

SARACEVIC, Tefko. **Information science**. *Journal of the American Society for Information Science*, Washington, v.50, n. 12, p. 1051-1063, 1999.

SARACEVIC, Tefko. **Ciência da informação: origem, evolução e relações. Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

SARACEVIC, Tefko. **Introduction to information science**. Nova Iorque: Bowker, 1970.

Semeler A.; Pinto, A. **Data librarianship Venn diagram handbook (beta version)**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2019b.

SETZER, V. W. **Dado, Informação, Conhecimento e Competência**. Revista de Ciência da Informação, São Paulo, dez. 1999.

SILVA, Alessandro S. & BARBOZA, Renato. **Exclusão social e consciência política: luta e militância de transgêneros no ENTLAIDS**. Cadernos CERU, v. 20, n. 1, 2009, pp. 257-276.

SOARES, S. S. D. **A demografia da cor: a composição da população brasileira de 1870 a 2007**. In: THEODORO, M. (org.). **As políticas públicas e a desigualdade racial no Brasil: 120 anos após a abolição**. Brasília: Ipea, 2008. p. 97-117.

TAYLOR, A. G. **The Organization of Information**. London: Libraries Unlimited. 2004. 417p.

TUKEY, J. W. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley, 1977.

URDANETA, I. P. **Gestión de la inteligencia, aprendizaje tecnológico y modernización del trabajo informacional: retos y oportunidades**. Caracas: Universidad Simón Bolívar, 1992.

ZIVIANI, A.; PORTO, F.; OGASAWARA, E. **Ciência de dados: desafio para a ciência, indústria e governo**. ComCiência: Revista Eletrônica de Jornalismo Científico, Campinas, v. 170, jul. 2015. Não paginado. Disponível em: <http://comciencia.br/comciencia/handler.php?section=8&edicao=115&id=1381>
Acesso em: 26 fev. 2023.

_____. **Les règles de la méthode sociologique**. Paris: Alcan, 1895.

WERSIG, Gernot. **Information science: the study of postmodern knowledge usage**. *Information Processing & Management*, Elmsford, v. 29, n. 02, p. 229-239, mar. 1993.