

UNIVERSIDADE FEDERAL DE SÃO CARLOS (UFSCar)
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS (CECH)

MARCOS TERUO OUCHI

UM ESTUDO DA *BLOCKCHAIN* APLICADO AO CONTEXTO DOS DADOS DE
PESQUISA

*BLOCKCHAIN STUDY APPLIED TO THE CONTEXT OF SCIENTIFIC RESEARCH
DATA*

São Carlos/SP
2019

MARCOS TERUO OUCHI

UM ESTUDO DA *BLOCKCHAIN* APLICADO AO CONTEXTO DOS DADOS DE
PESQUISA CIENTÍFICA

Dissertação apresentada à banca de defesa de Mestrado do Programa de Pós-graduação em Ciência da Informação da Universidade Federal de São Carlos, como requisito para obtenção do título de mestre em Ciência da Informação.

Área de Concentração: Conhecimento, Tecnologia e Inovação

Linha de Pesquisa: Tecnologia, Informação e Representação

Orientador(a): Profa. Dra. Ana Carolina Simionato Arakaki



grupo de pesquisa
dados e metadados

São Carlos/SP
2019

MARCOS TERUO OUCHI

UM ESTUDO DA *BLOCKCHAIN* APLICADO AO CONTEXTO DOS DADOS
DE PESQUISA CIENTÍFICA

BLOCKCHAIN STUDY APPLIED TO THE CONTEXT OF SCIENTIFIC
RESEARCH DATA

Dissertação apresentada à banca de defesa de Mestrado do Programa de Pós-graduação em Ciência da Informação da Universidade Federal de São Carlos, como requisito para obtenção do título de mestre em Ciência da Informação.

Área de Concentração: Conhecimento, Tecnologia e Inovação

Linha de Pesquisa: Tecnologia, Informação e Representação

BANCA EXAMINADORA

Prof^a Dr^a Ana Carolina Simionato Arakaki
Universidade Federal de São Carlos (UFSCar)

Prof. Dr. Guilherme Ataíde Dias Universidade
Federal da Paraíba (UFPB)

Prof. Dr. Rogério Aparecido Sá Ramalho
Universidade Federal de São Carlos (UFSCar)

Universidade Federal de São Carlos - UFSCar

São Carlos / SP, 13 de agosto de 2019.



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência da Informação

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Marcos Teruo Ouchi, realizada em 13/08/2019:

Prof.ª. Dra. Ana Carolina Simionato Arakaki
UFSCar

Prof. Dr. Guilherme Ataíde Dias
UFPB

Prof. Dr. Rogério Aparecido Sá Ramalho
UFSCar

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Guilherme Ataíde Dias e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ão) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.

Prof.ª. Dra. Ana Carolina Simionato Arakaki

*Nesses tempos obscuros para a Educação e a
Ciência Brasileira, àqueles em cujos ombros
me apoiei.*

*À Selma, amada esposa e Valentine, amada
filha. Sem vocês nada disso teria sentido.*

AGRADECIMENTOS

Essa dissertação é resultado de mais uma daquelas decisões tomadas em encruzilhadas repentinas da vida, quando você conhece pessoalmente alguém cujo trabalho há muito tempo admira e a sua intuição diz que é hora de abraçar mais um desafio. E, assim, você tem a oportunidade e a honra de ter como orientadora a Profa. Dra. Ana Carolina Simionato Arakaki a quem devo meus mais sinceros agradecimentos.

Impossível também de deixar de agradecer ao Prof. Dr. Rogério Aparecido Sá Ramalho por tantas contribuições desde minha graduação até o presente trabalho e pela compreensão pelas escolhas que fiz para o mestrado.

Agradeço também à Profa. Dra. Ariadne Chloe Mary Furnival, uma grande referência profissional e pessoal, pelas inumeráveis contribuições desde os tempos da minha graduação.

Pela leitura e valiosas contribuições a este trabalho, agradeço ao membro da banca de avaliação Prof. Dr. Guilherme Dias.

Aos mestres, por compartilharem seus conhecimentos.

À toda a equipe do Departamento de Ciência da Informação pela paciência e por facilitar nosso percurso.

Ao “Bond do Capurro”, grandes companheiros do Programa de Pós-Graduação em Ciência da Informação, pelas inúmeras oportunidades de trabalharmos e aprendermos juntos.

Às minhas companheiras e companheiros do Sistema de Bibliotecas do Município de São Carlos, em especial às minhas companheiras Karina Manfré Fröner, Kelly Marcomini e nosso diretor César Maragno por acreditarem nas ideias malucas que levo para o sistema.

Por fim e não menos importante agradeço às minhas amadas Selma e Valentine por serem meus motivos para buscar constantemente a minha melhor versão.

Aqui tem informação!
(Rogerinho do Ingá)

RESUMO

A possibilidade da discussão científica é o cerne da rotina prática de pesquisa e, para isso, os pesquisadores devem encontrar recursos e descrições suficientemente detalhadas nos trabalhos publicados a fim de obterem os mesmos resultados dentro das margens suficientes de erros experimentais. Dentre os recursos disponíveis, os dados coletados durante as pesquisas são essenciais, pois o resultado obtido deve ser o produto de sua rigorosa análise. Nesse escopo e associada a mecanismos de consenso distribuído, a estrutura de dados *Blockchain* pode garantir a confiabilidade dos dados de forma transparente, descentralizada e imutável, e surge como possibilidade para contribuir para o progresso da Ciência por meio de novas maneiras de tratamento das grandes quantidades de dados coletados em uma pesquisa científica. Neste sentido, o objetivo geral desse trabalho consiste em analisar a tecnologia *Blockchain* e suas possibilidades inseridas à Comunicação Científica e na Gestão de Dados de Pesquisa. Os objetivos específicos são descrever a gestão de dados de pesquisa e o ciclo de vida de dados; identificar os principais conceitos, a estrutura e o funcionamento, relacionados à tecnologia *Blockchain*; e apontar a tecnologia *Blockchain* e suas potenciais aplicações na Ciência. Trata-se de uma pesquisa exploratória e teórica de caráter qualitativo. Os resultados apresentam o levantamento realizado sobre a Comunicação Científica e sua relação com a Ciência da Informação, incidindo na importância da gestão de dados e os metadados para a descrição dos conjuntos de dados de pesquisa, e o seu impacto com o *Blockchain*. Considera-se que o momento é prematuro para afirmarmos que a *Blockchain* ou qualquer uma das tecnologias em seu entorno, em seus estágios atuais, resolverão todos os problemas relacionados à crise de reprodutibilidade. No curto prazo a *Blockchain* provavelmente seja aplicada a pontos específicos no fluxo de uma pesquisa científica ajudando a dirimir a desconfiança nos dados e nas metodologias utilizadas. Uma vez que as aplicações relacionadas ao tema ainda estão em fases embrionárias de desenvolvimento, implementação e aceitação, não é possível nem mesmo afirmar que suas implementações estejam próximo dos objetivos a que foram projetados, possibilitando assim inúmeras oportunidades de pesquisas e de desenvolvimentos para a Ciência da Informação.

Palavras-chave: Curadoria de dados científicos. Ciclo de vida dos dados científicos. Gestão dos dados científicos. Ciência aberta. *Blockchain*.

ABSTRACT

The possibility of scientific discussion is the core of the practical research routine and, for this, researchers must find sufficiently detailed resources and descriptions in the published papers in order to obtain the same results within the enough margins of experimental errors. Among the available resources, the data collected during the research e summing it is essential, because the result obtained should be the product of its rigorous analysis. In this scope and associated with distributed consensus mechanisms, the *Blockchain* data structure can ensure data reliability in a transparent, decentralized and immutable way, and emerges as a possibility to contribute to the progress of science through new ways of processing large amounts of data collected in a scientific research. In this sense, the general objective of this work is to analyze *Blockchain* technology and its possibilities inserted in scientific communication and research data management. The specific objectives are to describe the management of research data and the data lifecycle; identify the main concepts, structure and functioning, related to *Blockchain* technology; and point out *Blockchain* technology and its potential applications in Science. This is an exploratory and theoretical research of qualitative character. The results present the survey conducted on Scientific Communication and its relationship with Information Science, focusing on the importance of data management and metadata for the description of research datasets, and their impact with *Blockchain*. It is considered that the moment is premature to affirm that *Blockchain* or any of the technologies in its surroundings, in its current stages, will solve all problems related to the reproducibility crisis. In the short term *Blockchain* is likely to be applied to specific points in the flow of scientific research helping to resolve mistrust in the data and methodologies used. Since the applications related to the theme are still in embryonic phases of development, implementation and acceptance, it is not even possible to affirm that their implementations are close to the objectives to which they were designed, thus enabling numerous opportunities for research and developments for Information Science.

Keywords: Scientific data curation. Scientific data lifecycle. Scientific data management. Open Science. *Blockchain*.

LISTA DE QUADROS

Quadro 1 - Lista de fontes e termos para o levantamento bibliográfico (exceto os relacionados ao termo <i>Blockchain</i>)	18
Quadro 2 - Lista de fontes e termos para o levantamento bibliográfico relacionados ao termo <i>Blockchain</i>	20
Quadro 3 - Atributos dos Dados Abertos	29
Quadro 4 - Comparativo entre as propostas de Heim e Gray sobre a Comunicação Científica e os Dados de Pesquisa	32
Quadro 5 - Benefícios da Ciência Aberta	33
Quadro 6 - Ideias e projetos relevantes para as questões de reprodutibilidade da pesquisa	75

LISTA DE FIGURAS

Figura 1- O Modelo de Ciclo de Vida de Curadoria do Digital Curation Centre	54
Figura 2 - Sistema centralizado, descentralizado e distribuído	61
Figura 3 - Uma função <i>hash</i> criptográfico em funcionamento.	64
Figura 4 - A <i>Blockchain</i> representada graficamente	Erro! Indicador não definido.
Figura 5 - Estrutura genérica da <i>Blockchain</i> do Bitcoin	Erro! Indicador não definido.

LISTA DE SIGLAS E ABREVIações

AI	Artificial Intelligence
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
BGP	Byzantine Generals Problem
BOAI	Budapest Open Access Initiative
BRAPCI	Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CI	Ciência da Informação
CVD	Ciclo de Vida dos Dados
DCC	Digital Curation Centre
DDI	Data Documentation Initiative
DLF	Digital Library Foundation
DMP	Data Management Platform
DPoS	Delegate Proof-of-Stake
FOSTER	Facilitate Open Science Training For European Research
GPS	Global Positioning System
IBFS	International Society of <i>Blockchain</i> For Science
ICO	Initial Coin Offering
IoT	Internet of Things
LISA	Library and Information Science Abstracts
LPoS	Leased Proof of Stake
MAC	Media Access Control
MEMEX	Memory Index
MRDF	Machine-Readable Data Files
NASA	National Aeronautics and Space Administration
NESTA	The National Endowment for Science, Technology and the Arts
OECD	Organisation for Economic Co-Operation and Development
P2P	Peer to Peer
PBFT	Practical Byzantine Fault Tolerance
PIN	Personal Identification Number
POI	Proof of Importance
PoS	<i>Proof of Stake</i>
POW	Proof-of-Work
RIN	The Research Information Network
SciELO	Scientific Electronic Library Online
SETI@home	Search for extraterrestrial intelligence @home
SPOF	Single Point of Failures
SSD	Solid-State Drive
TIC	Tecnologia de Informação e Comunicação
UFSCar	Universidade Federal de São Carlos
WWW	World Wide Web

SUMÁRIO

1 INTRODUÇÃO	14
1.1 Objetivos	17
1.2 Procedimentos metodológicos	17
1.3 Justificativa	23
1.4 Estrutura do trabalho	24
2 A COMUNICAÇÃO CIENTÍFICA E A CIÊNCIA DA INFORMAÇÃO	25
3 GESTÃO DOS DADOS E METADADOS DE PESQUISA	44
3.1 Dados de pesquisa	48
3.2 O ciclo de vida de dados	51
4 <i>BLOCKCHAIN</i>	58
4.1 A estrutura da <i>Blockchain</i>	59
4.2 A <i>Blockchain</i> e suas potenciais aplicações na Ciência	72
5 CONSIDERAÇÕES FINAIS	79
REFERÊNCIAS	84

1 INTRODUÇÃO

"[...] if I have seen further, it is by standing on the shoulders of giants." (NEWTON, 1675).

Originalmente atribuída a Bernardo de Chartres e tendo como versão mais conhecida a que utilizamos como epígrafe neste capítulo - a enviada por Sir Isaac Newton em carta a Robert Hooke em 1675 - a expressão “sobre ombros de gigantes” sugere que, nas atividades científicas, as trocas entre o pesquisador e os membros de uma comunidade, contemporâneos ou não, configuram-se em uma das mais importantes atividades da metodologia científica: a comunicação.

Partindo de uma perspectiva positivista, melhorar significa tornar superior o novo em relação ao seu estado anterior e configura-se como o alicerce dos avanços do conhecimento, ou seja, encontrar a verdade a partir de descobertas realizadas por predecessores simboliza a mecânica da evolução científica e o progresso da Ciência.

Por outro lado, ao definir a Ciência como o “Conhecimento Público”, Ziman (1979, p. 24) nos esclarece que o objetivo da Ciência é mais do que obter informações, mas alcançar um consenso de opinião racional entre o pesquisador e seus pares, então, não podemos nos limitar a conceituar a Ciência apenas como “[...] uma questão de sequência de gênios a realizarem uma sequência de importantes descobertas” (PRICE, 1976, p. 93).

A Ciência avança com base no conhecimento adquirido em estudos prévios e a confiabilidade e reprodutibilidade dos resultados juntos constituem um dos pilares da pesquisa científica. (GIBB, 2014; MCNUTT, 2014; COLLABORATION, 2015; NASSI-CALÒ, 2017).

Mueller (2006, p. 28) por este motivo afirma que em uma pesquisa, “por definição, legitimidade depende de consenso” entre os pares de sua comunidade, onde e quando seus fatos e teorias

[...] têm de passar por um crivo, por uma fase de análises críticas e de provas, realizadas por outros indivíduos competentes e desinteressados, os quais deverão determinar se eles são bastante convincentes para que possam ser universalmente aceitos. (ZIMAN, 1979, p. 24).

A replicação dos resultados de uma pesquisa por um indivíduo ou por terceiros torna o resultado válido, entretanto, replicar uma pesquisa pode não ser algo trivial levando-se em conta o quão complexo se tornou a atividade de pesquisa científica.

Peng (2011), argumenta que a reprodutibilidade de uma pesquisa “[...] tem o potencial de servir como um padrão mínimo para julgar alegações científicas quando a replicação independente completa de um estudo não é possível.” (PENG, 2011, p. 1226). Tornar uma pesquisa reprodutível é, portanto, um objetivo razoável e desejável, pois “[...] descobertas científicas que acabam não sendo validadas assombram cientistas, publicações e empresas.” (MARQUES, 2014, p. 57).

A título de exemplo, em uma pesquisa realizada com 1.576 pesquisadores publicada pela Revista *Nature* no ano de 2016, revela atitudes contraditórias sobre a reprodutibilidade de pesquisas científicas. A pesquisa apontou que mais de 70% dos pesquisadores não obtiveram sucesso em reproduzir experimentos de outros cientistas e mais da metade não conseguiram reproduzir seus próprios experimentos.

É o caso de Gibb (2014) que, por sua vez, se mostrou particularmente frustrado frente ao seu insucesso em tentar reproduzir seus próprios experimentos químicos utilizando os mesmos instrumentos, insumos e laboratório utilizado durante o experimento original. Esses e outros casos contribuem para que a preocupação com a reprodutibilidade da pesquisa científica aumente “[...] constantemente com relatos de que os resultados de experimentos em vários domínios da Ciência não poderiam ser replicados.” (GOODMAN; FANELLI; IOANNIDIS, 2016, p. 1).

Para Mayer e Zeviani (2016) o ato de reproduzir uma pesquisa significa possibilitar que outros cientistas, de áreas e pensamentos distintos e independentemente, possam obter acesso aos dados e aos métodos da pesquisa original, inclusive aos códigos computacionais, para que, ao menos, a análise dos dados e o método utilizado possam ser validados tornando-se uma alternativa entre replicar uma pesquisa e nada fazer. Cruz et al. (2018) afirmam que embora existam inúmeros problemas que dificultam o compartilhamento dos dados, existe “[...] uma necessidade latente dos pesquisadores com relação ao compartilhamento e preservação dos seus dados.” (CRUZ et al., 2018, p. 2776).

Para isso, pesquisadores de algumas das áreas do conhecimento vêm buscando novas formas e métodos para o gerenciamento de seus dados de pesquisa científica buscando a garantia da integridade e, principalmente mantendo o acesso e reuso aos mesmos.

Dessa necessidade, surgem inúmeras oportunidades de contribuições por parte da Ciência da Informação que, segundo Santana (2014), pode e deve contribuir para melhorias no acesso e uso intenso de dados,

[...] buscando identificar e estudar fatores e características que propiciem ampliação do equilíbrio entre os atores envolvidos no processo e a máxima otimização do uso dos dados. (SANTANA, 2014, p. 119).

Desde 2015, o Grupo de Pesquisa "Dados e Metadados" da Universidade Federal de São Carlos (GP-DM UFSCar), vem realizando pesquisas que contribuem com essa visão tendo como escopo a relação entre dados e informação. O principal objetivo do grupo é o desenvolvimento de pesquisas teórico-aplicadas sobre a organização, representação e manutenção da informação e de dados nos contextos da Arquivologia, Biblioteconomia, Museologia e Ciência da Computação.

Sendo a *Blockchain* uma estrutura de dados, cujos componentes são constituídos de tecnologias oriundas das áreas de estudo abrangidas pelo Grupo de Pesquisa Dados e Metadados, não poderia deixar de ser objeto de estudos deste grupo, da qual o autor deste trabalho é membro.

A *Blockchain* é uma estrutura de dados distribuída (livro-razão ou ledger), que contém dados de transações, registrados e posteriormente armazenados como blocos ligados a outros blocos - como uma corrente - de forma que esses registros se tornam invioláveis e atualizáveis apenas por consenso ou acordo entre pares. (ANTONOPOULOS, 2014; BASHIR, 2017; ROUSE, 2017), podendo “[...] ser auditado a qualquer momento [...]” (OLIVEIRA; SANTARÉM SEGUNDO, 2018, p. 5371).

Sendo a *Blockchain* uma estrutura de dados, com características singulares, que promete atribuir a confiabilidade desejável aos dados e conjuntos de dados de forma transparente, descentralizada e imutável, vislumbramos nela a possibilidade de sua utilização no armazenamento de dados de pesquisa científicas coletados por pesquisadores que, atualmente os mantêm “[...] armazenados localmente em seus

computadores, em algum meio digital de armazenamento ou em repositórios dos laboratórios de pesquisa.” (CRUZ et al., 2018, p. 2776).

Este trabalho, portanto, busca investigar e responder a seguinte questão de pesquisa: **quais são os benefícios da aplicação da *Blockchain* no contexto dos Dados de Pesquisas Científicas?**

1.1 Objetivos

O objetivo geral da pesquisa consiste em analisar a tecnologia *Blockchain* e suas possibilidades frente à comunicação científica e na gestão de dados de pesquisa.

Os objetivos específicos são:

- Descrever a gestão de dados de pesquisa e o ciclo de vida de dados;
- Identificar os principais conceitos, a estrutura e o funcionamento, relacionados à tecnologia *Blockchain*;
- Identificar as potenciais aplicações da tecnologia *Blockchain* na Ciência.

1.2 Procedimentos metodológicos

Trata-se de uma pesquisa exploratória e teórica de caráter qualitativo, que aborda os temas de: comunicação científica, *Blockchain*, curadoria de dados de pesquisa, ciclo de vida de dados de pesquisa, reprodutibilidade da pesquisa, ciência aberta, *dados* e metadados.

Quanto à pesquisa exploratória, esta proporcionou uma “[...] maior familiaridade com o problema, com vistas a torná-lo mais explícito ou construir hipóteses (GIL, 2007, p. 41). A análise exploratória e descritiva da literatura disponível sobre o tema proposto, permitiu a construção de um conhecimento teórico sobre as práticas de comunicação científica, curadoria digital, proposições e abordagens de Ciclo de Vida de Dados, definição e construção de metadados e padrões de metadados, utilização de tecnologias em pesquisas na Ciência, além claro, da *Blockchain*.

Para a contextualização teórica foram utilizadas as fontes bibliográficas e documentais como fundamentação para os resultados, por essa razão, a pesquisa

refere-se a uma pesquisa bibliográfica e documental. Gil (2002, p. 44) conceitua que “A pesquisa bibliográfica é desenvolvida com base em material já elaborado, constituído principalmente de livros e artigos científicos” e a pesquisa documental, para ele, diferencia-se pela natureza das fontes que, no caso, “[...] vale-se de materiais que não recebem ainda um tratamento analítico, ou que ainda podem ser reelaborados de acordo com os objetos da pesquisa. (GIL, 2002, p. 45).

O recorte bibliográfico sobre os temas desta pesquisa foi realizado pelos idiomas português, espanhol e inglês, no período de 2010 a 2019. O recorte temporal foi definido tendo em consideração a concentração bibliográfica publicada sobre os temas desta pesquisa e do período em que ocorreu o surgimento do assunto principal, qual seja, a *Blockchain*, que foi apresentada publicamente em um artigo científico no ano de 2008 e tendo sua efetiva implementação e adoção em 2010 gerando, à partir de então, publicações que consideramos relevantes para o presente estudo, no entanto, alguns trabalhos clássicos ou de relevância histórica para a pesquisa foram citados.

Durante a pesquisa foram utilizadas as bases de dados: Biblioteca Digital Brasileira de Teses e Dissertações (BDTD); Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI); *Library and Information Science Abstracts* (LISA); Portal de Periódicos da CAPES; *Scientific Electronic Library Online* (SciELO), *Scopus*, e *Web of Science*. Além de outras fontes de informação como o Repositório Institucional da Universidade Federal de São Carlos (UFSCar) e, se possível, de outras Universidades, postagens em blogs, artigos em revistas e sites institucionais, anais de eventos científicos e *white papers* de instituições públicas e privadas.

As etapas de pesquisa foram divididas em:

1ª etapa: Levantamento bibliográfico e seleção do material obtido - buscando informações pertinentes ao tema proposto e que criaram embasamento teórico ao estudo. O levantamento bibliográfico foi efetuado em fontes bibliográficas (primárias e secundárias), levando em consideração materiais impressos e digitais, abrangendo todos os objetivos do trabalho.

As consultas foram realizadas de acordo com os termos, bem como seus correspondentes em português, elencados no Quadro 1 e no Quadro 2. A atualização mais recente foi realizada em maio de 2019. A escolha pela separação

dos termos em dois quadros visa destacar as buscas pelas relações do termo *Blockchain* com a pesquisa científica e mais especificamente com as questões relacionadas à reprodutibilidade da pesquisa. Ressalta-se também a utilização de operadores booleanos e filtros para o refinamento de acordo com os recursos e características disponíveis em cada fonte de informação. O levantamento documental se deu de forma similar, através dos mesmos conjuntos de termos em ferramentas genéricas de pesquisa na internet.

Quadro 1 - Lista de fontes e termos para o levantamento bibliográfico (exceto os relacionados ao termo *Blockchain*)

Fontes		Termos	Número de Publicações	Período com publicação
P o r t a l D e	Library and Information Science Abstracts (LISA)	<i>“scientific communication”</i>	388	2010 a 2019
		<i>“data curation”</i>	361	2010 a 2019
		<i>“research data curation”</i>	41	2010 a 2019
		<i>“scientific research data”</i>	35	2010 a 2019
		<i>“data life cycle”</i>	54	2010 a 2019
		<i>reproducibility</i>	308	2010 a 2019
		<i>“open science”</i>	249	2010 a 2019
P e r i ó d i c o s	Information Science & Technology Abstracts (ISTA)	<i>“scientific communication”</i>	148	2010 a 2019
		<i>“data curation”</i>	66	2010 a 2018
		<i>“research data curation”</i>	3	2014 a 2017
		<i>“scientific research data”</i>	7	2010 a 2019
		<i>“data life cycle”</i>	2	2011 a 2015
		<i>reproducibility</i>	49	2011 a 2019
		<i>“open science”</i>	48	2010 a 2019
C A P E S	Library, Information Science & Technology Abstracts	<i>“scientific communication”</i>	58	2015 a 2019
		<i>“data curation”</i>	181	2010 a 2019
		<i>“research data curation”</i>	7	2012 a 2019
		<i>“scientific research data”</i>	12	2010 a 2019
		<i>“data life cycle”</i>	13	2010 a 2019
		<i>reproducibility</i>	98	2010 a 2019
		<i>“open science”</i>	167	2010 a 2019

SCOPUS	<i>“scientific communication”</i>	1159	2010 a 2019
	<i>“data curation”</i>	998	2010 a 2019
	<i>“research data curation”</i>	8	2010 a 2019
	<i>“scientific research data”</i>	61	2010 a 2019
	<i>“data life cycle”</i>	153	2010 a 2019
	<i>reproducibility</i>	231539	2010 a 2019
	<i>“open science”</i>	1413	2010 a 2019
Web of Science	<i>“scientific communication”</i>	626	2010 a 2019
	<i>“data curation”</i>	298	2010 a 2019
	<i>“research data curation”</i>	4	2016 a 2019
	<i>“scientific research data”</i>	20	2010 a 2019
	<i>“data life cycle”</i>	51	2010 a 2019
	<i>reproducibility</i>	51638	2010 a 2019
	<i>“open science”</i>	600	2010 a 2019
Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)	<i>“scientific communication”</i>	276	2010 a 2019
	<i>“data curation”</i>	6	2012 a 2017
	<i>“research data curation”</i>	1	2014
	<i>“scientific research data”</i>	10	2011 a 2018
	<i>“data life cycle”</i>	10	2012 a 2019
	<i>reproducibility</i>	1284	2010 a 2019
	<i>“open science”</i>	20	2011 a 2018
Repositório Institucional da UFSCar	<i>“scientific communication”</i>	16	2010 a 2018
	<i>“scientific research data”</i>	1	2012
	<i>reproducibility</i>	159	2010 a 2019
	<i>“open science”</i>	1	2018
Scientific Electronic Library Online (SciELO)	<i>“data curation”</i>	15	2014 a 2018
	<i>“research data curation”</i>	9	2012 a 2017
	<i>“scientific research data”</i>	1732	2010 a 2019
	<i>“data life cycle”</i>	323	2010 a 2019
	<i>reproducibility</i>	1451	2010 a 2019
	<i>“open science”</i>	506	2010 a 2019
Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI)	<i>“data curation”</i>	17	2010 a 2019
	<i>“research data curation”</i>	12	2010 a 2019
	<i>“scientific research data”</i>	157	2010 a 2019
	<i>“data life cycle”</i>	8	2011 a 2018

	<i>reproducibility</i>	4	2010 a 2018
	<i>“open science”</i>	135	2010 a 2019

Fonte: elaborado pelo autor

Quadro 2 - Lista de fontes e termos para o levantamento bibliográfico relacionados ao termo *Blockchain*

Fontes		Termos	Qtd.	Período
P o r t a l d e P e r i ó d i c o s C A P E S	Library and Information Science Abstracts (LISA)	<i>Blockchain</i>	65	2015 a 2019
		<i>Blockchain AND research</i>	43	2015 a 2019
		<i>Blockchain AND reproducibility</i>	5	2018 e 2019
	Information Science & Technology Abstracts (ISTA)	<i>Blockchain</i>	35	2016 a 2019
		<i>Blockchain AND research</i>	15	2018 e 2019
		<i>Blockchain AND reproducibility</i>	3	2019
	Library, Information Science & Technology Abstracts	<i>Blockchain</i>	58	2015 a 2019
		<i>Blockchain AND research</i>	27	2016 a 2019
		<i>Blockchain AND reproducibility</i>	3	2019
	SCOPUS	<i>Blockchain</i>	342	2010 a 2019
		<i>Blockchain AND research</i>	7	
		<i>Blockchain AND reproducibility</i>	197	2013 a 2019
Web of Science	<i>Blockchain</i>	215	2013 a 2019	
	<i>Blockchain AND research</i>	6		
	<i>Blockchain AND reproducibility</i>	393	2015 a 2019	
Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)	<i>Blockchain</i>	4	2017 a 2019	
	<i>Blockchain AND research</i>	15	2016 a 2019	
Repositório Institucional da UFSCar	<i>Blockchain</i>	7	2017 a 2019	
	<i>Blockchain</i>	1	2018	

Scientific Electronic Library Online (SciELO)	<i>Blockchain</i>	4	2018
Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI)	<i>Blockchain</i>	3	2017 e 2018

Fonte: elaborado pelo autor

O material foi selecionado considerando-se critérios de relevância apontados pelos mecanismos de busca das bases de dados consultadas, título, resumo, palavra-chave, bem como, citação em fontes documentais.

2ª etapa: Leitura, interpretação, análise e sistematização das informações - a fim de oportunizar o desenvolvimento da base teórica para a discussão dos diferentes pontos de vista identificados na literatura sobre o tema proposto, abrangendo assim o primeiro objetivo dessa pesquisa.

3ª etapa: Análise e estabelecimento das características fundamentais extraídas da literatura - para estabelecimento da solução do problema de pesquisa, verificando as características principais encontradas na literatura, abrangendo assim o primeiro objetivo dessa pesquisa.

4ª etapa: Sistematização do estudo exploratório – a fim de traçar um histórico do desenvolvimento e do progresso da Comunicação Científica, a fim de identificar as iniciativas que vêm contribuindo para o progresso da Ciência frente às atuais configurações decorrentes do impacto da adoção das tecnologias de comunicação e informação; abrangendo assim os objetivos específicos dessa pesquisa.

5ª etapa: Elaboração e redação final da pesquisa - a ser desenvolvida para divulgação à comunidade científica dos resultados obtidos com o desenvolvimento do estudo em questão, abrangendo assim todos os objetivos dessa pesquisa.

6ª etapa: Divulgação da pesquisa – essa etapa não constitui unicamente do momento final da pesquisa, mas sim, de um processo construtivo e contínuo conjunto do orientador e aluno.

1.3 Justificativa

A Ciência da Informação, visto sua natureza interdisciplinar (BORKO, 1968), traz importantes contribuições para a Comunicação Científica fornecendo fontes de informações para auxiliar na obtenção de respostas específicas a questionamentos dos cientistas; concorrendo para a atualização do cientista em seu campo de atuação ou estimulando a descoberta e a compreensão de novos campos além de divulgar novas tendências de áreas emergentes de serviços de disseminação seletiva de informação e eventos; apresentando relatórios bibliométricos, cientométricos e ou alométricos atestando a relevância do trabalho do cientista, fornecendo um feedback para o aperfeiçoamento da produção do pesquisador.

A Ciência pode ser considerada uma jornada individual e ao mesmo tempo um empreendimento coletivo e da Comunicação Científica, provoca no indivíduo e na coletividade melhorias consensuais e avanços nas fronteiras do conhecimento: o progresso científico.

Os impactos das tecnologias de informação e comunicação na Ciência são irrefutáveis, no entanto, como argumenta Pinheiro (2014), os atuais avanços nas metodologias científicas, tendo os dados como elemento central, são resultados do aumento da complexidade das Ciências, “[...] com equipes mais numerosas trabalhando cada vez mais em colaboração, inclusive internacionalmente [...]. Portanto, não se trata apenas de um fenômeno tecnológico, mas científico e político, entre outras instâncias.” (PINHEIRO, 2014, p. 159).

[...] sem legitimação, sem o consenso da comunidade, nenhuma proposta terá efeito ou chances de provocar mudanças significativas no atual sistema de Comunicação Científica, não importa a quão inovadora ou o quão formidável sua contribuição potencial. (MUELLER, 2006, p. 37).

Justifica-se este trabalho, uma vez que “[...] ocorrências cada vez mais frequentes de retratação de pesquisas científicas sugerem que o erro, o falseamento de dados, a omissão e a manipulação ameaçam a qualidade das

evidências que embasam as publicações.” (DUDZIAK, 2018, *on-line*). Assim sendo, há a necessidade de estudos que ajudem a minimizar esses problemas.

A *Blockchain* pode ser uma possibilidade para, se não para resolver, minimizar os impactos dos problemas, uma vez que os dados armazenados na *Blockchain* adquirem características singulares, sendo armazenados univocamente e transparentemente, permitindo-se que auditorias sejam realizadas.

1.4 Estrutura do trabalho

A presente dissertação é apresentada em quatro capítulos, sendo este o primeiro. Neste capítulo, foi delimitado o objeto, o tema e a questão de pesquisa, além dos objetivos, justificativa e os procedimentos metodológicos que nortearam essa pesquisa.

No **Capítulo 2 - A Comunicação Científica e a Ciência da Informação** buscamos traçar um rápido panorama geral da Comunicação Científica, sua função e importância para o progresso científico, bem como a sua evolução, desde os gregos antigos até aos atuais movimentos de Acesso Livre e Ciência Aberta. Neste capítulo também abordamos as relações entre a Ciência da Informação e a temática dos dados abertos e, finalmente, abordamos as relações entre os dados de pesquisa, a repetição, a replicação e a reprodução de uma pesquisa científica.

Os temas relacionados à gestão dos dados são abordados no **Capítulo 3 - Gestão dos dados de pesquisa**. Tendo como pano de fundo as inúmeras inovações tecnológicas que, de forma incremental ou disruptiva, impactaram os temas relacionados ao conhecimento, à informação, dados e metadados, descrevemos alguns dos instrumentos e modelos criados para minimizar os impactos negativos e ampliar os benefícios da boa gestão dos dados, incluindo sua descrição e documentação.

A *Blockchain*, seus atributos e estrutura, em especial, os Mecanismos de Consenso são abordados no **Capítulo 4 - Blockchain**. A partir do conhecimento dos aspectos gerais da *Blockchain*, buscou-se identificar suas possíveis aplicações na Ciência, principalmente os com características promissoras no que se refere à reprodutibilidade das pesquisas científicas. E, finalmente, o **Capítulo 5** tece as Considerações Finais.

2 A COMUNICAÇÃO CIENTÍFICA E A CIÊNCIA DA INFORMAÇÃO

A pedra angular da Ciência é “[...] a possibilidade de avaliar criticamente a veracidade das afirmações científicas efetuadas e das conclusões tiradas por outros cientistas”. (PLESSER, 2018, p. 1). Idealmente, para ele, outros cientistas, com capacidades e meios, devem encontrar descrições detalhadas o suficiente em trabalhos publicados para seguirem as etapas descritas e obter os mesmos resultados dentro das margens de erro experimentais suficientes.

Nesse escopo, a Comunicação Científica é essencial para que os avanços das diversas áreas do conhecimento sejam identificados, bem como que os membros de sua comunidade sejam conhecidos, analisados, reconhecidos, aceitos ou negados, permitindo assim, novos conhecimentos e o progresso da Ciência. (LE COADIC, 1996; MONTEIRO; SANT’ANA; SANTAREM SEGUNDO, 2016).

A partir desse viés, a Comunicação Científica,

[...] inclui todos os espectros das atividades associados à produção, disseminação e uso da informação a partir do momento que o cientista tem a ideia para sua pesquisa, até as informações sobre os resultados desta pesquisa para que os mesmos sejam aceitos como um componente do conhecimento científico. (GARVEY, 1979 *apud* ROSA; GOMES, 2010, p. 19).

Rosa e Gomes (2010) enumeram as funções da Comunicação Científica que consistem em 1) fornecer respostas a perguntas específicas, 2) concorrer para a atualização profissional do cientista no campo específico de sua atuação, 3) estimular a descoberta e a compreensão de novos campos de interesse, 4) divulgar as tendências de áreas emergentes, fornecendo aos cientistas ideia da relevância de seu trabalho, testar a confiabilidade de novos conhecimentos diante da possibilidade de testemunhas e verificações, por fim, 5) redirecionar ou ampliar o interesse dos cientistas e fornecer um feedback para o aperfeiçoamento da produção do pesquisador (ROSA; GOMES, 2010, p. 19).

Meadows (1999) atribui aos gregos antigos o pioneirismo nas atividades que impactaram na Comunicação Científica Moderna. Os primeiros séculos da Ciência mantiveram as discussões no meio acadêmico pela troca de conhecimento em ambientes menos formais como os simpósios e por obras manuscritas como as de Aristóteles.

A Comunicação Científica sofreu e vem sofrendo diversos impactos tecnológicos: Primeiro com a imprensa e, depois, com a internet.

A criação da tecnologia de impressão foi muito importante no desenvolvimento das forças produtivas na sociedade, ao facilitar a circulação da mesma informação com um alcance sem precedentes. Inicia-se, então, um processo de Comunicação Científica, na medida em que a produção de conhecimentos gera, por sua vez, a necessidade de novos conhecimentos. Isso merece ser destacado, uma vez que nesse período foram iniciadas as bases da nossa sociedade atual, que começou a ser construída quando as associações científicas foram criadas e os primeiros periódicos científicos são publicados, dando início à formalização do processo de Comunicação Científica. A partir desse período a atividade científica, baseada no método de experimentação e na matemática, estabelece um novo paradigma social e tecnológico. Nesse contexto, desenvolve sua própria forma de expressão e canais de comunicação e são criadas as primeiras associações científicas. (FREIRE, 2006, p. 8).

Nesse contexto, afirma-se que a Comunicação Científica era transmitida por meio de cartas e copiadas manualmente entre os membros dessas associações, que passou a ser distribuir seus relatórios em grandes volumes para um número maior de pessoas, sendo em sua decorrência e à partir da segunda metade do século XVII, uma verdadeira revolução na forma como ideias eram disseminadas, sendo essa, portanto a origem dos primeiros periódicos científicos. “O motivo principal [do surgimento das revistas científicas] encontra-se nessa necessidade de comunicação, do modo mais eficiente possível [...]”. (MEADOWS, 1999, p. 7).

Os periódicos científicos estabeleceram hegemonia como principal veículo de comunicação entre pesquisadores. Como também, foi estabelecida uma política de destaque de cada periódico, Mueller (2006, p. 31) destaca que “[...] o periódico indexado costuma ser o veículo mais prestigiado [...]”. Pinheiro (2014) relembra que na década de 80 do século XX tais veículos passaram por uma crise decorrente dos elevados valores exigidos pelas editoras para acesso aos seus conteúdos, abrindo caminho para o surgimento do acesso livre à informação científica que eclodiu como uma reação ao domínio das editoras no processo editorial de periódicos científicos (PINHEIRO, 2014, p. 156).

O termo de acesso aberto foi utilizado pela primeira vez na publicação da Budapest Open Access Initiative (BOAI) em 2001, que define uma contemporaneidade para o termo, bem como a sua importância e práticas, e dentre as recomendações é registrado que haja o

[...] desenvolvimento de políticas de Acesso Aberto em instituições de educação superior e agências patrocinadoras, o acesso aberto de trabalhos acadêmicos, o desenvolvimento de infraestrutura como os repositórios de Acesso Aberto e a criação de um padrão de conduta profissional para a publicação em Acesso Aberto (BOAI FORUM, 2002, p.1, tradução nossa).

Nesse contexto, Pinheiro (2014, p. 156) advoga que “[...] a produção científica deveria ser livre de custos”, não bastando apenas acessar a informação como no caso do chamado Acesso Aberto (Open Access), mas que “[...] esse acesso não represente custos e esteja livre das restrições de direitos de autoria e licenciamentos”, trazendo inúmeros benefícios como “[...] acelerar a pesquisa, enriquecer a educação, compartilhar o aprendizado dos ricos com os pobres e os pobres com os ricos, tornando essa literatura tão útil quanto possível.” (BOAI FORUM, 2002, *on-line*) e, por isso, o “[...] movimento para acesso livre ao conhecimento científico pode ser considerado como o fato mais interessante e talvez importante de nossa época no que se refere à Comunicação Científica.” (MUELLER, 2006, p. 27).

Pinheiro (2014) declara que antes do surgimento do movimento de acesso livre à informação científica e da internet “[...] pouco ou quase nada se discutia ou pesquisava sobre a hoje denominada Ciência Aberta.” (PINHEIRO, 2014, p. 159).

Do lado das comunidades científicas, está a crescer a sensibilidade para as consequências da verdadeira explosão na produção de dados científicos, quer pelo crescimento global das atividades de investigação, quer como consequência dos novos métodos e instrumentos de pesquisa e registo que originam cada vez maiores volumes de dados. (RIBEIRO et al., 2010, p. 8).

Para Sayão e Sales (2014) o acesso livre é “[...] um pilar de importância crítica para a prática de uma Ciência mais aberta [...]”, atribuem à Declaração de Berlin sobre o acesso livre ao conhecimento em ciência e humanidades de 2003, a ampliação do movimento de acesso livre a outros formatos, em destaque aos dados brutos e metadados.

A Ciência Aberta representa um alargamento do acesso livre, tornando acessíveis dados científicos, únicos e insubstituíveis, dos mais diversos tipos, básicos para pesquisas, mas em geral não publicados, a outros e futuros pesquisadores, para a sua reutilização. (PINHEIRO, 2014, p. 163).

Em decorrência de novas alternativas para comunicar a informação científica, quase sempre envolvendo as Tecnologias de Comunicação e Informação (TIC), em especial à internet e a *World Wide Web* e que deram início ao surgimento dos primeiros periódicos eletrônicos na década de 90, além de servidores de *e-prints*, repositórios institucionais, repositórios de dados, publicações de dados e o próprio auto arquivamento em páginas pessoais, além dos resultados de pesquisas científicas em artigos publicados em periódicos, criaram as condições para a concretização do acesso livre e sua ampliação para uma Ciência Aberta. Pinheiro (2014) acrescenta ainda que “[...] é fundamental reconhecer que esta questão transcende as ferramentas tecnológicas [...] entre os muitos fatores e componentes envolvidos neste complexo fenômeno.” (PINHEIRO, 2014, p. 163) e preconiza que “[...] da sua convergência talvez seja possível a utopia da ‘Ciência de multidão’ tornar-se uma realidade no mundo contemporâneo. (PINHEIRO, 2014, p. 163).

Nesse contexto, os recentes avanços tecnológicos também transformaram a maneira de como comunicar, abarcando o formato textual tradicional além de outros objetos relacionados, denominando assim, a publicação ampliada como parte do processo de publicação de dados. “Nesse processo histórico, os profissionais de informação vão estar sempre próximos dos cientistas, seja na organização e na armazenagem das informações produzidas, seja na sua comunicação.” (FREIRE, 2006, p. 8).

Para isso, esse tipo de publicação possibilita que os conjuntos completos de dados sejam disponibilizados conjuntamente às publicações em repositórios de dados digitais disponíveis na internet, por universidades e editoras. “As comunidades científicas produzem dados e informações que são registrados através das publicações científicas.” (MONTEIRO; SANT’ANA; SANTAREM SEGUNDO, 2016, p.12).

A explosão informacional, mais especificamente das produções científicas, foi um dos fenômenos que impulsionaram a gênese da Ciência da Informação e o que percebemos é um retorno à problemática, agora, centrado na unidade “dado”, ou melhor, nos conjuntos de dados.

Para a *Open Knowledge International* (2018), os dados abertos podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa desde que se respeite as exigências de atribuição de fontes e compartilhamentos pelas mesmas

regras e que possuem as seguintes características, conforme apresentados no Quadro 3.

Quadro 3 - Atributos dos Dados Abertos

Atributos	Descrição
Disponibilidade e Acesso	Os dados devem estar disponíveis como um todo e sob custo não maior que um custo razoável de reprodução, preferencialmente possíveis de serem baixados pela internet. Os dados devem também estar disponíveis de uma forma conveniente e modificável.
Reutilização e Redistribuição	Os dados devem ser fornecidos sob termos que permitam a reutilização e a redistribuição, inclusive a combinação com outros conjuntos de dados.
Participação Universal	Todos devem ser capazes de usar, reutilizar e redistribuir - não deve haver discriminação contra áreas de atuação ou contra pessoas ou grupos. Por exemplo, restrições de uso 'não-comercial' que impediriam o uso 'comercial', ou restrições de uso para certos fins (ex.: somente educativos) excluem determinados dados do conceito de 'abertos'.

Fonte: Open Knowledge International (2018).

De acordo com Lawrence et al. (2011), a base do método científico é a coleta e a análise de dados, porém a qualidade da pesquisa é julgada pelas conclusões extraídas da revisão por pares e não pela qualidade desses dados. “Em geral as publicações tradicionais não coadunam com a agenda atual da Ciência Aberta [...]” (CURTY; AVENTURIER, 2017).

Isso resulta em comunidades que nem sempre avaliam muito o esforço necessário para produzir dados e que nem sempre podem reproduzir sua análise porque os dados não são bem descritos ou preservados. [...]. Geralmente, isso ocorre porque os detalhes de tais atividades são normalmente considerados muito volumosos, muito derivados e / ou muito técnicos para publicação em um periódico. (LAWRENCE et al., 2011, p. 6).

As limitações como a visibilidade parcial dos dados primários (que são os dados brutos originais de uma pesquisa), designam para um contexto e percursos

metodológicos condensados, formato de publicação herméticos que impossibilitam a interligação com recursos complementares. (CURTY; AVENTURIER, 2017). A partir disso fizeram surgir modelos como as publicações ampliadas, clarificando a idealização das pesquisas, seus métodos e materiais, bem como o conjunto de dados obtidos na pesquisa.

Do ponto de vista dos dados científicos, estamos muito interessados em definir um processo de publicação de dados que tenha a capacidade de fazer afirmações sobre a confiabilidade e adequação para fins específicos dos dados que são entendidos pelos consumidores de dados. (LAWRENCE et al., 2011, p. 7).

Além da publicação, há as questões relacionadas às citações dos conjuntos de dados, uma vez que ela ainda é a forma mais aceita de atribuição da qualidade e de confiabilidade de uma informação. Em 2014, a *FORCE11* publicou um manifesto assinado por grandes editores no qual, além de outros compromissos, assume que “[...] os dados devem ser considerados produtos de pesquisa legítimos e passíveis de citação” (DATA CITATION SYNTHESIS GROUP, 2013). Não surpreende tal posicionamento, uma vez que o trabalho de coleta, higienização, organização, estruturação e publicação de conjuntos de dados ou *datasets* munidos de suas respectivas descrições (metadados), podem ser considerados como produtos legítimos de esforço de um ou mais pesquisadores e passível de reconhecimento pela comunidade acadêmica como um produto de pesquisa.

A partir da complementação do artigo manuscrito por módulos conectados de arquivos de dados executáveis e interligados, as publicações ampliadas permitem que não só os artigos científicos sejam avaliados de modo mais interativo, mas também potencializam o reuso de dados científicos de modo mais eficiente e eficaz. (CURTY; AVENTURIER, 2017, p. 9).

Em relação ao reuso dos dados, Santos e Santana (2013, p. 208) destacam que a aplicação e o estudo dos dados no âmbito da Ciência da Informação ainda é uma novidade, mesmo que com potencial de gerar avanços significativos para as pesquisas da área no país.

No exterior, uma publicação da *Graduate School of Library and Information Science da Universidade de Illinois At Urbana-Champaign* na *Library Trends* de 1982, sob o título “*Data libraries for the Social Sciences*”, editada por Heim (1982, p. 322), já se preocupava com a temática e observava que “[...] o arquivo de dados

e o desenvolvimento de bibliotecas de dados foram o foco de intenso debate nacional e internacional no final dos anos 60 e início dos anos 70. ”.

Gray (2009), mais recentemente, expôs uma cronologia em que divide a história da Ciência em quatro fases, denominada como Paradigmas da Ciência. Para ele, em sua fase inicial há milhares de anos atrás, a Ciência nasce empírica, quando a descrição de fenômenos naturais pelos povos primitivos que faziam Ciência como “[...] um meio racional de correlacionar os fatos observados”, evoluindo posteriormente pelas primeiras noções de astronomia no Egito, a invenção da escrita na Mesopotâmia, a matemática na Grécia e assim por diante. Desta, a Ciência evoluiu nos últimos cem anos para um ramo teórico com a utilização de modelos e generalizações para gerar teorias que, nas últimas décadas evoluiu para o terceiro paradigma da Ciência, qual seja, a computacional, onde fenômenos complexos são simulados pelos computadores. Por fim, o quarto paradigma, no qual os dados são capturados por instrumentos ou gerados por simulações que unificam a teoria e a experimentação, possibilitando ao pesquisador o gerenciamento dos dados e a análise estatística.

O modelo apresentado por Gray (2009) possui diversos pontos em comum com a proposta publicada em 1982 por Heim. O trabalho de Heim (1982) enumera questões relacionadas aos dados e acrescenta que, para que os cientistas sociais pudessem postular teorias em níveis altamente gerais, os dados primários ou brutos necessários para desenvolvê-las tiveram que passar por quatro níveis: 1) Fatos e figuras desarticuladas; 2) Dados históricos; 3) Novos métodos para reunir dados parciais e setoriais, juntamente com técnicas quantitativas de organização e interpretação e 4) Surgimento de múltiplos métodos e bases de dados complexos que possibilitaram a análise estatística dos dados, armazenados em coleções de arquivos legíveis por máquinas, as *Machine-Readable Data Files* (MRDF) e acrescenta que:

A falha das bibliotecas tradicionais em considerar arquivos de dados legíveis por máquina como dentro de sua área de aplicação fez com que o desenvolvimento da biblioteca de dados ocorresse, principalmente, fora das configurações tradicionais da biblioteca. Pior ainda, a maioria dos bibliotecários em grandes bibliotecas de pesquisa nem sequer reconhece o desserviço que eles fazem à sua clientela ao omitir o fornecimento de informações nessa área. (HEIM, 1982).

Assim, o quadro 4 reflete sobre comparativo dos pontos mais importantes das duas propostas.

Quadro 4 - Comparativo entre as propostas de Heim e Gray sobre a Comunicação Científica e os Dados de Pesquisa

Nível ou Paradigma	Níveis de Heim (1982)	Paradigmas de Gray (2009)
1	Fatos e figuras desarticuladas.	Ciência nasce empírica.
2	Dados históricos.	Ciência evolui nos últimos cem anos para um ramo teórico.
3	Novos métodos para reunir dados parciais e setoriais, juntamente com técnicas quantitativas de organização e interpretação.	A Ciência evoluiu para o paradigma Computacional.
4	Surgimento de múltiplos métodos e bases de dados complexos que possibilitaram a análise estatística dos dados, armazenados em coleções de arquivos legíveis por máquinas, as <i>Machine-Readable Data Files</i> (MRDF).	Dados são capturados por instrumentos ou gerados por simulações que unificam a teoria e a experimentação, possibilitando ao pesquisador o gerenciamento dos dados e a análise estatística.

Fonte: Elaborado pelo autor.

Para tanto, o quadro 4 traz as características definidas à Comunicação Científica atual, comparada a forma do passado. As formas tecnológicas podem interagir como otimizar os métodos e vivências da Ciência, nota-se que estamos “Vivenciamos o quarto paradigma da Ciência, o qual tem redefinido o *modus operandi* da práxis científica como consequência dos desafios impostos pela produção de dados em larga escala”. (CURTY; CERVANTES, 2016, p. 1).

O desenvolvimento de novos aparatos científicos, instrumentos, sensores, escalas e o uso intensivo de modelos de simulação da natureza geram uma quantidade imensa de dados, delineiam as fronteiras de um novo paradigma científico – conhecida como eScience -, criam novas metodologias de produção de conhecimento científico, formas de compartilhamento e de socialização entre os pesquisadores e alteram o fluxo tradicional de

Ciência da Informação e de revisão por pares. (SAYÃO; SALES, 2017, p. 67).

Dentre os benefícios de uma Ciência Aberta estão, segundo o relatório da *The Research Information Network* (RIN) e *The National Endowment for Science, Technology and the Arts* (NESTA) (2010), estão:

Quadro 5 - Benefícios da Ciência Aberta

Nome	Benefício	Estudo de Caso (RIN/NESTA)
Eficiência da pesquisa	Evitando a duplicação de esforço, tempo e recursos por meio de boas práticas e desenvolvimento e adoção de ferramentas de pesquisa e padrões abertos.	Na bioinformática, a Ciência Aberta permitiu que cientistas recuperaram de três fontes de dados separadas, uma coleção de genes compartilhados com o mesmo perfil de expressão gênica, economizando o esforço de pesquisa em cada um dos 50 a 70 genes individualmente. Na tecnologia de linguagem, o acesso completamente aberto dos dados ao público trouxe para uma das iniciativas considerável visibilidade, uma vez que se tornou o tema de centenas de artigos publicados.
Qualidade e rigor acadêmico da pesquisa	Todo o preparo e documentação dos conjuntos de dados para a análise e reutilização incentivam a atenção aos detalhes e à garantia de qualidade, embora exijam tempo e esforço. Desta forma, o potencial de uma revisão mais rigorosa é um motivador para aqueles envolvidos uma vez que contribuiria para um aumento na qualidade das pesquisas.	O potencial para revisão e escrutínio mais rigoroso tem feito com que grupos de pesquisas se motivem a manter registros mais precisos de suas atividades. Em astronomia, por exemplo, acredita-se que a qualidade dos dados mantidos e disponíveis nos <i>data centers</i> melhoraram drasticamente na última década, embora os próprios centros precisem de mais capacidade para garantir a documentação de alta qualidade. E na química, a ferramenta <i>LabBlog</i> é vista como um registro de proveniência de dados, vinculando efetivamente laboratórios, <i>data centers</i> e artigos em periódicos acadêmicos.

<p>Visibilidade e abertura para engajamento do público geral</p>	<p>Uma abordagem aberta no processo de pesquisa abre oportunidade de envolvimento do público em geral, contribuindo para uma abordagem da “Ciência cidadã”, compartilhando a carga de trabalho de coleta de dados e obviamente, dando maior visibilidade à pesquisa, que pode se reverter, inclusive, em novos financiamentos.</p>	<p>A publicação de dados está se tornando amplamente aceita, particularmente em áreas como a astronomia, cristalografia e tecnologia da linguagem, ampliando a visibilidade dos pesquisadores e trazendo novas oportunidades de financiamento. Na astronomia, projetos vêm atraindo a colaboração do público numa abordagem de “Ciência cidadã”.</p>
<p>Novas questões de pesquisa</p>	<p>As estruturas colaborativas criadas por meio da Ciência Aberta, torna viável a pesquisa questionar ou adotar abordagens impraticáveis de outras formas.</p>	<p>Na astronomia, o projeto Astrogrid desenvolveu a infraestrutura de projetos de observatórios virtuais, permitindo uma abordagem muito mais ambiciosa à análise de dados e à busca de padrões.</p> <p>O trabalho do grupo de bioinformática de imagens por meio do <i>Linking Open Drug Data</i>, demonstra novas capacidades para encontrar correlações, integrando semanticamente a marcação de dados de fontes abertas e torná-los pesquisáveis, por meio da vinculação de dados.</p> <p>Em química, a mineração de dados de repositórios abertos é relatada como tendo impacto significativo em áreas como a modelagem preditiva. Um grupo de pesquisas em neuroimagem busca explorar o potencial para melhoria no diagnóstico, investigando as associações entre imagens e dados genômicos, mais especificamente, o grupo de epidemiologia visa, por meio da integração de fontes abertas de dados geoespaciais e de saúde pública, visualizar a incidência geográfica da doença do sono.</p>

Colaboração e fortalecimento da comunidade	O compartilhamento de dados, <i>software</i> e outras ferramentas reduz-se as barreiras ao trabalho interinstitucional e interdisciplinar do compartilhamento de dados, <i>software</i> e outras ferramentas. Este, talvez, seja um dos benefícios ainda não plenamente explorado por ainda não estar totalmente difundido.	As técnicas desenvolvidas dentro da área de astronomia, utilizando uma infraestrutura aberta teve enorme impacto em várias áreas de análise de imagens médicas. Da mesma forma, o trabalho de outro grupo de bioinformática da imagem levou a iniciativas interdisciplinares com pesquisadores e historiadores de artes interessados em novos recursos de visualização de dados em diferentes fontes de dados. O grupo de química vem colaborando com neurocientistas e o grupo de tecnologia linguística com psicólogos.
Impacto social e econômico	Grupos de pesquisa avançam para abordagens abertas a pesquisas e inovações comerciais. Indústrias tradicionais, como a farmacêutica, perceberam que o custo de se realizar internamente toda a pesquisa está se tornando proibitivo. Movimentos de acesso a dados abertos têm crescido e trazido bons resultados em áreas como saúde pública, farmacêutica, linguagens e astronomia.	Na astronomia, colaboração entre analistas de imagens médicas e astrônomos levou à automação de análise de imagens de microarranjos de tecidurais, novas técnicas de análise de imagem para detecção de câncer de mama e análise de imagens cerebrais por ressonância magnética. Nas tecnologias de linguagem, o compartilhamento aberto de corporas e de ferramentas, desempenhando um papel no desenvolvimento de <i>software</i> de reconhecimento de fala. Movimentos em direção ao acesso aberto aos dados do governo, em especial, dados epidemiológicos, influenciando, inclusive, diretamente a política de saúde de seus países.

Fonte: *The Research Information Network (RIN)* e *The National Endowment for Science, Technology and the Arts (NESTA)* (2010).

O projeto FOSTER, acrônimo da iniciativa *Facilitate Open Science Training for European Research*, que durou trinta meses, iniciada em fevereiro de 2014 e possibilitou o apoio em diversos segmentos, especialmente o financiamento aos jovens pesquisadores na adoção do acesso aberto na Europa, e principalmente, promovendo a integração de princípios e práticas de acesso aberto no atual fluxo de trabalho de pesquisa. Para a iniciativa a Ciência Aberta constitui de uma

[...] prática da Ciência de tal forma que outros possam colaborar e contribuir, onde dados de pesquisa, notas de laboratório e outros processos de pesquisa estão disponíveis gratuitamente, em termos

que permitam a reutilização, redistribuição e reprodução da pesquisa e seus dados e métodos subjacentes. (FOSTER, 2018a).

Atualmente, a iniciativa está em fase de implementação de sua atualização em conjunto ao projeto *Fostering the practical implementation of Open Science in Horizon 2020 and beyond (FOSTER Plus)*. O projeto *FOSTER Plus* possui dois anos financiados pela União Europeia, realizado por 11 parceiros em 6 países, cujo principal objetivo é “[...] contribuir para uma mudança real e duradoura no comportamento dos pesquisadores europeus para garantir que a *Open Science (OS)* se torne a norma.” (FOSTER, 2018b).

Nesse escopo, os investimentos para a concretização da Ciência Aberta se justificam pela garantia de sua própria sustentabilidade (CURTY; AVENTURIER, 2017). Para Wilkinson et al. (2016), ainda “[...] há uma necessidade urgente de melhorar a infraestrutura que suporta a reutilização de dados acadêmicos”. Além do mais, gerenciar os dados não é a meta, mas o “[...] canal principal que leva à descoberta e à inovação do conhecimento e à subsequente integração e reutilização desses dados e conhecimento pela comunidade após o processo de publicação de dados.” (WILKINSON et al., 2016).

Todavia, ressalta-se que a verdadeira adoção da Ciência Aberta deve ir além das questões tecnológicas, mas encontrar, acessar e manipular os dados, que são recursos valiosos, intensificando a sua disponibilidade e reformulando o *modus operandi* da Ciência, por meio de políticas que ampliem a divulgação e ainda que consigam definir melhores práticas para que os pesquisadores .

A sustentabilidade e implementação da ideia dependerá da “[...] busca por formas eficientes de maximizar o reuso de dados científicos, ao invés de meramente estocá-los como volumes ociosos em repositórios.” (CURTY; AVENTURIER, 2017).

A Ciência Aberta é um conceito relativamente novo e complexo e sua adoção exigirá uma mudança no comportamento dos pesquisadores em relação à realização de pesquisas e compartilhamento de informações e exigirá a adoção de novas práticas. (PONTIKA et al., 2015, p.1).

Wilkinson et al. (2016) expõem a existência de duas partes interessadas neste modelo: o humano e o computacional e que essa segunda parte “[...] são cada vez mais relevantes e exigem tanto, ou mais, atenção à medida que sua importância cresce”. Sua motivação é o estabelecimento de princípios que objetivam orientar

produtores de dados e editores em relação ao gerenciamento de dados a longo prazo.

A Ciência, portanto, é melhorar a descoberta de conhecimento através da assistência a seres humanos e seus agentes computacionais, na descoberta, acesso e integração e análise de tarefas apropriadas dados científicos e outros objetos digitais acadêmicos. (WILKINSON et al., 2016).

Para o encaminhamento destas tarefas científicas, a gestão de dados também vem se consolidando para a melhor viabilização da Comunicação Científica e o compartilhamento de dados. Nesse viés, os princípios FAIR, ou ainda, *Findable, Accessible, Interoperable and Reusable*, são diretrizes essenciais para o gerenciamento dos dados de pesquisa.

No trabalho *The FAIR Guiding Principles for scientific data management and stewardship* em que aparece como autor principal, Wilkinson et al. (2016) apoia a descoberta por meio de um bom gerenciamento de dados. No artigo, ele descreve as quatro diretrizes que constituem de que o dado:

Para ser ENCONTRADO: os (meta) dados são atribuídos a um identificador globalmente exclusivo e persistente; os dados são descritos com metadados precisos e relevantes; os metadados, de forma clara e explícita, incluem o identificador dos dados que descreve; os (meta) dados são registrados ou indexados em um recurso pesquisável. Para ser ACESSÍVEL: os (meta) dados são recuperáveis pelo seu identificador usando um protocolo de comunicação padronizado; o protocolo é aberto, gratuito e universalmente implementável; o protocolo permite um procedimento de autenticação e autorização, quando necessário; os metadados estão acessíveis, mesmo quando os dados não estão mais disponíveis; Para ser INTEROPERÁVEL: os (meta) dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento; os (meta) dados usam vocabulários que seguem os princípios do FAIR; os (meta) dados incluem referências qualificadas a outros (meta) dados. Para ser REUTILIZÁVEL: os meta (dados) são ricamente descritos com uma pluralidade de atributos precisos e relevantes; os (meta) dados são liberados com uma licença de uso de dados clara e acessível; os (meta) dados estão associados à proveniência detalhada; os (meta) dados atendem aos padrões da comunidade relevantes ao domínio. (WILKINSON et al. 2016, p. 4)

Nota-se, pela observação dos princípios FAIR, há uma nítida relação com as temáticas abordadas pela Ciência da Informação muito antes da proposição de definição da nova Ciência de Wilkinson et al. (2016).

Bush (1945) ao escrever que “[...] um registro, para ser útil à Ciência, deve ser continuamente estendido, armazenado e, acima de tudo, deve ser consultado”, remetendo-nos aos princípios elencados na FAIR. A Ciência da Informação, então, tem como escopo a investigação das propriedades e comportamentos da informação, podendo auxiliar no processo de

[...] transformar a imensa massa de dados operacionais disponíveis diariamente em informações consistentes que permitam a tomada de decisões e agreguem valor às atividades e aos negócios. (SANTOS; VIDOTTI, 2009, p. 4).

Araújo (2017) realizou um amplo mapeamento histórico onde identifica três grandes momentos na história da área de Ciência da Informação. O primeiro momento, o Momento Fundacional, agrupado em cinco dimensões: 1) o surgimento da bibliografia; 2) a relação entre a documentação e parte da biblioteconomia; 3) a atuação dos primeiros cientistas da informação; 4) desenvolvimento tecnológico e sua teorização; e 5) a fundamentação na teoria matemática.

Juntos, esses fatos conduziram à consolidação de uma primeira Ciência da Informação, que se manifestou no contexto anglo-saxão e soviético entre as décadas de 1940 e 1960, e daí se espalhou para diversas outras regiões do planeta. (ARAÚJO, 2017).

O segundo momento, a ampliação das problemáticas, constituiu-se para Araújo (2017) em seis subáreas: 1) o estudo dos fluxos informacionais da informação científica, 2) a representação da informação, 3) o estudo dos usuários da informação, 4) a gestão da informação e do conhecimento, 5) a economia política da informação e, por fim, 6) o estudo métrico da informação. (ARAÚJO, 2017).

O terceiro momento, surge para Araújo (2017) como consequência da evolução das subáreas e com o conceito de informação sendo objeto de teorizações e reflexões, para ele, há três grandes conceitos ou modelos de estudo da informação: o primeiro onde a informação é observada como um fenômeno físico, um sinal, um objetivo; o segundo, sendo entendido como fenômeno cognitivo, semântico, subjetivo; e o terceiro, entendido não como objetivo ou subjetivo, mas sim, pragmático, intersubjetivo, sociocultural. Um fenômeno coletivo, de uma construção social.

Neste terceiro momento, Araújo (2017) identifica treze perspectivas contemporâneas de estudos: 1) análise de domínio, 2) altmetria, 3) cultura organizacional, 4) curadoria digital, 5) *folksonomias*, 6) ética intercultural da informação, 7) neodocumentação, 8) humanidades digitais, 9) arqueologia da sociedade da informação, 10) estudo de usuários, 11) regimes de informações, 12) memória e 13) diálogo entre biblioteconomia, arquivologia e museologia.

Sant'Ana (2014) destaca a importância da Ciência da Informação conhecer e contribuir em todas as fases e fatores do processo de acesso a dados, construindo novas teorias e definindo caminhos para que seu fluxo contribua para o atendimento das necessidades informacionais, tornando-se fundamental, conhecer e contribuir em todas as fases e fatores do processo de acesso a dados.

A Ciência da Informação faz importantes contribuições para a Comunicação Científica fornecendo fontes de informações para auxiliar na obtenção de respostas a questionamentos dos cientistas; concorrendo para a atualização do cientista em seu campo de atuação ou estimulando a descoberta e a compreensão de novos campos, além de divulgar novas tendências de áreas emergentes por meio de serviços de disseminação seletiva de informação e eventos; apresentando relatórios bibliométricos, cientométricos ou altimétricos atestando a relevância do trabalho do cientista, fornecendo feedback para o aperfeiçoamento da produção do pesquisador.

A Comunicação Científica, em si, representa uma área do conhecimento de complexidade significativa no que diz respeito aos caminhos teóricos, epistemológicos e práticos, estabelecendo-se como uma disciplina de interesse da Ciência da Informação. (FERREIRA; MARCHIORI; CRISTOFOLI, 2010, p.81).

Por esse prisma, há a necessidade do engajamento das bibliotecas, museus e arquivos com o desafio de gerenciar o acesso, coleta, armazenamento e disponibilização de conjuntos de dados de pesquisa em seus sistemas de informações, pois, como defendem Sayão e Sales (2015, p. 6), “[...] a biblioteca de pesquisa é a custodiante natural dos registros acadêmicos, e este papel se estende agora para incluir os dados de pesquisa”.

Nota-se, portanto, que a Ciência da Informação abarca as questões relacionadas ao fluxo de dados, a gestão dos processos informacionais e a obtenção de conhecimentos, humanos e computacionais e precedeu os atuais

movimentos denominados Ciência Aberta, humanidades digitais, *Big Data*, data science ou e-Science.

No início de 2018, a NASA, a agência aeroespacial norte americana, anunciou a descoberta de um sistema com ao menos cinco exoplanetas por “cientistas cidadãos”, em uma publicação de Greicius (2018), no que foi considerado o primeiro sistema multiplanetário descoberto inteiramente por crowdsourcing.

A descoberta dos exoplanetas se deu pouco mais de vinte anos após um grupo de cientistas da computação e astrônomos proporem uma solução inédita para o grande volume de dados captados do espaço por grandes radiotelescópios em diversas partes do mundo. Howe (2009), descreve como o público foi convocado para doar seu tempo e recursos para a tarefa de analisar esses dados. Os voluntários faziam o download de um simples protetor de tela, baixaria um bloco de dados, fazia a varredura dos dados contidos nesse bloco e em seguida retornaria o resultado das análises ao servidor central. O projeto recebeu o nome de *SETI@home* e pretendia contar com a ajuda de 100 mil pessoas. Em 2005, 5,2 milhões de pessoas baixaram o protetor de tela, registrando quase 3 milhões de anos de tempo de computação. Entrou para o livro dos recordes, embora até onde sabemos, não encontrou vida extraterrestre.

Ainda em sua infância, para Arolas e González-Ladrón-De-Guevara (2012), o termo *crowdsourcing* vem passando por revoluções constantes à medida que novas aplicações surgem. “O termo *crowdsourcing*, é essencialmente associado à colaboração em massa, à produção coletiva, à processos cooperativos, colaborativos, voluntários, gratuitos e de autoria não identificada.” (ZANONA; PETEROSI; BRANDÃO, 2013, p. 46) .

O termo *crowdsourcing* foi cunhado por Jeff Howe em 2006 e tem sua origem na junção das palavras *crowd* (multidão) e *outsourcing* (terceirização). Inicialmente é definido por Howe como “[...] o ato de oferecer um trabalho que geralmente era executado por uma pessoa, [...] e oferecê-lo em uma chamada aberta para um grupo indefinido de pessoas”, sendo “[...] um modelo de solução de problemas legítimo e complexo [...]” (BRABHAM, 2008, p. 76).

A Ciência tem se beneficiado de dados produzidos não apenas por cientistas, mas também pelo público geral, por exemplo, Dugas et al. (2013), desenvolveram um modelo prático de previsão da propagação da gripe influenza baseado em dados em tempo gerados pela aplicação “*Google Flu Trends*” que, apesar de não estar

mais ativo, utilizava como base as buscas realizadas por pessoas comuns sobre a gripe influenza realizadas em sua plataforma de pesquisa.

Portanto, uma vez que participam elementos indefinidos, o modelo cria estruturas de informações onde “[...] a qualidade dos dados é uma grande preocupação.” (ANTON *et al.*, 2018).

Esta preocupação caminha em todos os sentidos. Ao mesmo tempo em que conjuntos de dados são disponibilizados para que cientistas amadores realizem análises, esses também produzem e disponibilizam seus próprios conjuntos de dados, trazendo enormes desafios relacionados à procedência, à qualidade e a confiança nos produtores e nos conjuntos de dados, aumentando ainda mais os desafios relacionados à validação de uma pesquisa científica, como exemplo a reprodutibilidade dos dados de pesquisa.

Nosek e Errington (2017) concordam com a afirmação de Plesser, entretanto, aprofundam o questionamento sobre o que significa, de fato, repetir uma metodologia e o que qualifica os “mesmos” resultados? Para eles, “Repetir a metodologia não significa uma replicação exata, mas sim a repetição do que se supõe importar para obter o resultado original.” (NOSEK; ERRINGTON, 2017).

Destaca-se também a existência de grande variedade de áreas do conhecimento com suas particularidades no que tange seus dados científicos, por essa razão, “[...] o padrão de compartilhamento varia enormemente no mundo heterogêneo e complexo da Ciência” (SAYÃO; SALES, 2017, p. 68) e, portanto, conhecer de perto as estratégias para gestão do ciclo de vida dos dados praticados pelos pesquisadores é essencial para o avanço da Ciência, do conhecimento e da atuação profissional na Ciência da Informação.

Ainda nesta década, inúmeros estudos de replicação vêm sendo realizados, obtendo resultados que colocam várias práticas científicas em questão.

Uma quantidade não desprezível de descobertas científicas acaba não sendo confirmada após sua publicação, por problemas que incluem erros e fraudes, mas que também se estendem a falsos resultados positivos ou negativos obtidos de boa-fé. O problema assombra pesquisadores e revistas científicas, obrigados a cancelar a publicação de trabalhos cujos resultados soavam promissores, e tornou-se um pesadelo para empresas farmacêuticas e de biotecnologia. (MARQUES, 2014).

Patil, Peng e Leek (2016) enumeram tentativas de reprodutibilidade cujos dados e código do estudo original foram disponibilizados, mas estavam incompletos e ou incorretos.

Num esforço colaborativo com aproximadamente 250 pesquisadores vinculados a dezenas de instituições, se reuniram no projeto *Open Science Collaboration* (2015) para conduzirem replicações de 100 estudos experimentais e correlacionais publicados em três periódicos da área de psicologia em 2008, sugerindo que apenas 39 dos estudos publicados poderiam ser reproduzidas e, mais gravemente, dos 61 não replicados, apenas 24 capazes de produzir resultados semelhantes aos originais, concluindo que:

Uma grande parcela de repetições produziu evidência mais fraca para as conclusões iniciais, apesar de utilizar materiais fornecidos pelos autores originais, revisão de antecedência para a fidelidade metodológica e alto poder estatístico para detectar os tamanhos dos efeitos originais. (OPEN SCIENCE COLLABORATION, 2015).

Em outro estudo, com resultados parciais publicados em 2017, o Projeto de Reprodutibilidade em Biologia do Câncer apresentou os resultados de cinco estudos de replicação e “[...] sugerem que há uma oportunidade substancial de melhorar a reprodutibilidade na Biologia do Câncer [...]” (NOSEK; ERRINGTON, 2017).

À medida que o movimento para examinar e aprimorar a confiabilidade da pesquisa se expande, é importante notar que alguns de seus termos básicos - reprodutibilidade, replicabilidade, confiabilidade, robustez e generalização - não são padronizados. (GOODMAN; FANELLI; IOANNIDIS, 2016).

A reprodutibilidade da pesquisa vem sendo questionada em muitos campos. (GIBB, 2014; HEWITT, 2016; CAMERER et al., 2018). "No momento atual não há consenso sobre o que a reprodutibilidade é ou deveria ser", ou ainda, o que pode ser feito para maximizar a reprodutibilidade? (CASAVADEL *apud* BAKER, 2016; GIBB, 2014). Gibb (2014) que percebe que os resultados da pesquisa “[...] capturam um retrato confuso das atitudes em torno dessas questões”.

No documento intitulado *Replication Studies: improving reproducibility in the empirical sciences*, a *Royal Netherlands Academy of Arts and Sciences* enumera vinte causas de não reprodutibilidade (ROYAL NETHERLANDS ACADEMY OF ARTS AND SCIENCES, 2018, p. 22), que são:

1. Desenho experimental ineficiente associado a controle de vieses falho; 2. Amostras de tamanho insuficiente; 3. Problemas em testes estatísticos que geram falsos resultados negativos; 4. Erro técnico ou humano na execução do estudo, associado a controle de qualidade ineficaz; 5. Fraude ou fabricação de dados; 6. Falta de rigor na análise estatística; 7. Análise estatística equivocada; 8. Falta de conhecimento sobre variáveis que influenciam o resultado; 9. Falhas do pesquisador em reproduzir os resultados antes da publicação; 10. Omissão de resultados nulos ou análise seletiva que faz os nulos parecerem positivos; 11. Não compartilhamento de dados ou de detalhes metodológicos; 12. Escolha de variáveis que se adequam aos resultados; 13. Formulação de hipótese depois que os resultados são conhecidos; 14. Discrepância entre os resultados registrados e os publicados; 15. Ausência de revisão por pares adequada; 16. Ênfase no incentivo a artigos de alto impacto; 17. Recompensas exageradas a resultados de pesquisa tidos como disruptivos; 18. Sistema de financiamento à pesquisa demasiadamente competitivo; 19. Falta de recompensa para práticas que favoreçam a replicação de estudos; 20. Crença de que o rigor no processo de pesquisa dificulta novas descobertas. (ROYAL NETHERLANDS ACADEMY OF ARTS AND SCIENCES, 2018, p. 22).

Nesse parâmetro, a gestão de dados científicos é de extrema importância para que os dados sejam disponíveis à comunidade com o mínimo de tratamento informacional. Para isso, o próximo capítulo conduzirá para uma reflexão sobre os procedimentos que envolvem a gestão dos dados, curadoria digital e os metadados.

3 GESTÃO DOS DADOS E METADADOS DE PESQUISA

A construção da Biblioteca de Alexandria foi envolta de interesses de transformar uma localização geograficamente marginalizada em uma “[...] capital com uma memória e raízes artificiais; compensar sua marginalidade geográfica por uma centralidade simbólica” (MEY, 2004), por meio da concentração de toda a memória do mundo em uma nova cidade, trazendo inúmeros ganhos políticos e, conseqüentemente, poder.

Mey (2004) ressalta que Demétrio de Faleros convenceu Ptolomeu I Sóter que este seria um governante mais estável e melhor se conhecesse os povos por ele governados acumulando suas obras, podendo erigir um reino mais duradouro.

As obras se adquiriam de formas variadas. A mais interessante constituía-se na cópia de todos os livros encontrados nos navios que aportavam em Alexandria: revistava-se o navio, os livros eram levados à Biblioteca, copiados, e então devolvidos para que o navio partisse. [...] O único fato marcante, neste primeiro “empréstimo entre bibliotecas”: nem sempre os originais eram devolvidos, mas as cópias... (MEY, 2004, p. 75).

Concentrar toda a memória do mundo, organizá-la ou melhor, curá-la para acessá-la é um ideal que retorna sistematicamente como um dos objetivos da humanidade. Não foram poucas as iniciativas que objetivaram prover acesso a todo o conhecimento humano.

Com a criação do Instituto Internacional de Bibliografia por Paul Otlet, juntamente de seu sócio Henry de La Fontaine, em 1895 e, posteriormente, a construção do *Palais Mondial* - novamente, uma cidade do conhecimento, a exemplo de Alexandria – que em seguida seria renomeada Mundaneum e que fora construída para abrigar diversas atividades, objetivava atingir tais objetivos, sob os princípios de “[...] organização, totalidade, simultaneidade, gratuidade, voluntariedade, universalidade e mundialidade” (RUSSO, 2010, p. 171), porém, as Guerras Mundiais, o fim do apoio governamental e empresarial bem como o rápido avanço científico e tecnológico, ofuscaram as importantes contribuições de Otlet.

As ideias americanas relacionadas à informação passaram a dominar o cenário pós-guerra. Em artigo publicado na revista *The Atlantic*, Bush, em 1945, criticava as limitações dos sistemas de organização e representação de informações, alertando para a necessidade de criar métodos que possibilitaram analisar, organizar, representar e recuperar informações, frente ao grande volume

produzido. Neste artigo, Bush (1945) propõe o uso do MEMEX para auxiliar o ser humano no uso da informação e vislumbra o uso das tecnologias eletromecânicas como suporte de registros do conhecimento.

Um registro, para ser útil à Ciência, deve ser continuamente estendido, armazenado, e acima de tudo, deve ser consultado. Hoje registramos convencionalmente pela escrita e fotografia, seguido de impressão, mas também gravaremos em filme, em discos de cera e em fios magnéticos. (BUSH, 1945).

De fato, tais dispositivos tornaram-se realidade. Os cartões perfurados, as tecnologias de armazenamento evoluíram para dispositivos magnéticos, ópticos e, atualmente, as tecnologias de memória sólida (SSD) são o padrão da indústria.

Os surgimentos dessas tecnologias apenas ampliaram os impactos do antigo problema da explosão informacional, antes representada pela explosão bibliográfica. Engana-se aquele que acredita que a explosão informacional e a abundância de dados (a característica mais marcante da *Big Data*) são problemas recentes e causados pelas tecnologias de informação e comunicação. Da mesma forma, ressalta-se que não é primeira vez que mudanças no âmbito dos suportes informacionais e da ampliação do volume de informação aparecem como problemas.

Como esses desafios, os bibliotecários já superaram por inúmeras vezes as dificuldades de organizar informações registradas em tábuas de argila, rolos de papiros e grandes volumes de papel, como mapas e partituras, antes de se concentrarem no tratamento de livros no formato que os conhecemos atualmente. Os instrumentos clássicos da biblioteconomia como os sistemas de classificação, indexação, catalogação são tecnologias que resolvem grande parte dos problemas relacionados à explosão informacional.

Cunha (2000) afirma que em todas as épocas “[...] as bibliotecas sempre acompanharam e venceram os novos paradigmas tecnológicos”. Como exemplo, o desenvolvimento e o uso de meios óticos, magnéticos, eletrônicos e mecânicos como o fonógrafo, os cartões perfurados, o disco de vinil e a fita magnética deram início a uma nova forma de armazenamento de informações na qual dispositivos eletromecânicos são necessários para o acesso ao conteúdo nestes suportes. Até o surgimento dessas tecnologias de armazenamento, bastavam os sentidos naturais para acessar o conteúdo armazenado nos mais diversos suportes. Mais

uma vez, o desafio de manter a organização para o acesso a esse conteúdo foi superado com a adesão dessas e de outras tecnologias pelas bibliotecas.

Essas e as atuais tecnologias acrescentaram outras dimensões e problemas relacionados ao conteúdo informacional. À medida que as tecnologias de armazenamento surgem e seu custo de produção e manutenção caem em relação às tecnologias precedentes, a indisponibilidade de dispositivos para o acesso aos registros antigos aumenta, gerando novos problemas para acesso.

As respostas a essas ameaças de continuidade digital que são baseadas em abordagens de preservação tradicionais não funcionam. Simplesmente capturar dados em mídias de armazenamento estáveis e depois copiá-las em novas mídias quando as anteriores se tornarem obsoletas é uma ameaça. Dados digitais devem ser geridos desde sua criação se a pretensão for de sobrevivência. (HARVEY, 2010, p. 10).

Arellano e Ángel (2008) afirmam que para que um arquivo digital seja recuperado, não basta apenas salvar o programa (algoritmo), mas preservar também o hardware em que foi gerado, buscando garantir sua autenticidade e sua acessibilidade.

É muito comum conferir às tecnologias de informações atributos que nem sempre são condizentes com a realidade. Diz-se muito da capacidade de tais tecnologias em gerenciar o conhecimento quando, em geral, são tecnologias extensoras de nossas atividades de coletar, processar, memorizar e transmitir dados, estando ainda, portanto, longe da capacidade humana de criar, intuir, sentir e raciocinar. Como Svenonius (2008, p. 1) afirma

Acesso eletrônico imediato à informação digital é o único atributo mais marcante da era da informação. O mecanismo de recuperação elaborado que suporta tal acesso é um produto de tecnologia. Mas a tecnologia não é suficiente. A eficácia de um sistema de acesso à informação é uma função direta da inteligência posta em organizá-lo.

Até mesmo os sistemas baseados em inteligência artificial (I.A.) “[...] não podem gerar insights sem treinamento.” (WEDEMANN, 2018, *on-line*). A inteligência posta em organizá-la, portanto, é um atributo do ser humano e o instrumento para aplicá-la pode ser encontrada nas atividades relacionadas à curadoria digital.

O acesso à informação de amanhã depende de ações cautelares tomadas hoje, pois um fato fundamental da sustentabilidade digital

é que, sem preservação, não há acesso. Neste cenário, vemos surgir um novo papel e campo de trabalho para os profissionais que lidam com informação: a curadoria digital. [...] A preservação segura vai garantir a conservação da informação ao longo do tempo, mas para isso devem-se usar as melhores práticas em todo o ciclo de vida dos dados: criação, descrição e curadoria, além de seu armazenamento seguro, que garanta seu uso e reuso. (SALES; CAVALCANTI, 2015, p. 92).

A curadoria digital segundo Abbott (2008, *on-line*) é “[...] a gestão e preservação de dados digitais a longo prazo”, garantindo a sustentabilidade dos dados a longo prazo.

Todas as atividades envolvidas no gerenciamento de dados desde o planejamento de sua criação, melhores práticas na digitalização e documentação e a garantia de sua disponibilidade e adequação para descoberta e reutilização no futuro fazem parte da curadoria digital. A curadoria digital também pode incluir o gerenciamento de vastos conjuntos de dados para uso diário, por exemplo, garantindo que eles possam ser pesquisados e continuar legíveis. Curadoria digital é, portanto, aplicável a uma ampla gama de situações profissionais desde o início do ciclo de vida da informação até o fim; digitalizadores, criadores de metadados, financiadores, criadores de políticas e gerentes de repositórios, para citar alguns exemplos. (ABBOTT, 2008, *on-line*).

A ideia de “curadoria” remete diretamente ao curador, àquele que tradicionalmente atua em museus, galerias de artes ou instituições de preservação, arte e cultura. Segundo o dicionário Michaelis (2018), curadoria” refere-se ao “Ato ou efeito de curar”. “Curar”, por sua vez, além de restabelecimento da saúde, significa também “[...] preparar algo, para facilitar o seu uso posterior” e “[...] ter por objetivo; cuidar, ocupar-se de, tratar”. (DIGITAL CURATION CENTRE, 2018a)

Nesse viés, é possível considerar as atividades de museólogos, bibliotecários e arquivistas dentro do rol de atividades de curadoria, uma vez que “[...] desde sua origem na Antiguidade Clássica, a biblioteca é um espaço de preservação dos conhecimentos gerados pela humanidade a partir de diferentes sociedades.” (ARAÚJO; OLIVEIRA, 2005, p. 24).

Mesmo que o escopo da curadoria digital seja a preservação de longo prazo, sendo um objetivo que muito se aproxima da preservação, criadores de dados e usuários podem obter benefícios imediatos de sua aplicação, pois mais do que obter acesso, seu propósito é garantir acesso persistente a dados digitais confiáveis.

A confiabilidade é garantida quando os dados se configuram como registro formal, por meio de verificações de autenticidade e uso de padrões abertos e livres

de licenças, que levam a maiores oportunidades de pesquisas cruzadas e colaboração. Com isso, é possível “[...] explorar o investimento inicial, garantindo que os dados estarão disponíveis para uso e reutilização e proteger o valor financeiro da informação.” (ABBOTT, 2008).

O reuso dos dados contribui para minimizar a duplicação de trabalho de coleta, otimizando custos e recursos; possibilitando a preservação de longo prazo mantendo a integridade dos dados; fornece salvaguardas contra má conduta científica, incluindo fraudes e ferramentas de treinamentos para novas gerações de pesquisadores. (OUCHI; SIMIONATO, 2018, p. 128).

A possibilidade de reutilização dos dados de pesquisa é, por isso, um dos fatores mais lembrados para que a curadoria seja implementada em qualquer processo de pesquisa, uma vez que, estando disponíveis e garantidos, os dados podem ser pesquisados, avaliados e recombinaados, criando novas oportunidades de colaboração e trazendo inúmeras vantagens como a redução de custo e tempo de coleta dos dados. (ABBOTT, 2008; SAYÃO; SALES, 2014; WILKINSON et al., 2016; CURTY; AVENTURIER, 2017).

3.1 Dados de pesquisa

Os dados de pesquisa são cruciais, pois provêm evidências para o conhecimento científico publicado, base de evidências para os resultados de pesquisas e que são a fundação para todo o progresso científico. (MOLLOY, 2011; INGRAM, 2016). São, portanto, essenciais, não apenas para o desenvolvimento, mas também para estabelecer a validade, replicação e reprodução da pesquisa, neste sentido, tornando-se um valioso produto da pesquisa.

Santos e Santana (2013, p. 205) definem o termo dado como “[...] uma unidade de conteúdo necessariamente relacionada a determinado contexto e composta pela tríade entidade, atributo e valor [...]”, qual seja, utilizaremos o termo dado como o “[...] elemento básico nos fluxos informacionais [...]” (SANTOS; SANT'ANA, 2013, p. 201).

Nos termos da *Organisation for Economic Cooperation and Development* (OECD), o termo “dado de pesquisa” é definido como:

[...] registros factuais (pontuações numéricas, registros textuais, imagens e sons) usados como fontes primárias para pesquisa científica, e que são comumente aceitos na comunidade científica como necessários para validar resultados da pesquisa. Um conjunto de dados de pesquisa constitui uma representação parcial e sistemática do assunto investigado. (OECD, 2007, p. 13).

Ingram (2016) define dados de pesquisa como “[...] todas as informações que foram coletadas, observadas, geradas ou criadas para validar os resultados da pesquisa original [...]”, incluindo “[...] formatos não digitais, como cadernos de laboratório e cadernos de esboços.” (INGRAM, 2016).

“Os dados de pesquisa podem ser caracterizados de várias formas.” (SAYÃO; SALES, 2015, p. 7). A título de exemplo, Sayão e Sales (2015) classificam os dados de pesquisa de acordo com sua natureza, origem ou de acordo com seu status no fluxo de trabalho da pesquisa. Segundo sua origem, os dados podem ser classificados como dados observacionais, quando “[...] obtidos por meio de observações diretas” (SAYÃO; SALES, 2015, p. 7), dados computacionais, quando “[...] resultados de execução de modelos computacionais ou de simulações” (SAYÃO; SALES, 2015, p. 8) e dados experimentais, quando “[...] provenientes de situações controladas em bancadas de laboratório” (SAYÃO; SALES, 2015, p. 8). Segundo a sua natureza, os dados seriam classificados como “[...] números, imagens, vídeos ou áudio, *software*, algoritmos, equações, animações ou modelos e simulações.” (SAYÃO; SALES, 2015, p.8).

Segundo Green, MacDonald e Rice (2009), conforme a fase da pesquisa, os dados de pesquisa podem ser classificados como Dados Brutos, Crus ou Preliminares (*RAW DATA* em inglês), que são aqueles que são produzidos “[...] diretamente dos instrumentos científicos” (SAYÃO; SALES, 2015, p.8), Dados Derivados, que são “[...] resultado do processamento ou combinação de dados brutos ou de outros dados” (SAYÃO; SALES, 2015, p. 8) e Dados Canônicos ou Dados Referenciais, aquelas “[...] coleções de dados consolidados e arquivados geralmente em grandes centros de dados, por exemplo, sequência genética, estrutura química, etc. [...] “. (SAYÃO; SALES, 2015, p. 9).

Considerando que os “[...] dados de pesquisa são os dados produzidos ou utilizados para o desenvolvimento de uma pesquisa.” (COSTA, 2017, p. 15), Fernandes e Ribeiro (2011) concluem que “[...] os dados produzidos em contexto de investigação são reconhecidamente de grande valor.”. Complementa-se ainda

pela citação de Simmons, Nelson e Simonsohn (2011, p. 1359) “Nosso trabalho como cientistas é descobrir verdades sobre o mundo. Nós geramos hipóteses, coletamos dados e examinamos se os dados são ou não consistentes com essas hipóteses.” (SIMMONS; NELSON; SIMONSOHN, 2011, p. 1359).

O custo acessível, aumento da capacidade e velocidade com que dados são coletados, registrados e acessados pelas tecnologias de informação e comunicação, tornaram possíveis a popularização e utilização de tecnologias de digitalização, de equipamentos automatizados de coleta, virtualização, dispositivos embarcados e abriram amplas possibilidades, antes de uso extremamente custoso e trabalhoso e que demandava muitos recursos humanos, financeiros e de tempo para sua manutenção e utilização e que agora, permitem uma miríade de opções ao pesquisador moderno, “[...] feitas no âmbito de projetos financiados e cuja reprodução requer novos custos em equipamentos e mão-de-obra.” (FERNANDES; RIBEIRO, 2011).

As decisões a serem tomadas por um pesquisador no decorrer da coleta e da análise de dados de suas pesquisas são muitas. Escolher quais e que tipos de dados devem ser coletados, em que quantidade e frequência, quais variáveis de controle devem ser consideradas, quais condições devem ser combinadas ou comparadas, se devo excluir e qual parte delas, são raras e “[...] às vezes impraticável, que os pesquisadores tomem todas essas decisões de antemão.” (SIMMONS; NELSON; SIMONSOHN, 2011, p. 1359).

Em vez disso, é comum (e prática aceita) que os pesquisadores explorem várias alternativas analíticas, busquem uma combinação que produza “significância estatística” e, então, relatem apenas o que “funcionou”. (SIMMONS; NELSON; SIMONSOHN, 2011, p. 1359).

Neste sentido, o tratamento dos Dados de Pesquisa depende, segundo Ingram (2016), do tipo de dados envolvidos, como são criados ou coletados e como devem ser usados no presente e no futuro.

O entendimento da necessidade de gerir o acesso e a utilização dos dados produzidos ou recolhidos no quadro das atividades de investigação, garantindo a sua preservação, parece ter sido mais precoce e mais profundo em áreas como a astronomia e a climatologia, onde o trabalho de investigação é baseado na análise de dados recolhidos de forma

distribuída, mas tem vindo a alargar-se a outros domínios. (RIBEIRO et al., 2010, p. 8).

No caso específico das informações científicas, “[...] a importância do registo e arquivamento dos dados produzidos durante a pesquisa é fundamental.” (ARELLANO; ÁNGEL, 2008, p. 23), pois “[...] quaisquer resultados ou dados de pesquisa podem ser usados para evidenciar descobertas publicadas ou podem ser combinados com outros dados para produzir novos tipos de registo de dados”. (INGRAM, 2016, *on-line*).

Destaca-se também a existência da variedade de áreas do conhecimento com suas particularidades no que tange seus dados científicos, por essa razão, “[...] o padrão de compartilhamento varia enormemente no mundo heterogêneo e complexo da Ciência” (SAYÃO; SALES, 2017, p. 68).

3.2 O ciclo de vida de dados

Conhecer de perto as estratégias para gestão dos dados praticados pelos pesquisadores é essencial para o avanço da Ciência, do conhecimento e da atuação profissional dos profissionais da Ciência da Informação.

Pesquisadores e formuladores de políticas consideram cada vez mais a gestão de dados como uma necessidade crítica em muitas áreas da Ciência, nas Ciências sociais e nas humanidades. (KANSA; KANSA; ARBUCKLE, 2014, p. 58).

Santana (2013) esclarece que conhecer e contribuir em todas as fases e fatores do processo de acesso a dados é, para isso, fundamental elaborar modelos que sirvam de compreensão sobre as fases, os relacionamentos, os fatores envolvidos, recursos disponíveis para tirar o melhor proveito e realizar melhorias nos aspectos relacionados ao acesso aos dados.

A Ciência da Informação pode e deve contribuir para que este cenário de acesso e uso intenso de dados se desenvolva da melhor maneira possível, buscando identificar e estudar fatores e características que propiciem ampliação do equilíbrio entre os atores envolvidos no processo e a máxima otimização do uso dos dados. (SANTANA, 2014, p. 119).

Sayão e Sales (2015) enumeram diversas concepções de modelos de ciclo de vida de dados de pesquisa, afirmando a importância desses modelos, já que

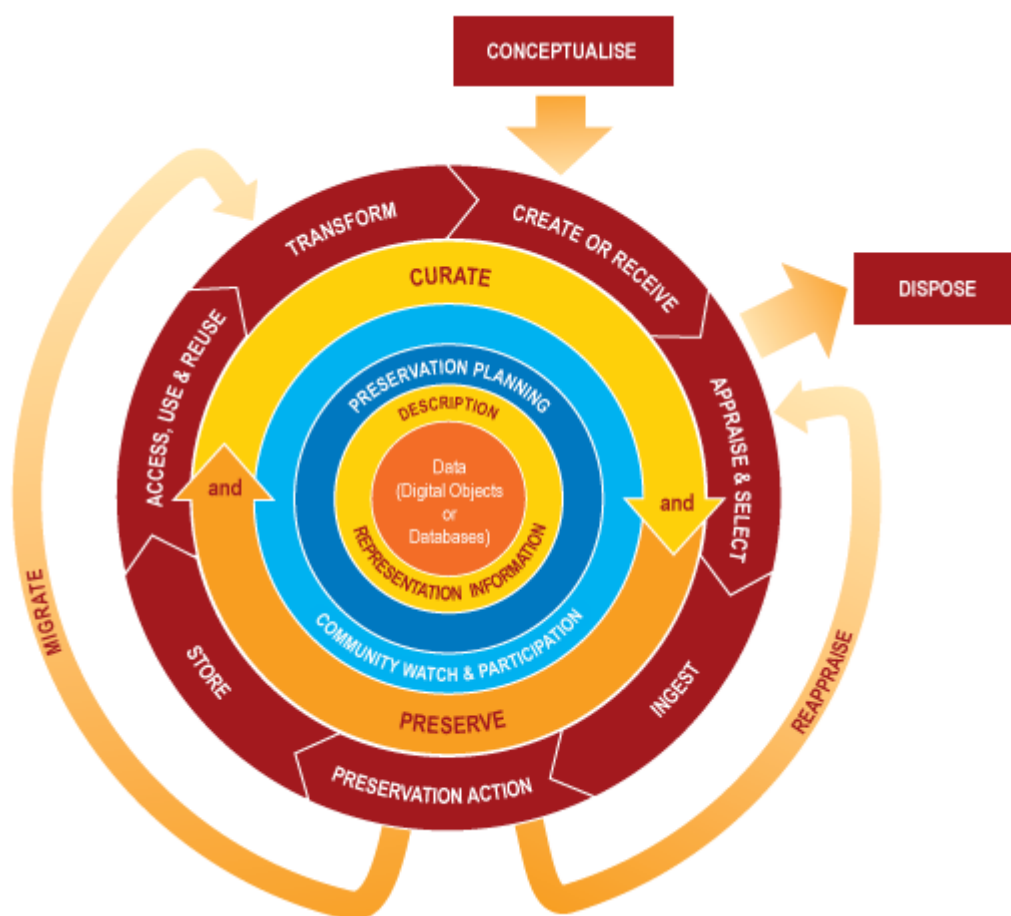
oferecem uma estrutura às diversas operações necessárias ao gerenciamento durante a vida dos recursos digitais. Dentre os modelos, Sayão e Sales (2015) citam o *Digital Curation Centre (DCC) Curation Lifecycle Model*, o *Dataone Data Lifecycle*, o *Data Documentation Initiative (DDI)*, o *Combined Lifecycle Model* e o *UK Data Archive Data Lifecycle*.

A gestão do ciclo de produção de dados científicos em uma pesquisa reduz as ameaças ao seu valor de pesquisa de longo prazo, reduzindo o risco de obsolescência digital. Harvey (2010, p. 10) afirma que “Dados digitais devem ser geridos desde sua criação se a pretensão for de sobrevivência”.

O ciclo de vida de curadoria dos dados digitais são processos contínuos que fornecem uma visão geral e de alto nível dos estágios envolvidos no gerenciamento e preservação de dados para uso e reutilização. Em geral, segundo o *Digital Curation Centre (2018)* os ciclos assumem as seguintes etapas: conceitual, criação, acesso e uso, avaliação e seleção, descarte, ingestão, preservação, revalorização, armazenamento, reutilização e transformação, que serão detalhados posteriormente.

Por se tratar de um ciclo, o Ciclo de Vida da Curadoria Digital, a exemplo do Modelo de Ciclo de Vida de Curadoria do DCC, pode ser representado como um fluxo, como representado na Figura 1.

Figura 1 - O modelo de ciclo de vida de curadoria do Digital Curation Centre



Fonte: Digital Curation Centre (2018a).

Em seu núcleo encontram-se os dados. O modelo considera como dado “[...] qualquer informação em formato digital binário [...]” (DIGITAL CURATION CENTRE, 2018a), incluindo objetos digitais simples como arquivos de texto, imagem som, juntamente com seus identificadores e metadados; objetos digitais complexos como websites; e bancos de dados, que são considerados coleções estruturadas de registros ou dados armazenados em um sistema de computador.

Em torno dos dados, o modelo prevê o que denomina como “ações completas do ciclo de vida” que inclui a descrição e representação da informação, atribuindo metadados administrativos, descritivos, técnicos, estruturais e de preservação, usando padrões apropriados, assegurando a descrição e controle adequado a longo prazo; o planejamento da preservação, incluindo planos para gerenciamento e administração de todas as ações de ciclo de vida; a vigilância e participação da comunidade, mantendo vigilância sobre as atividades comunitárias apropriadas e participação no desenvolvimento e compartilhamento de padrões, ferramentas e

software; e organizar e preservar, mantendo Ciência e realizando ações administrativas planejadas para promover a curadoria e a preservação.

As demais ações como as ações sequenciais e ocasionais já citadas anteriormente, são as etapas normalmente consideradas pela maioria dos modelos de curadoria de dados e são, segundo o *Digital Curation Centre (2018a, on-line)* :

- **Etapa conceitual:** concepção e planejamento da criação de objetos digitais, inclusive a captura dos dados e opções de armazenamento.
- **Criação:** produção de objetos digitais e atribuição de metadados arquivísticos administrativos, descritivos, estruturais e técnicos.
- **Acesso e uso:** assegurar-se de que os usuários designados possam acessar facilmente os objetos digitais no dia a dia. Alguns objetos digitais podem ser disponibilizados publicamente, enquanto outros podem ser protegidos.
- **Avaliação e Seleção:** avaliação de objetos digitais e seleção daqueles que precisam de curadoria e preservação de longo prazo. Adesão à documentação, políticas e requisitos legais.
- **Descarte:** eliminar conjunto de objetos digitais não selecionados para conservação de longo prazo. Documentação, políticas e requisitos legais podem exigir a destruição segura desses objetos.
- **Ingestão:** transferência de objetos digitais para um arquivo, repositório digital confiável, data center ou similar, aderindo, novamente, às orientações documentadas, políticas e requisitos legais.
- **Preservação:** realizar ações para assegurar a preservação e retenção de longo prazo dos objetos digitais.
- **Revalorização:** retornar objetos digitais que falham nos procedimentos de validação para nova avaliação e seleção adicional.
- **Armazenamento:** armazenamento dos objetos digitais de forma segura, conforme descrito em padrões de alta relevância.
- **Reutilização:** assegurar-se de que os dados sejam acessíveis aos usuários designados para uso e reutilização pela primeira vez. Alguns materiais podem estar disponíveis publicamente, enquanto outros dados podem ser protegidos.
- **Transformação:** criação de novos objetos digitais a partir do objeto original. (DIGITAL CURATION CENTRE, 2018a, *on-line*).

No entanto, é importante destacar que o modelo proposto pelo *Digital Curation Centre* parte de uma idealização do gerenciamento de dados, sendo envolvido em um contexto cultural presente na própria instituição. Como os autores Sayão e Sales (2015, p. 13) destacam:

Entretanto, alguns projetos podem usar apenas parte do ciclo de vida, por exemplo, um projeto envolvido com meta-análise pode se concentrar nas etapas descobrir, integrar e analisar e desconsiderar as outras etapas, ou seja, alguns projetos podem não seguir de forma linear o caminho delineado pelo modelo.

Assim, o modelo do *Digital Curation Centre* pode e deve ser adaptado às diversas realidades, principalmente às circunstâncias e desafios particulares apresentados em cada projeto de gestão de dados, portanto, o modelo deve ser usado em todo ou em partes, de acordo com a necessidade.

Sayão e Sales (2015) orientam aos indivíduos envolvidos, como pesquisadores, professores, estudantes e outros profissionais, que ao iniciar pela gestão do Ciclo de Vida dos Dados de Pesquisa (CVD) é necessário que eles saibam as possibilidades para que possam continuar a trabalhar com os dados de pesquisas anteriores, uma vez que os dados e as coleções de dados dessas pesquisas normalmente possuem tempo de vida maior que os projetos que as criaram. Os autores esclarecem que o que possibilita o futuro reuso desses recursos é também a existência de uma documentação exaustiva dos conjuntos de dados (*datasets*) na forma de metadados.

[...] os metadados são um subconjunto da documentação central padronizada e estruturada de dados que explica a origem, finalidade, referência de tempo, localização geográfica, criador, condições de acesso e termos de uso de uma coleta de dados. (EYNDEN et al., 2011).

Green, MacDonald e Rice (2009) acrescentam que, além da descrição, é especialmente importante que os processos usados para criar os dados também sejam documentados, trazendo como exemplos de documentação: arquivo de código de dados para dados estatísticos; código para um programa de *software*; uma especificação de formato; um relatório técnico explicando o protocolo ou a

metodologia da pesquisa e defendem que “[...] sem essa documentação, um conjunto de dados pode não ser adequado para reutilização.”

No contexto de dados científicos, Qin et al. (2010, p. 128) esclarecem que “O termo metadado científico é frequentemente usado para se referir aos dados que descrevem os conjuntos de dados coletados ou gerados a partir de pesquisas científicas”.

Para Sayão e Sales (2013), o rigor da aplicação e a qualidade e precisão dos esquemas de metadados são cruciais para garantir o acesso, o uso e o reuso interdisciplinar e atemporal dos dados de pesquisas.

Segundo Green, MacDonald e Rice (2009), a *Digital Library Foundation* (DLF), define três categorias de metadados: metadados descritivos, metadados administrativos e metadados estruturais. Metadados descritivos descrevem o conteúdo intelectual do objeto. Metadados administrativos são aqueles que os repositórios utilizam para o gerenciamento do objeto armazenado, podendo incluir informações de como foi criado, seu formato de armazenamento, políticas de compartilhamento e licenciamento além de informações necessárias para a preservação de longo prazo, os metadados de preservação. Metadados estruturais são aqueles que vinculam um objeto a outros para compor unidades lógicas.

Já Gilliland-Swetland (1999) apud Santos, Simionato e Arakaki (2014) apresentou os metadados em cinco níveis: administrativos, descritivos, de conservação, técnicos e de uso.

Os metadados explicitam os diferentes aspectos do recurso que descreve: sua estrutura, conteúdo, qualidade, contexto, origem, propriedade e condição. E auxiliam na organização, favorecem a interatividade, validam as identificações e asseguram a preservação e principalmente, otimizam o fluxo informacional melhorando o acesso aos dados e a localização dos recursos informacionais. (SANTOS; SIMIONATO; ARAKAKI, 2014, p. 150).

Ademais, o que seriam metadados de qualidade? Para Lawrence et al. (2011, p. 14), os fatores que determinam se os metadados são de qualidade são:

- Existência de metadados suficientes para descrever o formato e o conteúdo.
- Existência de metadados suficientes para descrever a procedência e o contexto de sua criação, mudanças ocorridas desde sua medição, fluxo

de processamento claro e registros das interações humanas com os dados antes da publicação e submissão dos dados.

- Existência de Metadados de referência atestando e afirmando a qualidade dos dados, caso já haja tais informações.
- Existência de metadados de descoberta, como por exemplo os do Dublin Core.
- Uso de vocabulários controlados apropriados.
- Referências internas e externas consistentes.

Embora haja grande consenso em relação à importância do papel dos metadados para subsidiar a organização, representação, descoberta, seleção e reuso dos conjuntos de dados, Tenopir et al. (2011) demonstram que há ainda uma lacuna entre o que é dito e o que é feito. Em pesquisa realizada sobre o compartilhamento de dados entre pesquisadores, concluíram que 56% dos pesquisadores entrevistados relataram que não usam nenhum padrão de metadados e cerca de 22% dos entrevistados indicaram usar seu próprio padrão. (TENOPIR et al., 2011)

Esses dados corroboram as percepções apontadas por Castro, Simionato e Zafalon (2016), as quais inúmeras áreas criam instrumentos próprios de descrição sem o conhecimento de métodos, processos e peculiaridades do tratamento informacional, gerando uma variedade de padrões que não atendem satisfatoriamente à complexidade de cada tipo de recurso.

4 BLOCKCHAIN

A partir do estudo apresentado no capítulo anterior, destaca-se uma outra forma de tratamento de dados, a *Blockchain* e suas demais tecnologias periféricas como os mecanismos de consenso e os contratos inteligentes (*smartcontracts*), que possuem potencial para alterar o processo de comunicação e divulgação científica.

A tecnologia *Blockchain* surge como um meio de implantar sistemas que fornecem funções de banco de dados descentralizadas protegidas de adulteração e revisão e que são capazes de operar em redes abertas de maneira completamente descentralizada. Isto é, a *Blockchain* é uma estrutura de dados distribuída (livro-razão ou *ledger*), que contém dados de transações, registrados e posteriormente armazenados como blocos ligados a outros blocos - como uma corrente - de forma que esses registros se tornam invioláveis e atualizáveis apenas por consenso ou acordo entre pares. (ANTONOPOULOS, 2014; BASHIR, 2017; ROUSE, 2017), podendo “[...] ser auditado a qualquer momento [...]” (OLIVEIRA; SANTARÉM SEGUNDO, 2018, p. 5371).

Pela sua capacidade teórica de prover segurança, anonimato e integridade de dados sem a necessidade de um intermediário (terceira parte) que valide as transações, a *Blockchain* foi inicialmente aplicada ao mercado financeiro, mais especificamente na criação da moeda virtual “bitcoin” (YLI-HUUMO et al., 2016).

Criada e publicada por um autor cujo pseudônimo era Satoshi Nakamoto em um artigo de nove páginas cujo título *Bitcoin: A Peer-to-Peer Electronic Cash System*, não deixava transparecer a verdadeira revolução a que estava destinada a causar.

Nakamoto combinou várias das invenções anteriores tais como *b-money* e *HashCash* para criar um sistema de dinheiro eletrônico completamente descentralizado que não dependesse de uma autoridade central para a emissão de moeda ou para a liquidação e validação de transações. (ANTONOPOULOS, 2014, p. 3).

A implementação da ideia se deu em 2009 com a participação e revisão de muitos programadores. Satoshi Nakamoto liderou a implementação até o mês de abril de 2011, quando se afastou do público. “A identidade da pessoa ou pessoas por trás do bitcoin ainda é desconhecida.” (ANTONOPOULOS, 2014, p. 4).

No entanto, nem Satoshi Nakamoto e nem qualquer outra pessoa exerce controle sobre o sistema bitcoin, que opera baseado em

princípios matemáticos totalmente transparentes. A invenção em si é revolucionária e já criou um campo de estudos nas áreas da computação distribuída, economia e econometria. (ANTONOPOULOS, 2014, p. 4)

Hoje, a *Blockchain* é demarcada por três categorias: a *Blockchain* 1.0 criou as aplicações em criptomoedas, o bitcoin; com o *Blockchain* 2.0, os usos relacionados a contratos de todo o tipo, que vão muito além de transações em dinheiro; e ainda, a *Blockchain* 3.0 compreende as aplicações em áreas governamentais, de saúde, ciência, literatura, cultura e artes. (SWAN, 2015).

Um *Blockchain* funciona como um banco de dados descentralizado que é gerenciado por computadores pertencentes a uma rede ponto a ponto (P2P). Cada um dos computadores na rede distribuída mantém uma cópia do livro razão para evitar um único ponto de falha (SPOF) e todas as cópias são atualizadas e validadas simultaneamente. (ROUSE, 2017).

Bartling et al. (2017) atribuem à sua natureza descentralizada, distribuída, imutável e transparente as características singulares que os dados na *Blockchain* adquirem, sendo armazenados sem ambiguidades e de forma transparente, permitindo-se assim que auditorias sejam realizadas.

4.1 A estrutura da *Blockchain*

A estrutura tecnológica básica de uma rede *Blockchain* é composta por uma rede ponto a ponto (P2P) e uma estrutura de dados distribuída. Uma rede ponto a ponto é um conjunto de computadores, que compartilham recursos e tarefas, sem a necessidade de um controle centralizado – ao contrário de uma rede que contenha servidores - onde cada computador é denominado um “nó” desta rede.

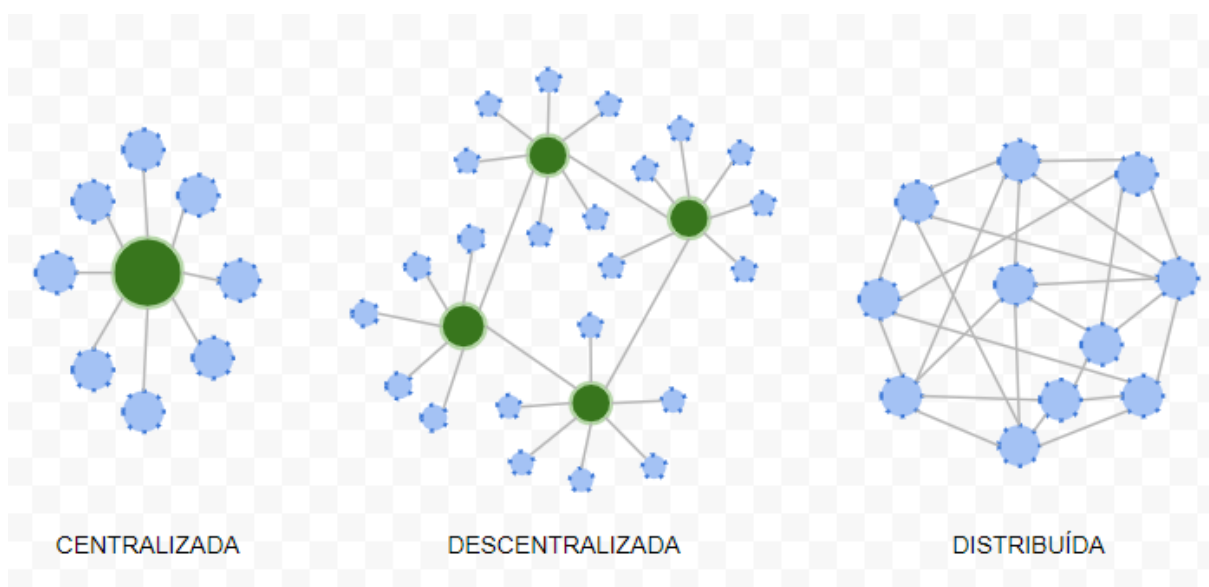
Cada um desses ‘nós’ pode armazenar uma cópia exata de todos os dados das transações realizadas, então, quando um novo ‘nó’ é adicionado, ele pode receber uma cópia dos dados armazenados em outros nós, garantindo que quando o inverso acontece, não haja impactos na rede.

Em sua concepção inicial, para que uma transação seja realizada na *Blockchain*, por exemplo da Bitcoin, cada parte envolvida possui duas chaves, uma pública e uma privada (secreta). Podemos entender a primeira como um endereço público e a segunda um meio de autenticação. “Imagine que a Chave Pública é

similar ao número de uma conta bancária e a Chave Privada similar a um PIN secreto ou uma assinatura em um cheque que provê controle sobre a conta.” (ANTONOPOULOS, 2014, p. 1).

Desta forma, em uma estrutura centralizada ou descentralizada, há a necessidade de um ou mais agentes que realizam as validações de uma ou mais transações, enquanto que, em uma estrutura descentralizada, todos ou pelo menos um número escolhido por consenso de ‘nós’ desta rede possui a autonomia de validar essa transação por meio do atendimento de pré-requisitos acordados, processados e registrados por cada um desses nós, como ilustrado na Figura 2.

Figura 2 - Sistema centralizado, descentralizado e distribuído



Fonte: Elaborado pelo autor.

Nesse sentido, Bartling et al. (2017, p. 5) destaca que

Hoje, o proprietário (ou pesquisador, editor acadêmico, repositório de dados, etc.) tem controle total sobre o computador, os dados e os serviços que executa (por exemplo, um banco de dados) e pode tecnicamente alterar o conteúdo de maneiras arbitrárias. Após a revolução do *Blockchain*, esse não é mais o caso, já que os sistemas de fornecimento de confiança descentralizados fornecem "poder criptográfico" para garantir a integridade de um serviço de computador e a autenticidade do banco de dados subjacente.

Sendo assim, a *Blockchain* é considerada como “[...] um livro-razão compartilhado distribuído” (BASHIR, 2017). Esse atributo permite uma única versão da realidade acordada entre todas as partes da rede sem a exigência de uma autoridade central e, uma vez que os dados são gravados em uma *Blockchain*, é extremamente difícil alterá-los novamente. A dificuldade de alterar os dados é conhecida como imutabilidade que, na verdade, “[...] não é verdadeiramente imutável [...]” (BASHIR, 2017). Esse atributo foi tão bem explicado em uma analogia por Antonopoulos (2014) que optamos por transcrevê-la integralmente:

Uma maneira de imaginar a *Blockchain* seria como um solo, onde os blocos seriam camadas de uma formação geológica, ou como uma amostra do núcleo de uma geleira. As camadas da superfície podem mudar com as estações, ou mesmo serem destruídas antes de terem tempo para se assentarem. Mas quanto mais profundo escavarmos, veremos que maior será a estabilidade das camadas geológicas. Quando você escavar algumas centenas de metros de profundidade, você estará olhando para uma fotografia do passado que permaneceu intocada por milhões de anos. Na *Blockchain*, pode acontecer de os poucos blocos mais recentes tenham que ser revisados/corrigidos, caso haja um novo cálculo da corrente devido a uma bifurcação. Os seis blocos do topo são como os centímetros mais superficiais do solo. Mas quanto mais fundo você penetrar na *Blockchain*, além dos seis blocos, se cada vez menor se torna a probabilidade desses blocos se modificarem. Após 100 blocos de profundidade há tanta estabilidade que a transação Coinbase - a transação contendo os bitcoins recém-minerados - pode ser gasta. Alguns milhares de blocos atrás (um mês) e a *Blockchain* é uma história estabelecida, para todos os fins práticos. Apesar de o protocolo sempre permitir que uma corrente seja desfeita por uma corrente mais comprida, e apesar de sempre existir a possibilidade de qualquer bloco ser revertido, a probabilidade de ocorrência de um evento como esse diminui à medida que o tempo passa, até que ela se torne ínfima. (ANTONOPOULOS, 2014, p. 1).

Da mesma forma, há o atributo de transparência, que estabelece que todos podem ver o conteúdo da *Blockchain*. Sendo compartilhado, o sistema se torna transparente. Essas propriedades podem ser percebidas facilmente aplicadas ao *bitcoin*.

No início da criptomoeda, um bloco gênese foi criado e serviu como o estado inicial do sistema. Ele conteve informações sobre as regras ou instruções sobre a estrutura de dados restante. Conforme as transações foram adicionadas, um novo bloco vai sendo formado e ao atender aos requisitos um novo bloco é adicionado.

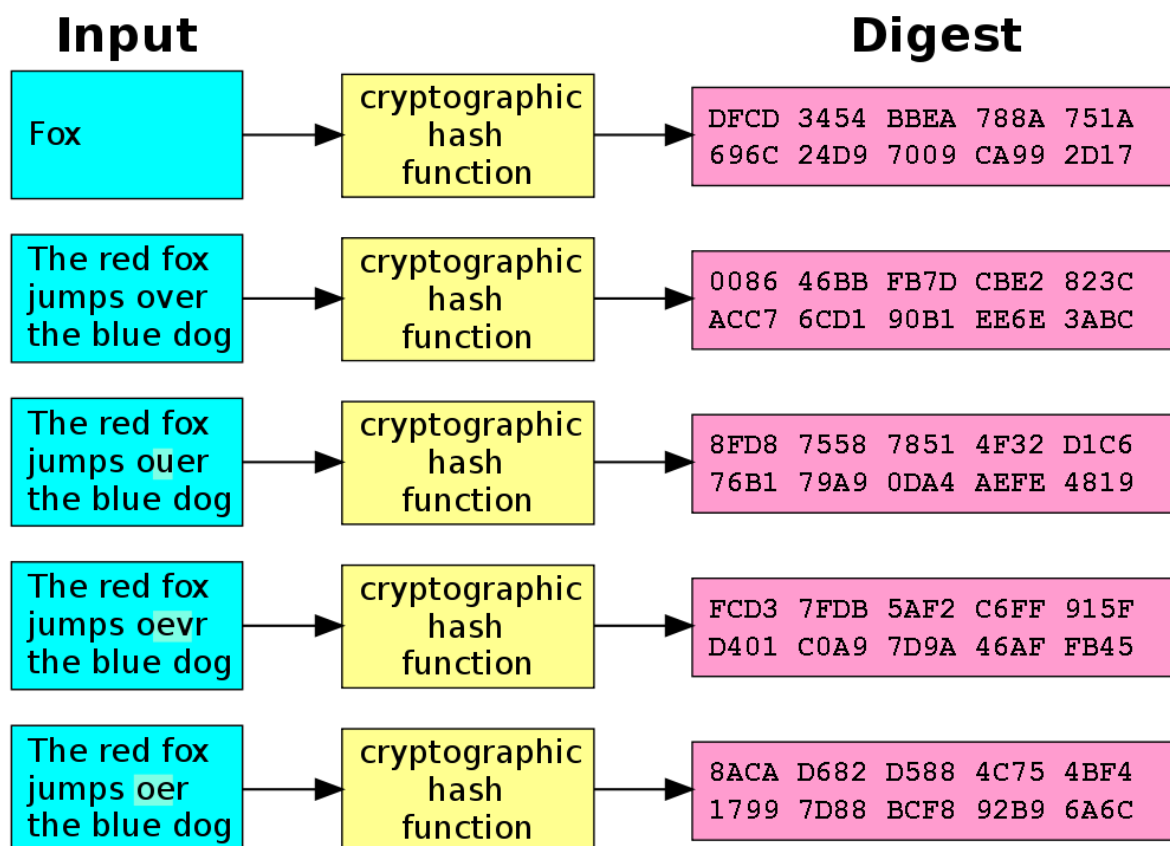
Essa validação, na *bitcoin*, é realizada pelo conceito de “prova de trabalho”, onde um “nó” da rede deve executar o processamento da transação e em seguida

validar seu resultado com os outros “nós” da rede. Ao haver um consenso entre os resultados encontrados por esses “nós”, essa transação é validada e inserida em um bloco. Esse bloco recebe uma identificação, uma assinatura que é construída criptograficamente, contendo dados do bloco adicional e do bloco anterior e que é denominada “*hash* criptográfico”. Então uma “cadeia de blocos” é formada e, daí sua denominação: *Blockchain*.

A função *hash* (resumo) são algoritmos cuja entrada se dá pela cadeia de comprimento arbitrário (uma mensagem) e cuja saída é uma cadeia de comprimento fixo (o valor *hash*), apresentando-se como “[...] um tipo de assinatura para essa mensagem.” (STEVENS, 2007, p. 4).

A Figura 3 ilustra uma função *hash* em funcionamento. Note-se que independentemente do tamanho da entrada (*input*) o comprimento de saída (*digest*) é sempre o mesmo. Outro ponto de atenção é que uma pequena mudança, por exemplo, num único caractere da cadeia de entrada provoca uma saída consideravelmente diferente. As funções *hash* possuem, segundo Schneier (2004), duas propriedades vitais: a primeira é que elas são unidirecionais. Isso significa que “[...] é fácil pegar uma mensagem e calcular seu valor de *hash*, mas é impossível obter um valor de *hash* e recriar a mensagem original.” (SCHNEIER, 2004, on-line). A segunda é que funções *hash* são livres de colisão. Isso significa que “[...] é impossível encontrar duas mensagens com o mesmo valor de *hash*.” (SCHNEIER, 2004, on-line).

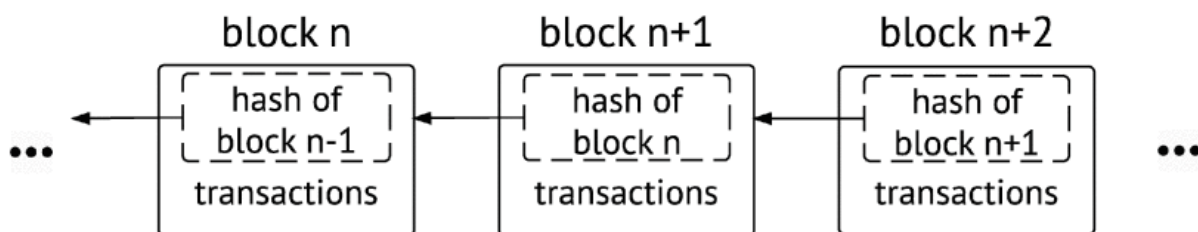
Figura 3 - Uma função *hash* criptográfico em funcionamento.



Fonte: Stolfi (2008).

Christidis e Devetsikiotis (2016) ilustram a *Blockchain* como um registro sequencial de dados de transações agrupados em blocos contendo a data e hora em que cada um foi realizado.

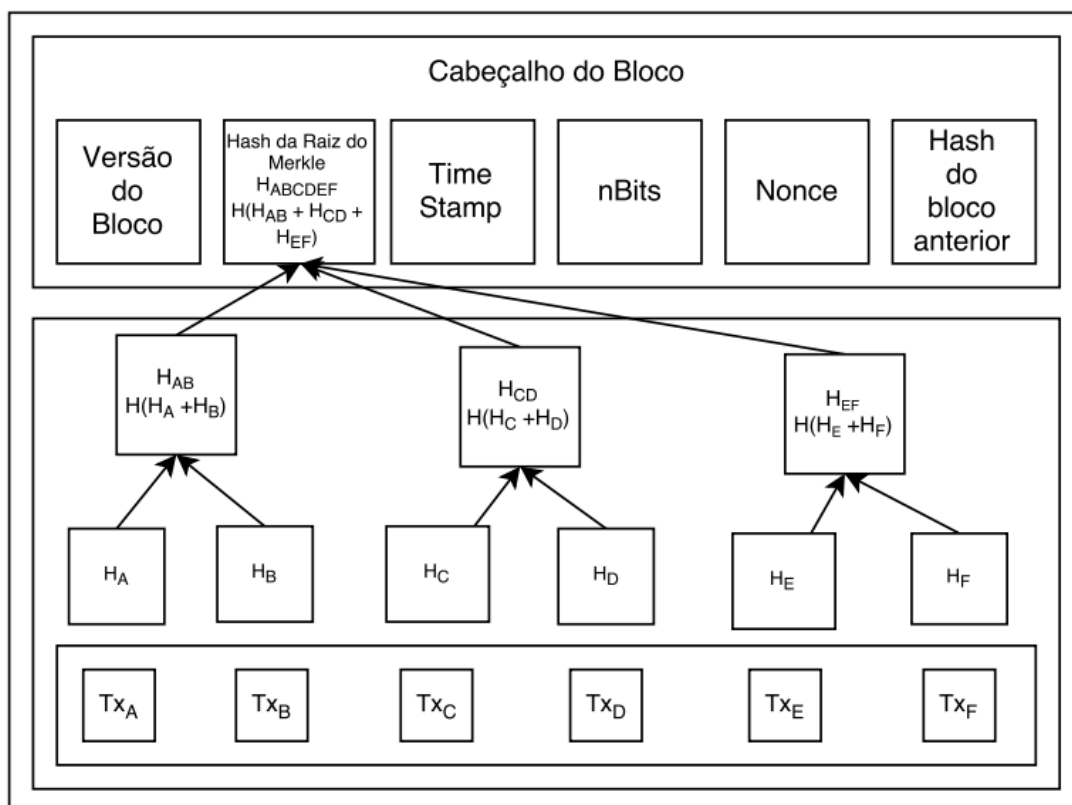
Cada bloco é identificado por seu *hash* criptográfico e cada novo bloco faz referência ao *hash* do bloco anterior. Isso estabelece um link entre os blocos, criando assim uma cadeia de blocos, ou *Blockchain*, conforme ilustrado na Figura 4

Figura 4 - A *Blockchain* representada graficamente

Fonte: Bartling et al. (2017).

Qualquer nó da rede, com acesso a essa lista de blocos ordenados e vinculados poderá lê-lo e descobrir qual é o estado global dos dados que estão sendo transacionados. Em geral, a estrutura de um bloco na *Blockchain*, como por exemplo, o da Bitcoin tem duas partes, conforme ilustrado na Figura 5.

Figura 5 - Estrutura da bitcoin



Fonte: Aliaga e Henriques (2017)

Como visto na figura 5, a primeira parte, seu cabeçalho, é composta de metadados que indicam as regras de validação dos blocos (Versão), a *hash* que representa todas as transações (*Hash Raiz da Árvore de Merkle*¹), a hora universal em que o bloco foi criado (TimeStamp), o indicativo de dificuldade para realização da mineração (Nbits), o valor arbitrário usado na mineração para se produzir o *hash* do desafio (*Nonce*), e o *hash* do bloco anterior que é usado no cálculo do *hash* de desafio. A segunda parte, o Corpo do Bloco, com o conteúdo propriamente dito, incorpora a Árvore de Merkle das transações e as transações propriamente ditas.

Tais características, acrescida da capacidade de interpretação de regras de negócios, têm possibilitado o uso da *Blockchain* em quaisquer outras aplicações onde partes desejam realizar transações automaticamente, sem a necessidade de intermediários.

O que é necessário é um sistema de pagamento eletrônico baseado em prova criptográfica em vez de confiança, permitindo que duas partes interessadas negociem diretamente entre si sem a necessidade de uma terceira parte confiável. (NAKAMOTO, 2008).

Embora sua aplicação inicial tenha sido como uma moeda virtual (criptomoeda), esta estrutura de dados provê inúmeras possibilidades de aplicações que vão muito além das transações financeiras (ULIERU, 2016; BARTLING; et CONTRIBUTORS TO LIVING DOCUMENT, 2017).

Basicamente, a *Blockchain* resolve um dos grandes problemas relacionados a sistemas distribuídos. Conhecido como o Problema dos Generais Bizantinos (*Byzantine Generals Problem – BGP*), que consiste em “[...] tentar tomar uma decisão através do intercâmbio de informações sobre uma rede pouco confiável e potencialmente comprometida.” (ANTONOPOULOS, 2014). Resumidamente, o BGP pode ser enunciado da seguinte forma:

O comandante de um exército tem algumas tropas posicionadas ao redor da cidade. Cada tropa está sob o comando de um general. O comandante dá as ordens para as tropas avançarem ou recuarem, sabendo que existem alguns generais e mensageiros que são traidores. Os generais discutem suas decisões através dos mensageiros. Se todos os generais seguirem as ordens do comandante, eles terão uma grande chance de sucesso. Se

¹ Em criptografia e ciência da computação, árvores de dispersão ou árvores de Merkle são um tipo de estrutura de dados que contém uma árvore de informações resumidas sobre um pedaço maior de dados.

algumas tropas ataquem enquanto outras recuam, o fracasso é iminente. Como os generais garantem que através da troca de mensagens entre eles, todos os generais tomarão a mesma decisão e que essa decisão coincide com a ordem dada pelo comandante? (SALUSTIANO, 2006, p. 88).

Ainda, segundo Salustiano (2006), as questões relativas às correções das falhas de comunicação, evitando que possíveis alterações nas mensagens sejam detectadas pelo inimigo, além de que todos os nós (generais) executem a ordem em concordância com o conteúdo das mensagens recebidas, devem ser resolvidas para garantir a confiabilidade do sistema. “Entre as novidades que o *Blockchain* traz, destacam-se os mecanismos de consenso [...]” (ALIAGA et al., 2018). Esses mecanismos devem permitir que partes desconhecidas cheguem a um consenso sobre o estado dos dados armazenados em uma *Blockchain*.

O consenso é um processo de concordância entre nós desconfiados em um estado final de dados. “A chave para o funcionamento da cadeia de blocos é que a rede deve concordar coletivamente sobre o conteúdo do livro-razão.” (KOSTAREV, 2017). Para alcançar consenso, diferentes algoritmos podem ser usados. É fácil chegar a um acordo entre dois nós (por exemplo, em sistemas cliente-servidor), mas quando vários nós participam de um sistema distribuído e precisam concordar em um único valor, torna-se muito difícil chegar a um consenso. Esse conceito de alcançar consenso entre vários nós é conhecido como consenso distribuído. (BASHIR, 2017).

O Algoritmo de Prova de Trabalho ou *The Proof-of-Work* (POW) em inglês é, sem dúvida, o mecanismo de consenso mais conhecido, em decorrência do pioneirismo e aceitação da criptomoeda *Bitcoin*. Bashir (2017) explica que este mecanismo depende da prova de que recursos computacionais suficientes foram gastos antes de propor um valor para aceitação pela rede. Popularmente é conhecido como “mineração”, “O *proof-of-work* é o mecanismo por meio do qual o consenso descentralizado é alcançado. É a invenção que tornou o *Blockchain* e o *Bitcoin* especiais.” (MATOS, 2018).

Trata-se de um processo pelo qual “uma parte de um dado requer um esforço computacional considerável para ser encontrada.” (ANTONOPOULOS, 2014, p. 2).

A principal inovação [do *Bitcoin*] foi usar um sistema de computação distribuído (chamado algoritmo de “prova de trabalho” ou “*proof of work*”) para conduzir uma “eleição” global a cada 10 minutos,

permitindo à rede descentralizada chegar em um consenso sobre o estado das transações. Isto resolve de forma elegante o problema de gasto duplicado, onde uma única unidade de moeda poderia ser gasta duas vezes. Antes do *Bitcoin*, o problema de gasto duplicado era uma fraqueza do dinheiro digital, e sua solução envolvia a transmissão e verificação de todas as transações através de uma entidade central. (ANTONOPOULOS, 2014, p. 3).

Nesse sentido, “Os computadores competem para encontrar um *hash* com propriedades específicas.” (KOSTAREV, 2017). Esse processo é conhecido como “mineração”. Segundo (MATOS, 2018), o algoritmo adiciona um número arbitrário chamado “*nonce*” ao final do bloco e exige que o minerador encontre um valor que, adicionada aos dados do bloco e ao conteúdo das novas transações, seja gerado um *hash* com uma determinada quantidade de números zeros em seu início. A POW garante o consenso na rede por meio da solução deste problema criptográfico. Quando esse problema é resolvido, um novo bloco na *Blockchain* é criado. Esse bloco é então transmitido a todos os outros nós que armazenam uma cópia completa da *Blockchain* e verificam individualmente sua validade. À medida que um bloco é adicionado, a dificuldade para modificar dados de blocos anteriores vai aumentando, pois, o esforço para mudar um bloco exige a mudança de todos os blocos posteriores. E isso exige um grande poder computacional para resolver o algoritmo de cada um desses blocos. Por esse motivo, a cadeia considerada válida é sempre a mais longa.

Existem várias críticas relacionadas a esse mecanismo, mas as mais contundentes são a ecológica e a de segurança. Para realizar todo o processo de consenso, a POW exige grande poder computacional e, conseqüentemente, grande gasto de energia o que está levando o processo de mineração ser realizado apenas por grandes centros de dados. Por esse motivo, há o risco de que em algum momento, segundo Kostarev (2017) o poder de mais de 51% do processamento possa estar concentrado no poder de uma única corporação, o que daria a ela o poder de realizar alterações arbitrárias nos dados da *Blockchain*. Porém esta seria uma possibilidade remota, uma vez que obter esse poder computacional seria extremamente caro, o minerador corre o risco de não suportar a corrente válida e ser penalizado com o não pagamento, tendo gasto a soma considerável para realizar o ataque. “Conseqüentemente, é melhor (e mais lucrativo) que os mineradores permaneçam honestos.” (KOSTAREV, 2017).

O Mecanismo de Prova de Participação ou *Proof of Stake* (PoS) funciona de forma diferente. Nela, a validação das transações é realizada por um algoritmo de consenso que leva em consideração o valor de coparticipação. Neste modelo, o poder de validação de um nó na rede é equivalente ao percentual de unidades de criptomoedas pertencentes a ela. Assim como na POW, um *token* de recompensa é criado sempre que um bloco é gerado, entretanto, a recompensa é distribuída aleatoriamente entre todos os participantes da rede levando em conta o valor da rede e a duração da propriedade. “É como uma loteria: O vencedor é determinado por acaso, porém quanto mais moedas (bilhetes de loteria) eles possuem, maiores são as chances.” (KOSTAREV, 2017).

Prova de Participação Alugada ou *Leased Proof of Stake* (LPoS) é uma derivação do Mecanismo de Prova de Participação e surgiu para resolver o problema da improbabilidade de pequenos mineradores com baixos índices de participação minerarem blocos, deixando a mineração da rede a cargo de um número limitado de participantes. “Como a segurança da rede é melhor quando há mais participantes, é importante incentivar esses pequenos proprietários a participarem.” (KOSTAREV, 2017).

A Prova Delegada de Participação (DPoS) ou *Delegate Proof-of-Stake* (DPOS) também é derivada da POS. Nela, os detentores de moedas utilizam seus saldos para eleger os nós que poderão minerar e adicionar blocos à *Blockchain*.

O mecanismo de Tolerância a falhas bizantinas práticas ou *Practical Byzantine Fault Tolerance* (PBFT) é usado quando mais de 1/3 dos nós podem ser maliciosos. Daí seu nome. Em cada uma de suas operações, um nó deve receber os votos de 2/3 de todos os outros nós para poder dar continuidade no processo.

O *Ripple*, segundo Aliaga e Henriques (2017) é um algoritmo de consenso que utiliza sub redes coletivamente confiáveis que inclui um novo bloco na *Blockchain* se uma transação atinge uma aprovação relativa de 80% dos nós que participam do processo de consenso, dentre os nós já previamente denominados como confiáveis.

A Prova de Importância ou *Proof of Importance* (PoI) se baseia na “[...] ideia de que a atividade de rede produtiva, e não apenas a quantidade de moedas, deve ser recompensada.” (KOSTAREV, 2017). A chance de minerar um bloco leva em consideração diversos fatores, incluindo o equilíbrio, a reputação e o número de transações feitas para e a partir do endereço de um minerador. “Isso fornece uma imagem mais holística de um membro da rede “útil”.” (KOSTAREV, 2017, online).

Esta não é uma lista exaustiva dos tipos de mecanismos. Importante lembrar que uma *Blockchain* pode usar um ou mais mecanismos, individualmente ou simultaneamente para atingir seus objetivos.

Almeida (2012) atribui ao direito natural e ao surgimento do capitalismo a forma como entendemos o conceito de contrato na atualidade. No princípio, buscava-se eliminar barreiras e garantir os princípios que nortearam os ideais capitalistas como a autorregulação do mercado, neste sentido, no contexto do direito contratual

[...] a principal ideia traçada nesse período era a da liberdade de contratar. O sujeito era livre para escolher contratar ou não contratar, escolher o seu parceiro contratual, além de estabelecer o conteúdo desse contrato. Não era dado ao Estado impor às partes um determinado tipo de contrato ou a contratar com determinado parceiro contratual. (ALMEIDA, 2012).

Com o passar do tempo os contratos “[...] passam a ser uma forma de opressão para os economicamente mais fracos.” (ALMEIDA, 2012) forçando o Estado a passar “[...] a dirigir os contratos para que esses mantenham o equilíbrio. Passa-se a admitir a revisão de um contrato que passasse a ser desequilibrado.” (ALMEIDA, 2012), assim, o contrato

[...] continua fazendo lei entre as partes, mas com respeito à dignidade da pessoa humana e de todas as normas de ordem pública que o capacitam a ser instrumento de circulação de riquezas, mas destinado a ser um instrumento mais democrático e justo do direito privado. (LEITE, 2007).

Assim, surge a figura do terceiro confiável como entidade mediadora aceita pelas partes de um contrato e que vem sendo o alicerce de todas as atividades envolvendo relações, em especial, as econômicas. Esta situação prevalece ainda hoje, entretanto com o surgimento de tecnologias como os Contratos Inteligentes temos novamente amplas possibilidades.

O conceito de Contrato Inteligente foi formulado por Nick Szabo, definindo-o como “[...] um protocolo de transação informatizado que executa os termos de um contrato.” (SZABO, 1994). A ideia básica por trás dos contratos inteligentes é que muitos tipos de cláusulas contratuais “[...] podem ser incorporados no hardware e *software* com os quais lidamos, de modo que violar o contrato seja proibitivamente caro para o infrator.” (SZABO, 1997).

Não é um conceito novo. Szabo (1997) ilustra o conceito utilizando como exemplo uma máquina de venda automática (*vending machine*) que, para ele pode ser considerada a ancestral dos Contratos Inteligentes.

Dentro de uma quantidade limitada de possibilidade de perda (o valor disponível no caixa deve ser menor do que o custo de romper o mecanismo), a máquina recebe moedas e, por meio de um mecanismo simples, que cria um problema de informática no projeto com autômatos finitos, entrega o produto de acordo com o preço exibido. A máquina de venda automática é um contrato com o portador: qualquer pessoa com moedas pode participar de uma troca com o vendedor. O cofre e outros mecanismos de segurança protegem as moedas e o conteúdo armazenados dos invasores, o suficiente para permitir a implantação lucrativa de máquinas de venda automática em uma ampla variedade de áreas. (SZABO, 1997, *on-line*).

Evidentemente que Szabo (1997) esclarece que os Contratos Inteligentes vão muito além das possibilidades de uma máquina automática de vendas, apresentando vários exemplos hipotéticos de aplicação para automóveis, como: segurança, transferência de propriedade, garantia de crédito, locação, arrendamento, etc.

Com o surgimento da *Blockchain*, dos Mecanismos de Consenso e dos Contratos Inteligentes que são tecnologias que, no âmbito jurídico, devem “[...] transformar a resolução de disputas [...]” (HOGEMANN, 2018, p. 111), não restando dúvidas de que a combinação de tecnologia *Blockchain* aos contratos inteligentes estão “[...] afetando os pressupostos, doutrinas e conceitos legais tradicionais.” (HOGEMANN, 2018, p. 109), ao ponto em que “[...] os algoritmos substituirão os juízes em alguns casos, com documentos escritos em código legível por máquina, como contratos inteligentes auto impositivos.” (HOGEMANN, 2018, p. 111).

Ao menos, em teoria e até o presente momento, Contratos Inteligentes são muito eficazes como “[...] instrumentos portadores digitais em plataformas descentralizadas como o *Bitcoin*.” (SONG, 2018, *on-line*), mas problemáticos quando há a necessidade de, por exemplo, vincular um ativo digital a um ativo físico (tokenização),

[...] seja frutas, carros ou casas, pelo menos em um contexto descentralizado. Os ativos físicos são regulados pela jurisdição em que você se encontra e isso significa que eles estão, de certo modo, confiando em algo além do contrato inteligente que você criou. (SONG, 2018).

Isso não significa, portanto, que mesmo que a posse de um ativo seja transferida digitalmente, a posse física tenha sido realizada no mundo real, fazendo com que o Contrato Inteligente, neste caso, sofra “[...] do mesmo problema de confiança dos contratos normais.” (SONG, 2018, *on-line*). O problema só seria resolvido se o ativo físico (hardware) em questão fosse ou tivesse incorporado em um Contrato Inteligente. Isso seria possível se cada objeto se tornasse um dispositivo conectado (IoT) ao mesmo sistema que executaria o contrato inteligente, facultando, por exemplo, o acionamento e uso do dispositivo apenas ao seu proprietário.

Em relação aos metadados da *Blockchain*, os autores Elena García-Barriocanal, Salvador Sánchez-Alonso e Miguel-Angel Sicilia (2017, p. 02) afirmam que o

[...] *Blockchain* e as tecnologias associadas fornecem um novo tipo de plataforma para superar alguns dos problemas da atual tecnologia de depósito de dados e, portanto, sugerem aos usuários a construção de uma abordagem descentralizada para o arquivamento de recursos digitais. No entanto, a reformulação do atual sistema de arquivos, agregadores e serviços digitais requer uma consideração cuidadosa das funções dos metadados, suas propriedades desejáveis e até que ponto as diferentes tecnologias são capazes de apoiá-los. Além disso, sistemas descentralizados baseados em *Blockchain* não são isentos de riscos, especialmente porque são construídos em torno de sistemas de incentivos para os participantes da rede, portanto a sustentabilidade deve ser incorporada em seu design.

Nesse sentido, aponta-se a importância dos metadados e padrões de metadados na organização e representação desses dados, inclusive na tecnologia *Blockchain*. Sua importância fica mais clara quando é observado os requisitos que são ou não suportados pela rede *Blockchain*. Os requisitos não suportados, incluem a digitação de mídia, relacionado à capacidade de longo prazo de entender, processar e renderizar objetos digitais. O armazenamento descentralizado só é referido aos metadados ao *software* de especificações necessárias, mas isso continua a ser uma questão de preservar a capacidade de processamento que foi adicionada antes, tipicamente via emulação ou migração. (GARCÍA-BARRIOCANAL; SÁNCHEZ-ALONSO; SICILIA, 2017, p. 04). Os requisitos que são impactados por soluções descentralizadas, são percebidos por meio da indexação, e não é um recurso diretamente suportado pelas *Blockchains* públicas, assim, estes requisitos precisam de uma infraestrutura adicional para a proveniência inviolável e

ainda, ocorre com referenciamento (*link*) que assume uma forma diferente se referência recursos em sistemas descentralizados. (GARCÍA-BARRIOCANAL; SÁNCHEZ-ALONSO; SICILIA, 2017, p. 04). Por fim, os requisitos que são diretamente suportados pela rede *Blockchain*, que incluem a identificação, desreferenciação e prova de declaração como funções que são diretamente suportadas por uma combinação da *Blockchain* e um sistema de arquivo descentralizado. (GARCÍA-BARRIOCANAL; SÁNCHEZ-ALONSO; SICILIA, 2017, p. 04).

A partir disso, nota-se que a *Blockchain* pode atingir níveis mais altos de disponibilidade, transparência e resistência a adulterações, o que resolveria alguns dos problemas dos sistemas de metadados atuais construídos em bancos de dados convencionais e, normalmente, em sistemas presentes da *Web*.

4.2 A *Blockchain* e suas potenciais aplicações na Ciência

A despeito do uso em seu aspecto monetário ser o mais frequente, a *Blockchain* tem sido usada em diversas outras aplicações. Por exemplo, como base para um portfólio de dados para produtores musicais (músicos, cantores, compositores, estúdios, consumidores, produtores, artistas, etc.) e que contém dados sobre seus perfis, premiações, obras, direitos, etc. que podem ser transacionados pela aplicação. Um contrato inteligente (*Smart Contract*) está programado para efetuar as validações necessárias de acordo com cada tipo de transação. (MYCELIA, 2018).

No campo científico, o movimento *Blockchain for Science* tem realizado diversos estudos colaborativos do uso da *Blockchain* na Ciência. Liderados pelo fundador do movimento, o Dr. Sönke Bartling, radiologista, pesquisador em Ciências básicas em imagiologia médica e editor do livro *Opening Science* da editora Springer Open, o movimento vem realizando estudos teóricos, publicando seus estudos em revistas e eventos e desenvolvendo projetos utilizando a *Blockchain* nos mais diversos aspectos da Ciência, principalmente relacionados à Ciência Aberta. O movimento tem por missão tornar a Ciência e a criação de conhecimento realmente aberta por meio da “(r)evolução *Blockchain*” (*BLOCKCHAIN FOR SCIENCE*, 2017).

O movimento possui uma denominação legal chamada *International Society of Blockchain For Science* (IBFS), que

[...] é formada por uma ampla gama de especialistas em *Blockchain*, incluindo pesquisadores, bibliotecários, defensores de direitos, empreendedores de tecnologia, especialistas em publicação acadêmica, especialistas em comunicação e gerenciamento comunitário e economistas criptográficos. (*BLOCKCHAIN FOR SCIENCE*, 2018, *on-line*).

A missão desta organização é conduzir a comunidade científica no uso da tecnologia *Blockchain* para “tornar o ecossistema de pesquisa e a criação de conhecimento mais eficiente”. (*BLOCKCHAIN FOR SCIENCE*, 2018). Em seu documento constituinte, a IBFS visa atingir seus objetivos empregando princípios científicos abertos de modo a acelerar a disseminação de descobertas científicas, observações e ideias; impedindo o controle centralizado ou fechado sobre quaisquer partes do ciclo de pesquisa, especificamente dados, publicações, incentivos e avaliações, de modo a evitar modelos de negócios artificiais (intermediários); evitando-se fraudes de pesquisas, garantindo a privacidade do sujeito de pesquisa, assegurando uma conduta ética e moral de alto nível e a contínua discussão de implicações éticas e garantindo a criação de conhecimento aberto e autogovernado na internet, servindo assim, a comunidade global de usuários.

A IBFS enxerga na *Blockchain* novos potenciais por meio de uma combinação de tecnologia, teoria dos jogos, mudanças sócio legais, incentivos econômicos e criptografia forte que se manifesta de uma infraestrutura descentralizada de computadores e redes ponto a ponto; economia de criptografia e tokenização; investimentos em *crowdsourcing* usando ofertas de moedas iniciais (ICO); novas proposições de valor para a criação de conhecimento; mercados de dados; hardware confiável; novos modelos de pesquisa de privacidade (proteção) e novos modelos de métodos de processamento de dados.

Van Rossum (2017) aponta que o potencial da *Blockchain* está em transformar a comunicação acadêmica e a pesquisa em geral, ressaltando a solução para alguns aspectos críticos da comunicação acadêmica, como a transparência, confiança, reprodutibilidade e direitos autorais.

As ideias e projetos mais relevantes para as questões da reprodutibilidade da pesquisa pela IBFS podem ser resumidas no Quadro 6, abaixo:

Quadro 6 - Ideias e projetos relevantes para as questões de reprodutibilidade da pesquisa

Autenticação	Serviço de registro de data e hora para conteúdo digital para provar que possuía determinada informação, arquivo em um horário específico criando uma trilha de dados imutáveis. Aplicações como essas poderão impactar nos registros de propriedade intelectual.
Planejamento / DMP / Projetos	O projeto de pesquisa e o Plano de gestão de dados (DMP) pode ser pré-registrado em uma <i>Blockchain</i> , de modo a evitar alterações arbitrárias no projeto. Isso impactaria principalmente na supressão do viés de publicação.
Aquisição, gestão, análise de dados, compartilhamento e publicação de dados de pesquisa.	Embora o uso da <i>Blockchain</i> possa garantir que a auditoria de dados possa ser realizada - uma vez que todos os dados da pesquisa podem ser registrados em uma <i>Blockchain</i> – o simples registro por si só não garante a confiabilidade dos dados. Os dados poderiam ser coletados por dispositivos em uma internet das coisas (IoT). O sensor criptografaria diretamente os dados e os armazenaria em <i>Blockchain</i> . Assim, os dados seriam armazenados de forma imutável, podendo impedir manipulações arbitrárias de dados, conscientemente ou inadvertidamente. A análise, pós-processamento e estatística dos dados pode ser protocolada em uma <i>Blockchain</i> , revisada, comprovada pelos pares, publicada e compartilhada.
Avaliação da Pesquisa	Sistemas de reputação científica podem ser construídos usando <i>Blockchain</i> sem um terceiro confiável.
Financiamento da Pesquisa	<i>Blockchain</i> poderia fornecer muitas novas maneiras de distribuir dinheiro para pesquisa. Tokens de criptomoeda poderiam ser usados para relacionar fundos de pesquisa a algumas condições, mesmo se o mecanismo de distribuição fosse anônimo e "caixa preta" na <i>Blockchain</i> . A <i>Blockchain</i> ajudaria os pesquisadores a conduzir e publicar melhor suas investigações, mas também poderá envolver o público por meio de um processo de pesquisa mais transparente. Em última análise, essa tecnologia poderia abrir o processo acadêmico ao público para investigação e até participação, ao mesmo tempo em que salvaguardaria a integridade de suas pesquisas. Esse acesso aberto pode inspirar e permitir que pesquisadores amadores colaborem com pesquisadores profissionais em um esforço para crowdsourcing usando os princípios da Ciência do cidadão.

Fonte: *Blockchain for Science* (2018).

No Brasil, onze formandos da Universidade Federal da Paraíba (UFPB) tiveram seus documentos registrados em *Blockchain*, sendo o primeiro caso brasileiro de diplomas universitários registrados com o uso da tecnologia.

Outras duas iniciativas baseadas em *Blockchain* são o *Scienceroot* (disponível em: <https://scienceroot.com/>) e o *Pluto* (disponível em: <https://pluto.network/>). Mesmo sendo plataformas iniciais, elas já começam a trazer

as primeiras aplicações de um sistema que funciona para os protocolos da Comunicação Científica, baseado em *Blockchain*.

Nota-se, portanto, uma miríade de aplicações potencialmente capazes de minimizar os problemas relacionados à reprodutibilidade da pesquisa, comprovando o interesse da comunidade científica na aplicação da *Blockchain*.

Idealmente, uma pesquisa científica seguiria um fluxo contínuo e coordenado de etapas que poderiam ser previamente acordadas e planejadas para obter um resultado irrefutável, avançando com base em estudos prévios e na confiabilidade e reprodutibilidade dos resultados, portanto, por definição, a legitimidade de uma pesquisa depende de consenso entre os pares de uma comunidade. Entretanto, como aponta Simmons, Nelson e Simonsohn (2011), a respeito de resultados em pesquisas em psicologia que geram falsos-positivos, descobrir verdades sobre o mundo gerando hipóteses, coletando dados, examinando-as e verificando se elas são consistentes com as hipóteses, normalmente não acontece de forma tão fluida, “[...] embora aspiremos a ser sempre precisos, os erros são inevitáveis.” (SIMMONS; NELSON; SIMONSOHN, 2011, p. 1359).

Infelizmente, a eliminação total do erro humano só pode ocorrer se não houver seres humanos; o que não é realmente um estado desejável. Tudo o que podemos esperar é minimizar os erros enquanto maximizamos sua rastreabilidade. (GIBB, 2014).

Claro que não podemos ser ingênuos, uma vez que “[...] é tentador para os pesquisadores enviar conjuntos de dados selecionados para publicação, ou mesmo para manipular dados para se adequarem à hipótese subjacente.” (BEGLEY; ELLIS, 2012).

Daí advém a necessidade de uma estrutura consistente de Curadoria de Dados. Abbott (2008) levanta uma série de questões a serem consideradas em relação à curadoria digital. Lembra que é um processo contínuo e não uma ação pontual e que é necessário ter uma infraestrutura financeira e política adequada. Observa também que as atividades envolvidas na curadoria digital podem ser compartilhadas entre diferentes instituições e comunidades. Pontua que a propriedade dos dados digitais é complexa e que mecanismos de controle de qualidade, autenticação e validação de dados devem fazer parte de sua estrutura. Finalmente, lembra que backup de dados não deve ser considerado como curadoria digital.

Uma vez que a *Blockchain* é um mecanismo de confiança, sua aplicação se efetivaria na adoção de uma filosofia baseada na abertura e descentralização da Ciência, permitindo a reprodução dos resultados da pesquisa de forma transparente, inclusive, com o uso de dados gerados fora do âmbito acadêmico, aumentando o impacto social do pesquisador e economizando tempo e dinheiro tanto para pesquisadores quanto para instituições de pesquisa.

Uma possibilidade de processo de reuso dos dados poderia ser com o uso de diversas tecnologias tendo como base a *Blockchain*. Sabemos que, além do projeto, algumas agências vêm solicitando a preparação de documentos que demonstrem como os dados da pesquisa serão coletados, gerados, tratados, analisados, armazenados e compartilhados. Esse documento, denominado Plano de Gestão de Dados, do inglês *Data Management Plan* (DMP), é majoritariamente apresentado como um texto e que responde a duas questões: Quais dados serão gerados pelo projeto e como eles serão preservados e disponibilizados, considerando questões éticas, legais, de confidencialidade e outras? (FAPESP, 2019, *on-line*).

A *Blockchain* já poderia ser utilizada neste momento. Todas as diretivas principais do projeto poderiam ser convertidas em uma Smart Contract que conteria parte ou até mesmo todas as cláusulas referentes a cada uma das etapas do projeto, como num contrato, com a diferença de que os termos devem ser descritos de forma a serem processados automaticamente por sistemas computacionais.

A agência pode ainda estabelecer *tokens* referentes ao projeto e repassar os valores ao grupo de pesquisa interessado por meio de criptomoedas. Isso pode garantir a transparência nas transações financeiras. Adicionalmente, o projeto poderia captar fundos para uma oferta inicial de criptomoeda (ICO).

O mais importante para o presente trabalho é a coleta dos dados. Vamos nos limitar a um único conjunto de dados: a temperatura da água ao longo do tempo. Esta coleta seria feita por um dispositivo eletrônico computacional (um hardware) que poderia conter diversos sensores, exemplarmente neste caso, um termômetro.

Esse aparelho, ligado à internet, seria considerado um dispositivo de IoT (*Internet of Things* ou Internet das coisas). Esse dispositivo conteria um GPS e um relógio interno. O GPS informaria a exata localização e o relógio a data e hora exata da coleta. Ele seria previamente programado com as diretivas dos *Smart Contracts* que poderiam conter a frequência com que os dados seriam coletados, o volume,

as condições, etc. O dispositivo a partir de então seria “lacrado” e sua identificação, um endereço físico único associado ao dispositivo eletrônico de comunicação em uma rede (um *MAC Address*) por exemplo, relacionada ao projeto em questão.

Ao iniciar a coleta, outros dados relevantes como as condições climáticas, posição do sol, da lua, entre outros poderiam ser coletados pela internet e serem adicionados ao registro coletado.

Esses dados seriam enviados a uma *Blockchain*, cujo mecanismo de consenso poderia ser similar ao Mecanismo de Prova de Importância (*Proof of Importance - PoI*) que se baseia na “ideia de que a atividade de rede produtiva, e não apenas a quantidade de moedas, deve ser recompensada.” (KOSTAREV, 2017, *on-line*). A chance de minerar um bloco levaria em consideração diversos fatores, incluindo a reputação, índices bibliométricos, altimétricos e cientométricos do minerador e o número de transações feitas para e a partir do seu endereço.

Ao serem minerados, esses dados fariam parte da *Blockchain* e tornando-se imutáveis ao longo do tempo, garantindo que os dados brutos para a análise sejam comprovadamente autênticos e teoricamente à prova de manipulações.

As consequências de um sistema como esse, permitiria para o desenvolvimento de uma nova pesquisa, entretanto podemos vislumbrar impactos positivos relacionados à reprodutibilidade da pesquisa e ao reuso dos dados dessas pesquisas.

Com os dados coletados e gravados a partir de um dispositivo “lacrado” e cujos parâmetros foram configurados a partir de um *Smart Contract* que representa o projeto de pesquisa, ficará muito difícil que a pesquisa tome rumos que não os acordados no início do projeto, sem que uma nova concordância entre o pesquisador e o grupo de pesquisa e a agência financiadora. Além disso, cada alteração ficaria registrada, gerando assim um registro transparente de toda e qualquer alteração necessária ao projeto e que pode ser considerado em pesquisas de reprodutibilidade posteriores. Isso acontece pela natureza descentralizada da *Blockchain*, tendo cada participante da rede uma “cópia” dos registros que são distribuídos, de forma absolutamente sincronizada e derivada de um processo de consenso entre eles.

Essa proposta ajudaria a minimizar os riscos gerados pelas causas de não reprodutibilidade elencados no documento *intitulado Replication Studies: Improving reproducibility in the empirical sciences*, da *Royal Netherlands Academy of Arts and*

Sciences (2018), especialmente os causados pelo projeto experimental ineficiente associado a controle de vieses falho, fraude ou fabricação de dados, omissão de resultados nulos ou análise seletiva que faz os nulos parecerem positivos, não compartilhamento de dados ou de detalhes metodológicos, escolha de variáveis que se adequam aos resultados, formulação de hipótese depois que os resultados são conhecidos, discrepância entre os resultados registrados e os publicados, sistema de financiamento à pesquisa demasiadamente competitivo e falta de recompensa para práticas que favoreçam a replicação de estudos.

5 CONSIDERAÇÕES FINAIS

A possibilidade de avaliar criticamente as afirmações e conclusões tiradas por outros cientistas é o cerne da prática científica e, para isso, devem poder encontrar recursos e descrições suficientemente detalhadas nos trabalhos publicados a fim de obterem os mesmos resultados dentro das margens suficientes de erros experimentais. Dentre os recursos, os dados coletados durante as pesquisas são essenciais, pois o resultado obtido deve ser o produto de sua rigorosa análise.

Ao contrário do esperado, experimentos vêm obtendo baixos índices de reprodutibilidade. Dentre as causas, grande parte estão relacionadas aos dados das pesquisas, ao rigor metodológico, à capacidade e à ética do pesquisador e de seu grupo ou comunidade de pesquisas.

Blockchain não é apenas sobre dinheiro e a Ciência não é apenas sobre obter informações. Ambos são sobre confiança, ou melhor, sobre confiabilidade e consenso. Em Ciência buscamos consenso racional entre os pares e em *Blockchain* consenso sobre o estado de um conjunto de registros de transações.

Em ambos os casos o consenso pode ser obtido através de mecanismos que possam garantir transparência por todo o ciclo da pesquisa.

Assim, por seus atributos, vislumbramos na tecnologia *Blockchain* possibilidades de elevar os índices de reprodutibilidade, através de mecanismos que possam aumentar a transparência e a confiabilidade dos resultados de uma pesquisa, alcançando assim o objetivo geral da pesquisa que consiste em analisar a *Blockchain* nas iniciativas de reprodutibilidade dos resultados das pesquisas científicas.

Abordando a Comunicação Científica para obtermos uma contextualização da *Blockchain*, no Capítulo 2, buscamos identificar os papéis da Comunicação Científica traçando um panorama histórico desde a Grécia antiga até os atuais paradigmas envolvendo cenários científicos de alta complexidade, passando pelas trocas de cartas manuscritas pelos pesquisadores, pelo surgimento da imprensa, pelos periódicos científicos e sua crise nos anos de 1980 que culminaram no surgimento de movimentos de Acesso Livre e Ciência Aberta.

A crise dos periódicos nos anos 1980 ilustra os efeitos da centralização nas mãos de algumas poucas empresas que monopolizavam o mercado da

Comunicação Científica e a necessidade de se adotar mecanismos mais abertos e descentralizados de gestão nos fluxos das informações científicas.

A Ciência Aberta possibilitou enormes avanços, ampliando a visibilidade e abertura para o engajamento do público em geral e a colaboração e fortalecimento de comunidades não tradicionais, trazendo níveis complexos de desafios no que concerne à confiabilidade e transparência dos processos envolvidos. Para que alcance todo o seu potencial, é necessário implementar mecanismos que considerem todas essas novas características da Ciência Aberta, uma vez que o objetivo da prática científica continua sendo possibilitar avaliar criticamente as afirmações e conclusões tiradas por outros cientistas, sendo que para isso, replicar a pesquisa é essencial para a maioria áreas.

Se replicar uma pesquisa de alta complexidade se tornou praticamente impossível, resta-nos verificar as evidências reunidas nela e é este o ponto que acreditamos mais vulnerável do atual modelo de prática científica.

Por fim, é importante destacar a importância da Ciência da Informação nas questões relativas à prática científica, em especial às preocupações muito anteriores aos de outras áreas com relação aos paradigmas e aos dados gerados pela ciência como, por exemplos a notável semelhança entre os paradigmas de Gray de 2009 em relação ao de Heim em 1982 e os princípios *FAIR* de Wilkinson de 2016 em relação ao propósito de Bush em 1945.

Portanto, é nítida a procedência dos conhecimentos gerados dentro da Ciência da Informação em relação aos atuais movimentos denominados Ciência Aberta, Humanidades Digitais, *Big Data*, *Data Science* ou *e-Science*.

O primeiro objetivo específico, identificar os atuais temas relacionados à gestão dos dados científicos, é atendido no Capítulo 3, onde procuramos demonstrar através de evidências históricas as relações entre o ato de curar, cuidar e preservar o conhecimento com o Poder. Da Biblioteca de Alexandria, passando pelo Mundaneum de Paul Otlet e Henry de La Fontaine às preocupações de Vannevar Bush com a organização da produção científica e sua recuperação tiveram como objetivo prover acesso ao conhecimento humano, afinal, conhecimento é poder.

Cada um desses empreendimentos pode ser considerado como manifestações das tecnologias de suas respectivas épocas que causaram aumentos consideráveis na produção de registros em relação à tecnologia anterior,

ou seja, o problema da explosão informacional não é recente e nela incluímos a sua mais recente versão: a *Big Data*. Diz-se que o que caracteriza a *Big Data* são seus atributos representados pela letra “v”: volume, variedade, veracidade e valor, mas, o que de fato difere a *Big Data* de suas versões antecessoras é que o registro e o acesso a essas informações são realizadas em rede e em escala atômica do conceito *ACID* (acrônimo de Atomicidade, Consistência, Isolamento e Durabilidade - do inglês: *Atomicity, Consistency, Isolation, Durability*) exigindo para seu acesso, poder computacional maior que a de um ser humano.

Assim, os conjuntos completos dos dados deveriam estar disponíveis e abertos, pois não há as tradicionais limitações do passado, como os custos econômicos do papel ou restrição de memórias eletrônicas.

Neste capítulo também procuramos entender o conceito de dados de pesquisa, bem como entender os modelos propostos para a gestão do Ciclo de Vida dos Dados, que têm em comum, o destaque para a importância da documentação e descrição dos conjuntos de dados (metadados), os recursos e os processos envolvidos na sua geração.

O segundo objetivo específico, descrever a estrutura e os atributos da *Blockchain*, e o terceiro objetivo, elencar possíveis aplicações da *Blockchain* no contexto da Comunicação Científica, são atendidos no Capítulo 4, onde descrevemos sua estrutura tecnológica, atributos e aplicações.

É muito cedo para afirmarmos que a *Blockchain* ou qualquer uma das tecnologias em seu entorno, devido aos custos de sua implementação e em seus estágios atuais de desenvolvimento, resolverão todos os problemas relacionados à crise de reprodutibilidade. No curto prazo, o mais provável é que a *Blockchain* seja aplicada a pontos específicos dentro do fluxo de uma pesquisa científica e ajude a dirimir a desconfiança nos dados e nas metodologias utilizadas.

Mas não podemos ser tão céticos, uma vez que as aplicações relacionadas ao tema ainda estão em fases embrionárias de desenvolvimento, implementação e aceitação, uma vez que não é possível nem mesmo afirmar que suas implementações esteja próximo dos objetivos a que foram projetados, trazendo inúmeras oportunidades de pesquisas e desenvolvimentos para a Ciência da Informação.

Assim, são nítidas as relações entre a Ciência da Informação e as temáticas relacionadas aos dados científicos e à *Blockchain*. A Ciência, e em especial, a Ciência da Informação devem contribuir melhorando

[...] a descoberta de conhecimento através da assistência a seres humanos e seus agentes computacionais, na descoberta, acesso e integração e análise de tarefas apropriadas aos dados científicos e outros objetos digitais acadêmicos. (WILKINSON et al., 2016).

Portanto, são inúmeras as oportunidades de pesquisas futuras relacionadas ao presente trabalho a começar pela própria implementação das tecnologias *Blockchain* e *Smartcontracts* apresentadas nesta pesquisa em um protótipo funcional.

Entender os impactos causados por esta mentalidade descentralizada e aberta geram uma série de questionamentos passíveis de pesquisa. A revisão por pares, por exemplo, apontada como uma das vinte causas de não reprodutibilidade pela *Royal Netherlands Academy of Arts and Sciences*, poderia ser aprimorada ou talvez até mesmo suprimida do processo através de novos modelos de revisão automatizada.

Outro aspecto a ser verificado é se a utilização de contratos inteligentes poderiam motivar o pesquisador a realizar um melhor planejamento de sua pesquisa, uma vez que cada um dos aspectos seriam firmados em acordos que seriam executados automaticamente à partir do cumprimento das condições acordadas, permitindo, inclusive o desenvolvimento de novos modelos de sistemas de reputação científica que podem ser construídos usando *Blockchain* sem um terceiro confiável.

Por outro lado, o uso de criptomoedas associadas aos *smartcontracts* permitiriam um modelo mais justo e distribuído de recompensas de acordo com a contribuição de cada participante de uma pesquisa e, assim, o financiamento descentralizado das pesquisas talvez seja possível através de uma participação mais democrática no desenvolvimento de pesquisas com real interesse para a sociedade.

Aspectos como a descrição dos conjuntos de dados na *Blockchain* suscitam pesquisas mais detalhadas, uma vez que uma ampla variedade de tipos de dados de pesquisa está envolvida em inúmeras áreas.

Para nós, o acervo é o mundo e acreditamos que o futuro da biblioteconomia, da Ciência da Informação será maravilhoso quando entendermos que somos muito mais do que paredes com algo dentro. Somos uma mentalidade. Um *mindset*. Um modelo mental que causa, há séculos, transformações de alto impacto na sociedade através da informação. É preciso resgatar esse *mindset*.

REFERÊNCIAS

- ABBOTT, D. **What is Digital Curation?** Disponível em: <<http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/what-digital-curation>>. Acesso em: 27 nov. 2018.
- ALIAGA, Y. E. M. et al. **Avaliação de mecanismos de consenso para *Blockchains* em busca de nova estratégia mais eficiente e segura | Anais do Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (SBSeg)**. Anais do Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (SBSeg). **Anais...** In: SIMPÓSIO BRASILEIRO EM SEGURANÇA DA INFORMAÇÃO E DE SISTEMAS COMPUTACIONAIS (SBSEG). Natal, RN: 2018 Disponível em: <<http://portaldeconteudo.sbc.org.br/index.php/sbseg/article/view/4267>>. Acesso em: 31 dez. 2018
- ALIAGA, Y. E. M.; HENRIQUES, M. A. A. Uma comparação de mecanismos de consenso em *Blockchains*. p. 4, 2017.
- ALMEIDA, J. E. DE. A evolução histórica do conceito de contrato: em busca de um modelo democrático de contrato. **Âmbito Jurídico**, v. 99, abr. 2012.
- ANTON, V. et al. Monitoring the mammalian fauna of urban areas using remote cameras and citizen science. **Journal of Urban Ecology**, v. 4, n. 1, 1 jan. 2018.
- ANTONOPOULOS, A. M. **Mastering Bitcoin: Unlocking Digital Cryptocurrencies**. Tradução André Torres et al. [s.l.] O'Reilly Media, Inc., 2014.
- ARAÚJO, C. A. Á. Uma história intelectual da Ciência da Informação em três tempos. **RACIn**, v. 5, n. 2, p. 10–29, 15 dez. 2017.
- ARAÚJO, E. A.; OLIVEIRA, M. DE. A produção de conhecimento e a origem das bibliotecas. In: **Ciência da Informação e Biblioteconomia**. Belo Horizonte: UFMG, 2005.
- ARELLANO, M.; ÁNGEL, M. Critérios para a preservação digital da informação científica. 4 dez. 2008.
- AROLAS, E. E.; GONZÁLEZ-LADRÓN-DE-GUEVARA, F. **Towards an integrated crowdsourcing definition**. Journal of Information Science. **Anais...** In: JOURNAL OF INFORMATION SCIENCE. SAGE Publications (UK and US), abr. 2012 Disponível em: <<https://riunet.upv.es/handle/10251/56904>>. Acesso em: 4 jan. 2019
- BAKER, M. 1,500 scientists lift the lid on reproducibility. **Nature**, v. 533, n. 7604, p. 452–454, 2016.
- BARTLING, S.; ET CONTRIBUTORS TO LIVING DOCUMENT. *Blockchain for Open Science and Knowledge Creation*. 20 mar. 2017.
- BASHIR, I. **Mastering Blockchain**. Birmingham: Packt Publishing Limited, 2017.

BEGLEY, C.; ELLIS, L. Drug development: Raise standards for preclinical cancer research. **Nature**, v. 483, n. 7391, p. 531–533, 2012.

BLOCKCHAIN FOR SCIENCE. **Mission Statement** *Blockchain for Science*, 25 abr. 2017. Disponível em:

<<https://www.Blockchainforscience.com/2017/04/25/mission-statement/>>. Acesso em: 30 dez. 2018

BLOCKCHAIN FOR SCIENCE. **IBFS - International Society of Blockchain For Science**. Institucional. Disponível em:

<<https://www.Blockchainforscience.com/ibfs-international-society-Blockchain-science/>>. Acesso em: 30 dez. 2018.

BOAI FORUM. **Budapest Open Access Initiative**. Disponível em:

<<https://www.budapestopenaccessinitiative.org/read>>. Acesso em: 20 dez. 2018.

BORKO, H. Information science: What is it? **American Documentation**, v. 19, p. 3–5, 1 jan. 1968.

BRABHAM, D. C. Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. **Convergence**, v. 14, n. 1, p. 75–90, 1 fev. 2008.

BUSH, V. As we may think. **The atlantic monthly**, v. 176, n. 1, p. 101–108, 1945.

CAMERER, C. F. et al. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. **Nature Human Behaviour**, v. 2, n. 9, p. 637, set. 2018.

CASTRO, F. F. DE; SIMIONATO, A. C.; ZAFALON, Z. R. **ASPECTOS RELACIONAIS ENTRE ONTOLOGIA E METADADOS: CONSIDERAÇÕES INTERDISCIPLINARES**. XVII Encontro Nacional de Pesquisa em Ciência da Informação. **Anais...** In: XVII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO. 15 set. 2016 Disponível em:

<<http://www.ufpb.br/evento/Iti/ocs/index.php/enancib2016/enancib2016/paper/view/4205>>. Acesso em: 31 mar. 2017

CHRISTIDIS, K.; DEVETSIKIOTIS, M. *Blockchains* and Smart Contracts for the Internet of Things. **IEEE Access**, v. 4, p. 2292–2303, 2016.

COSTA, M. P. DA. Fatores que influenciam a comunicação de dados de pesquisa sobre o vírus da zika, na perspectiva de pesquisadores. 22 mar. 2017.

CRUZ, J. C. et al. **Tecnologia Blockchain: um novo paradigma nas ciências abertas**. XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (XIX ENANCIB). **Anais...** In: XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (XIX ENANCIB). 19 set. 2018 Disponível em:

<<http://enancib.marilia.unesp.br/index.php/XIXENANCIB/xixenancib/paper/view/1522>>. Acesso em: 20 maio. 2019

CUNHA, M. B. DA. Building the future: the Brazilian university library in 2010. **Ciência da Informação**, v. 29, n. 1, p. 71–89, abr. 2000.

Curadoria. Michaelis On-line. São Paulo, SP: Melhoramentos, 2018. Disponível em: <<http://michaelis.uol.com.br/busca?r=0&f=0&t=0&palavra=curadoria>>. Acesso em: 27 nov. 2018

CURTY, R. G.; AVENTURIER, P. **O paradigma da publicação de dados e suas diferentes abordagens.** XVIII Encontro Nacional de pesquisa em ciência da informação ENANCIB 2017 GT-7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação. **Anais...** Marília, Brazil: out. 2017 Disponível em: <<https://hal.archives-ouvertes.fr/hal-01637987>>. Acesso em: 19 dez. 2018

CURTY, R. G.; CERVANTES, B. M. N. Data Science: Ciência orientada a dados. **Informação & Informação**, v. 21, n. 2, p. 1–4, 20 dez. 2016.

DATA CITATION SYNTHESIS GROUP. **Joint Declaration of Data Citation Principles - FINAL.** San Diego, 30 out. 2013. Disponível em: <<https://www.force11.org/group/joint-declaration-data-citation-principles-final>>. Acesso em: 3 abr. 2017.

DIGITAL CURATION CENTRE. **DCC Curation Lifecycle Model.** Institucional. Disponível em: <<http://www.dcc.ac.uk/resources/curation-lifecycle-model>>. Acesso em: 28 dez. 2018a.

DIGITAL CURATION CENTRE. **What is digital curation?** Institucional. Disponível em: <<http://www.dcc.ac.uk/digital-curation/what-digital-curation>>. Acesso em: 28 dez. 2018b.

DUDZIAK, E. A. **Crise na Ciência ou Crise na Reprodutibilidade de Pesquisas? SIBiUSP - Sistema Integrado de Bibliotecas da Universidade de São Paulo.**, 1 out. 2018. Disponível em: <<http://www.sibi.usp.br/noticias/crise-na-ciencia-ou-crise-na-reprodutibilidade-de-pesquisas/>>. Acesso em: 6 jan. 2019

DUGAS, A. F. et al. Influenza Forecasting with Google Flu Trends. **PLOS ONE**, v. 8, n. 2, p. e56176, 14 fev. 2013.

EYNDEN, V. V. DEN et al. **Managing and sharing data: best practice for researchers.** Colchester: UK Data Archive, 2011.

FAPESP. **FAPESP :: Gestão de Dados.** Disponível em: <<http://www.fapesp.br/gestaodedados/>>. Acesso em: 7 maio. 2019.

FERNANDES, M. E. M.; RIBEIRO, C. **Curadoria de Dados na U.Porto: Identificação de práticas em diversas áreas disciplinares.** . In: 2ª CONFERÊNCIA LUSO-BRASILEIRA SOBRE ACESSO ABERTO. Rio de Janeiro: 2011 Disponível em: <<http://hdl.handle.net/10216/73435>>. Acesso em: 17 out. 2017

FERREIRA, S. M. S. P.; MARCHIORI, P. Z.; CRISTOFOLI, F. Percepção e motivação para publicar em revistas tradicionais e de acesso aberto: um estudo nas ciências da comunicação. **Comunicação & Sociedade**, v. 31, n. 52, p. 79–125, 11 jan. 2010.

FOSTER. **Introduction to Open Science.** Portal. Disponível em: <<https://www.fosteropenscience.eu/node/2076>>. Acesso em: 20 dez. 2018a.

- FOSTER. **About FOSTER**. Institucional. Disponível em: <<https://www.fosteropenscience.eu/about>>. Acesso em: 20 dez. 2018b.
- FREIRE, G. H. DE A. Ciência da Informação: temática , histórias e fundamentos. **Perspectivas em Ciência da Informação**, v. 11, n. 1, 2006.
- GARVEY, W. D. **Communication: The Essence of Science: Facilitating Information Exchange Among Librarians, Scientists, Engineers and Students**. 1. ed. Oxford: Pergamon International Library, 1979.
- GIBB, B. C. **Reproducibility**. Comments and Opinion. Disponível em: <<https://www-nature.ez31.periodicos.capes.gov.br/articles/nchem.2017>>. Acesso em: 1 nov. 2018.
- GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.
- GOODMAN, S. N.; FANELLI, D.; IOANNIDIS, J. P. A. What does research reproducibility mean? **Science Translational Medicine**, v. 8, n. 341, p. 341ps12-341ps12, 1 jun. 2016.
- GRAY, J. Jim Gray on eScience: A Transformed Scientific Method. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Eds.). . **The Fourth Paradigm: Data-Intensive Scientific Discovery**. Richmond, WA: Microsoft Research, 2009.
- GREEN, A.; MACDONALD, S.; RICE, R. **Policy-making for Research Data in Repositories: A Guide**. Edinburgh, UK: EDINA and University Data Library, maio 2009. Disponível em: <<https://www.coar-repositories.org/files/guide.pdf>>. Acesso em: 26 dez. 2018.
- GREICIUS, T. **Multi-planet System Found Through Crowdsourcing**. Text. Disponível em: <<http://www.nasa.gov/feature/jpl/multi-planet-system-found-through-crowdsourcing>>. Acesso em: 3 jan. 2019.
- HARVEY, D. R. **Digital curation: a how-to-do-it manual**. New York: Neal-Schuman Publishers, 2010.
- HEIM, K. M. Data Libraries for the Social Sciences. **Library Trends**, v. 30, n. 3, Winter 1982.
- HEWITT, S. M. Reproducibility: It Is Just Good Science. **Journal of Histochemistry & Cytochemistry**, v. 64, n. 4, p. 223–223, 1 abr. 2016.
- HOGEMANN, E. R. O futuro do Direito e do ensino jurídico diante das novas tecnologias. **Revista Interdisciplinar de Direito**, v. 16, n. 1, p. 105–115, 20 jun. 2018.
- HOWE, J. **The Rise of Crowdsourcing**. Blog. Disponível em: <<https://www.wired.com/2006/06/crowds/>>. Acesso em: 4 jan. 2019.
- HOWE, J.; ARAUJO, A. M. **Poder Das Multidoes, O**. [s.l.] Elsevier Brasil, 2009.

- INGRAM, C. **How and why you should manage your research data: a guide for researchers**JISC, , 7 jan. 2016. Disponível em: <<https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data>>. Acesso em: 20 maio. 2019
- KANSA, E. C.; KANSA, S. W.; ARBUCKLE, B. Publishing and Pushing: Mixing Models for Communicating Research Data in Archaeology. **International Journal of Digital Curation**, v. 9, n. 1, p. 57–70, 30 maio 2014.
- KOSTAREV, G. **Review of Blockchain consensus mechanisms**. Disponível em: <<https://blog.wavesplatform.com/review-of-Blockchain-consensus-mechanisms-f575afae38f2>>. Acesso em: 7 jan. 2019.
- LAWRENCE, B. et al. Citation and Peer Review of Data: Moving Towards Formal Data Publication. **International Journal of Digital Curation**, v. 6, n. 2, p. 4–37, 26 jul. 2011.
- LE COADIC, Y. **A Ciência da Informação**. Brasília: Briquet de Lemos Livros, 1996.
- LEITE, G. A evolução doutrinária do contrato. **Âmbito Jurídico**, v. 45, set. 2007.
- MARQUES, F. Ciência transparente. **Revista Pesquisa Fapesp**, n. 218, abr. 2014.
- MATOS, M. **Como funciona o Proof of Work na Blockchain do BitcoinLivecoins**, 11 abr. 2018. Disponível em: <<https://livecoins.com.br/proof-of-work-Blockchain-bitcoin/>>. Acesso em: 2 jan. 2019
- MAYER, F.; ZEVIANI, W. **Pesquisa Reproduzível**. Disponível em: <<http://cursos.leg.ufpr.br/prr/capPesqRep.html>>. Acesso em: 27 maio. 2018.
- MCNUTT, M. Reproducibility. **Science**, v. 343, n. 6168, p. 229–229, 17 jan. 2014.
- MEADOWS, A. J. **A comunicação científica**. [s.l.] Briquet de Lemos/livros, 1999.
- MEY, E. S. A. Bibliotheca Alexandrina. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 2, n. 1, p. 71–91, 13 dez. 2004.
- MOLLOY, J. C. The Open Knowledge Foundation: open data means better science. **PLoS Biology**, v. 9, n. 12, p. e1001195, 6 dez. 2011.
- MONTEIRO, E. C. DE S. DE A.; SANT'ANA, R. C. G.; SANTAREM SEGUNDO, J. E. e-Science Semântica: integração dos dados na comunicação científica. **Informação em Pauta**, v. 1, n. 1, p. 9–29, 30 jun. 2016.
- MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, v. 35, n. 2, 2006.
- MYCELIA. **Mycelia for music**. Disponível em: <<http://myceliaformusic.org/>>. Acesso em: 10 set. 2018.

NASSI-CALÒ, L. **Avaliação sobre a reprodutibilidade de resultados de pesquisa traz mais perguntas que respostas.** Blog. Disponível em: <<https://blog.scielo.org/blog/2017/02/08/avaliacao-sobre-a-reprodutibilidade-de-resultados-de-pesquisa-traz-mais-perguntas-que-respostas/#.W-1lFOhKjDc>>. Acesso em: 15 nov. 2018.

NEWTON, I. **Isaac Newton letter to Robert Hooke, 1675**, 1675. Disponível em: <<https://digitallibrary.hsp.org/index.php/Detail/objects/9792>>. Acesso em: 7 jul. 2018

NOSEK, B. A.; ERRINGTON, T. M. Making sense of replications. **eLife**, v. 6, p. e23383, 19 jan. 2017.

OLIVEIRA, J. A. M. M. DE; SANTARÉM SEGUNDO, J. E. **A possibilidade de identificação de violações a direitos autorais com base em metadados gerados na Blockchain: avaliação da plataforma original.my.** . In: XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2018. Londrina: Universidade Estadual de Londrina, out. 2018 Disponível em: <<http://www.brapci.inf.br/index.php/res/v/103108>>. Acesso em: 20 maio. 2019

OPEN KNOWLEDGE INTERNATIONAL. **O que são dados abertos?** Guia. Disponível em: <<http://opendatahandbook.org/guide/en/what-is-open-data/>>. Acesso em: 23 dez. 2018.

OPEN SCIENCE COLLABORATION. Estimating the reproducibility of psychological science. **Science**, v. 349, n. 6251, p. aac4716, 28 ago. 2015.

OUCHI, M. T.; SIMIONATO, A. C. Descrição de conjuntos de dados na Web com schema.org. **Informação & Tecnologia**, v. 5, n. 1, p. 128–140, 2018.

PATIL, P.; PENG, R. D.; LEEK, J. A statistical definition for reproducibility and replicability. **bioRxiv**, p. 066803, 29 jul. 2016.

PENG, R. D. Reproducible Research in Computational Science. **Science**, v. 334, n. 6060, p. 1226–1227, 2 dez. 2011.

PINHEIRO, L. V. R. Do acesso livre à ciência aberta: conceitos e implicações na comunicação científica. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 8, n. 2, 30 jun. 2014.

PLESSER, H. E. Reproducibility vs. Replicability: A Brief History of a Confused Terminology. **Frontiers in Neuroinformatics**, v. 11, 18 jan. 2018.

PONTIKA, N. et al. **Fostering open science to research using a taxonomy and an eLearning portal.** Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business - i-KNOW '15. **Anais...** In: THE 15TH INTERNATIONAL CONFERENCE. Graz, Austria: ACM Press, 2015 Disponível em: <<http://dl.acm.org/citation.cfm?doid=2809563.2809571>>. Acesso em: 20 dez. 2018

PRICE, D. DE S. (ED.). **A ciência desde a Babilônia**. Belo Horizonte; São Paulo (SP): Itatiaia : Ed. Univ. S. Paulo, 1976.

QIN, J. et al. Linking Entities in Scientific Metadata. **International Conference on Dublin Core and Metadata Applications**, v. 0, n. 0, p. 128–136, 20 set. 2010.

RIBEIRO, C. et al. Os Repositórios da Dados Científicos: Estado da Arte. 2010.

RIN; NESTA. **Open to All?: Case studies of openness in research**. London: RIN/NESTA, set. 2010. Disponível em:

<http://www.rin.ac.uk/system/files/attachments/NESTA-RIN_Open_Science_V01_0.pdf>. Acesso em: 20 dez. 2018.

ROSA, F. G; GOMES, M. J.. Comunicação Científica: das restrições ao acesso livre. 2010.

ROUSE, M. **What is Blockchain?** Disponível em: <<https://searchcio.techtarget.com/definition/Blockchain>>. Acesso em: 10 set. 2018.

ROYAL NETHERLANDS ACADEMY OF ARTS AND SCIENCES. **Replication Studies: Improving reproducibility in the empirical sciences**. Amsterdam: Royal Netherlands Academy of Arts and Sciences (KNAW), 2018.

RUSSO, M. **Fundamentos de Biblioteconomia e Ciência da Informação**. [s.l.] Editora E-papers, 2010.

SALES, L. F.; CAVALCANTI, M. T. SELEÇÃO E AVALIAÇÃO DE COLEÇÕES DE DADOS DIGITAIS DE PESQUISA: uma possível abordagem metodológica. **Informação & Tecnologia**, v. 2, n. 2, p. 88–105, 2015.

SALUSTIANO, R. E. Aplicação de técnicas de fusão de sensores no monitoramento de ambientes. 2006.

SANTANA, R. C. G. Ciclo de vida dos dados e o papel da ciência da informação. 24 jan. 2014.

SANTOS, P. L. V. A. DA C.; SANTANA, R. C. G. Dado e Granularidade na perspectiva da Informação e Tecnologia: uma interpretação pela Ciência da Informação. **Ciência da Informação**, v. 42, n. 2, p. 199–209, 2013.

SANTOS, P. L. V. A. DA C.; SIMIONATO, A. C.; ARAKAKI, F. A. Definição de metadados para recursos informacionais: apresentação da metodologia BEAM. **Informação & Informação**, v. 19, n. 1, p. 146, 25 fev. 2014.

SANTOS, P. L. V. A. DA C.; VIDOTTI, S. A. B. G. Perspectivismo e tecnologias de informação e comunicação: acréscimos à Ciência da Informação? **DataGramZero: revista de Ciência da Informação**, n. 3. v. v. 10, 2009.

SANTOS, T. N. C. Curadoria digital e preservação digital: cruzamentos conceituais. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 14, n. 3, p. 450–464, 30 set. 2016.

SAYÃO, L. F.; SALES, L. F. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 8, n. 2, 2013.

SAYÃO, L. F.; SALES, L. F. Dados abertos de pesquisa: ampliando o conceito de acesso livre. 2014.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, v. 21, n. 2, p. 90, 20 dez. 2016.

SAYÃO, L. F.; SALES, L. F. Curadoria digital e dados de pesquisa. **AtoZ: novas práticas em informação e conhecimento**, v. 5, n. 2, p. 67–71, 9 jan. 2017.

SCHNEIER, B. **Cryptanalysis of MD5 and SHA: Time for a New Standard - Schneier on Security**Schneier on Security, 19 ago. 2004. Disponível em: <https://www.schneier.com/essays/archives/2004/08/cryptanalysis_of_md5.html>. Acesso em: 20 maio. 2019

SIMMONS, J. P.; NELSON, L. D.; SIMONSOHN, U. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. **Psychological Science**, v. 22, n. 11, p. 1359–1366, 1 nov. 2011.

SONG, J. **The Truth about Smart Contracts**Jimmy Song, 11 jun. 2018. Disponível em: <<https://medium.com/@jimmysong/the-truth-about-smart-contracts-ae825271811f>>. Acesso em: 21 maio. 2019

STEVENS, M. M. J. On Collisions for MD5. **International Journal of Human-computer Studies / International Journal of Man-machine Studies - IJMMS**, 1 jan. 2007.

STOLFI, J. **Shows a typical cryptographic hash function (SHA-1) at work. Note that small differences in the input result in very different digests.**, 27 nov. 2008. Disponível em: <https://commons.wikimedia.org/wiki/File:Cryptographic_Hash_Function.svg>. Acesso em: 20 maio. 2019

SVENONIUS, E. **The intellectual foundation of information organization**. [s.l.] MIT Press, 2008.

SZABO, N. **Smart Contracts**. Disponível em: <<http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart.contracts.html>>. Acesso em: 21 maio. 2019.

SZABO, N. Formalizing and Securing Relationships on Public Networks. **First Monday**, v. 2, n. 9, 1 set. 1997.

TENOPIR, C. et al. Data Sharing by Scientists: Practices and Perceptions. **PLOS ONE**, v. 6, n. 6, p. e21101, 29 jun. 2011.

ULIERU, M. **Blockchain Enhances Privacy, Security and Conveyance of Data**. Disponível em: <<https://www.scientificamerican.com/article/Blockchain-enhances-privacy-security-and-conveyance-of-data/>>. Acesso em: 29 dez. 2018.

WEDEMANN, K. **O que a Inteligência Artificial significa para seu negócio**. Corporativo. Disponível em: <https://www.sas.com/pt_br/insights/articles/analytics/o-que-inteligencia-artificial-significa-para-seu-negocio.html>. Acesso em: 17 dez. 2018.

WILKINSON, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, p. 160018, 15 mar. 2016.

YLI-HUUMO, J. et al. Where Is Current Research on *Blockchain* Technology?—A Systematic Review. **PLOS ONE**, v. 11, n. 10, p. e0163477, 3 out. 2016.

ZANONA, R. C.; PETEROSI, H. G.; BRANDÃO, J. A. **Crowdsourcing: cenário, aplicação e seus desdobramentos**. . In: VIII WORKSHOP DE PÓS-GRADUAÇÃO E PESQUISA DO CENTRO PAULA SOUZA. São Paulo: Centro Paula Souza, 2013

ZIMAN, J. M. (JOHN M.), 1925-. **Conhecimento publico**. Belo Horizonte; São Paulo: Itatiaia; Ed. Univ. São Paulo, 1979.