

SÉRIE TECNOLOGIA E ORGANIZAÇÃO DA INFORMAÇÃO

# CURADORIA DIGITAL

Contribuições para a Ciência da Informação

Francisco Carlos Paletta  
Audilio Gonzalez Aguilar  
(Organizadores)

**Atena**  
Editora  
Ano 2023

SÉRIE TECNOLOGIA E ORGANIZAÇÃO DA INFORMAÇÃO

# CURADORIA DIGITAL

Contribuições para a Ciência da Informação

Francisco Carlos Paletta  
Audilio Gonzalez Aguilar  
(Organizadores)

  
Atena  
Editora  
Ano 2023

**Editora chefe**

Profª Drª Antonella Carvalho de Oliveira

**Editora executiva**

Natalia Oliveira

**Assistente editorial**

Flávia Roberta Barão

**Bibliotecária**

Janaina Ramos

**Projeto gráfico**

Camila Alves de Cremo

Ellen Andressa Kubisty

Luiza Alves Batista

Nataly Evilin Gayde

**Imagens da capa**

iStock

**Edição de arte**

Luiza Alves Batista

2023 by Atena Editora

Copyright © Atena Editora

Copyright do texto © 2023 Os autores

Copyright da edição © 2023 Atena

Editora

Direitos para esta edição cedidos à Atena Editora pelos autores.

Open access publication by Atena

Editora



Todo o conteúdo deste livro está licenciado sob uma Licença de Atribuição *Creative Commons*. Atribuição-Não-Comercial-NãoDerivativos 4.0 Internacional (CC BY-NC-ND 4.0).

O conteúdo dos artigos e seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva dos autores, inclusive não representam necessariamente a posição oficial da Atena Editora. Permitido o *download* da obra e o compartilhamento desde que sejam atribuídos créditos aos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

Todos os manuscritos foram previamente submetidos à avaliação cega pelos pares, membros do Conselho Editorial desta Editora, tendo sido aprovados para a publicação com base em critérios de neutralidade e imparcialidade acadêmica.

A Atena Editora é comprometida em garantir a integridade editorial em todas as etapas do processo de publicação, evitando plágio, dados ou resultados fraudulentos e impedindo que interesses financeiros comprometam os padrões éticos da publicação. Situações suspeitas de má conduta científica serão investigadas sob o mais alto padrão de rigor acadêmico e ético.

**Conselho Editorial**

**Ciências Humanas e Sociais Aplicadas**

Prof. Dr. Adilson Tadeu Basquerote Silva – Universidade para o Desenvolvimento do Alto Vale do Itajaí

Prof. Dr. Alexandre de Freitas Carneiro – Universidade Federal de Rondônia

Prof. Dr. Alexandre Jose Schumacher – Instituto Federal de Educação, Ciência e Tecnologia do Paraná

Prof. Dr. Américo Junior Nunes da Silva – Universidade do Estado da Bahia

Profª Drª Ana Maria Aguiar Frias – Universidade de Évora

Profª Drª Andréa Cristina Marques de Araújo – Universidade Fernando Pessoa  
 Prof. Dr. Antonio Carlos da Silva – Universidade de Coimbra  
 Prof. Dr. Antonio Carlos Frasson – Universidade Tecnológica Federal do Paraná  
 Prof. Dr. Antonio Gasparetto Júnior – Instituto Federal do Sudeste de Minas Gerais  
 Prof. Dr. Antonio Isidro-Filho – Universidade de Brasília  
 Prof. Dr. Arnaldo Oliveira Souza Júnior – Universidade Federal do Piauí  
 Prof. Dr. Carlos Antonio de Souza Moraes – Universidade Federal Fluminense  
 Profª Drª Caroline Mari de Oliveira Galina – Universidade do Estado de Mato Grosso  
 Prof. Dr. Crisóstomo Lima do Nascimento – Universidade Federal Fluminense  
 Profª Drª Cristina Gaio – Universidade de Lisboa  
 Prof. Dr. Daniel Richard Sant’Ana – Universidade de Brasília  
 Prof. Dr. Deyvison de Lima Oliveira – Universidade Federal de Rondônia  
 Profª Drª Dilma Antunes Silva – Universidade Federal de São Paulo  
 Prof. Dr. Edvaldo Antunes de Farias – Universidade Estácio de Sá  
 Prof. Dr. Elson Ferreira Costa – Universidade do Estado do Pará  
 Prof. Dr. Eloi Martins Senhora – Universidade Federal de Roraima  
 Profª Drª Geuciane Felipe Guerim Fernandes – Universidade Estadual de Londrina  
 Prof. Dr. Gustavo Henrique Cepolini Ferreira – Universidade Estadual de Montes Claros  
 Prof. Dr. Humberto Costa – Universidade Federal do Paraná  
 Profª Drª Ivone Goulart Lopes – Istituto Internazionale delle Figlie de Maria Ausiliatrice  
 Prof. Dr. Jadilson Marinho da Silva – Secretaria de Educação de Pernambuco  
 Prof. Dr. Jadson Correia de Oliveira – Universidade Católica do Salvador  
 Prof. Dr. Jodeyson Islony de Lima Sobrinho – Universidade Estadual do Oeste do Paraná  
 Prof. Dr. José Luis Montesillo-Cedillo – Universidad Autónoma del Estado de México  
 Profª Drª Juliana Abonizio – Universidade Federal de Mato Grosso  
 Prof. Dr. Julio Candido de Meirelles Junior – Universidade Federal Fluminense  
 Prof. Dr. Kárpio Márcio de Siqueira – Universidade do Estado da Bahia  
 Profª Drª Kátia Farias Antero – Faculdade Maurício de Nassau  
 Profª Drª Keyla Christina Almeida Portela – Instituto Federal do Paraná  
 Profª Drª Lina Maria Gonçalves – Universidade Federal do Tocantins  
 Profª Drª Lucicleia Barreto Queiroz – Universidade Federal do Acre  
 Prof. Dr. Luis Ricardo Fernandes da Costa – Universidade Estadual de Montes Claros  
 Prof. Dr. Lucio Marques Vieira Souza – Universidade do Estado de Minas Gerais  
 Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte  
 Profª Drª Marianne Sousa Barbosa – Universidade Federal de Campina Grande  
 Profª Drª Marcela Mary José da Silva – Universidade Federal do Recôncavo da Bahia  
 Prof. Dr. Marcelo Pereira da Silva – Pontifícia Universidade Católica de Campinas  
 Prof. Dr. Marcelo Pereira da Silva – Pontifícia Universidade Católica de Campinas  
 Profª Drª Maria Luzia da Silva Santana – Universidade Federal de Mato Grosso do Sul  
 Prof. Dr. Miguel Rodrigues Netto – Universidade do Estado de Mato Grosso  
 Prof. Dr. Pedro Henrique Máximo Pereira – Universidade Estadual de Goiás  
 Prof. Dr. Pablo Ricardo de Lima Falcão – Universidade de Pernambuco  
 Profª Drª Paola Andressa Scortegagna – Universidade Estadual de Ponta Grossa  
 Profª Drª Rita de Cássia da Silva Oliveira – Universidade Estadual de Ponta Grossa  
 Prof. Dr. Rui Maia Diamantino – Universidade Salvador

Profª Drª Sandra Regina Gardacho Pietrobon – Universidade Estadual do Centro-Oeste

Prof. Dr. Saulo Cerqueira de Aguiar Soares – Universidade Federal do Piauí

Prof. Dr. Urandi João Rodrigues Junior – Universidade Federal do Oeste do Pará

Profª Drª Vanessa Bordin Viera – Universidade Federal de Campina Grande

Profª Drª Vanessa Ribeiro Simon Cavalcanti – Universidade Federal da Bahia /  
Universidade de Coimbra

Prof. Dr. William Cleber Domingues Silva – Universidade Federal Rural do Rio de Janeiro

Prof. Dr. Willian Douglas Guilherme – Universidade Federal do Tocantins

Série tecnologia e organização da informação –  
Curadoria digital: contribuições para a ciência da informação

**Diagramação:** Ellen Andressa Kubisty  
**Correção:** Maiara Ferreira  
**Indexação:** Amanda Kelly da Costa Veiga  
**Revisão:** Os autores  
**Organizadores:** Francisco Carlos Paletta  
 Audilio Gonzalez Aguilar

<b>Dados Internacionais de Catalogação na Publicação (CIP)</b>	
S485	<p>Série tecnologia e organização da informação - Curadoria digital: contribuições para a ciência da informação / Organizadores Francisco Carlos Paletta, Audilio Gonzalez Aguilar. – Ponta Grossa - PR: Atena, 2023.</p> <p>Formato: PDF                      Requisitos de sistema: Adobe Acrobat Reader                      Modo de acesso: World Wide Web                      Inclui bibliografia                      ISBN 978-65-258-1444-5                      DOI: <a href="https://doi.org/10.22533/at.ed.445233005">https://doi.org/10.22533/at.ed.445233005</a></p> <p>1. Ciência da informação. I. Paletta, Francisco Carlos (Organizador). II. Aguilar, Audilio Gonzalez (Organizador). III. Título.</p> <p style="text-align: right;">CDD 020</p>
<b>Elaborado por Bibliotecária Janaina Ramos – CRB-8/9166</b>	

**Atena Editora**  
 Ponta Grossa – Paraná – Brasil  
 Telefone: +55 (42) 3323-5493  
[www.atenaeditora.com.br](http://www.atenaeditora.com.br)  
[contato@atenaeditora.com.br](mailto:contato@atenaeditora.com.br)

## DECLARAÇÃO DOS AUTORES

Os autores desta obra: 1. Atestam não possuir qualquer interesse comercial que constitua um conflito de interesses em relação ao artigo científico publicado; 2. Declaram que participaram ativamente da construção dos respectivos manuscritos, preferencialmente na: a) Concepção do estudo, e/ou aquisição de dados, e/ou análise e interpretação de dados; b) Elaboração do artigo ou revisão com vistas a tornar o material intelectualmente relevante; c) Aprovação final do manuscrito para submissão.; 3. Certificam que os artigos científicos publicados estão completamente isentos de dados e/ou resultados fraudulentos; 4. Confirmam a citação e a referência correta de todos os dados e de interpretações de dados de outras pesquisas; 5. Reconhecem terem informado todas as fontes de financiamento recebidas para a consecução da pesquisa; 6. Autorizam a edição da obra, que incluem os registros de ficha catalográfica, ISBN, DOI e demais indexadores, projeto visual e criação de capa, diagramação de miolo, assim como lançamento e divulgação da mesma conforme critérios da Atena Editora.

## DECLARAÇÃO DA EDITORA

A Atena Editora declara, para os devidos fins de direito, que: 1. A presente publicação constitui apenas transferência temporária dos direitos autorais, direito sobre a publicação, inclusive não constitui responsabilidade solidária na criação dos manuscritos publicados, nos termos previstos na Lei sobre direitos autorais (Lei 9610/98), no art. 184 do Código Penal e no art. 927 do Código Civil; 2. Autoriza e incentiva os autores a assinarem contratos com repositórios institucionais, com fins exclusivos de divulgação da obra, desde que com o devido reconhecimento de autoria e edição e sem qualquer finalidade comercial; 3. Todos os e-book são *open access*, *desta forma* não os comercializa em seu site, sites parceiros, plataformas de *e-commerce*, ou qualquer outro meio virtual ou físico, portanto, está isenta de repasses de direitos autorais aos autores; 4. Todos os membros do conselho editorial são doutores e vinculados a instituições de ensino superior públicas, conforme recomendação da CAPES para obtenção do Qualis livro; 5. Não cede, comercializa ou autoriza a utilização dos nomes e e-mails dos autores, bem como nenhum outro dado dos mesmos, para qualquer finalidade que não o escopo da divulgação desta obra.

**TECNOLOGIA E ORGANIZAÇÃO DA INFORMAÇÃO ..... 1**

Francisco Carlos Paletta

 <https://doi.org/10.22533/at.ed.4452330051>**A CONSTRUÇÃO DE REPOSITÓRIOS DIGITAIS COMO FERRAMENTAS DE PRESERVAÇÃO E APROPRIAÇÃO DAS INFORMAÇÕES DE SAÚDE.....4**

Hélio Salomão Silva Cordoeira

 <https://doi.org/10.22533/at.ed.4452330052>**A CURADORIA DIGITAL NAS REDES SOCIAIS: O TWITTER DO PRESIDENTE AMERICANO .....24**

Rosália Maria Silva Oliveira

 <https://doi.org/10.22533/at.ed.4452330053>**CURADORIA DIGITAL: MAPA DE BIBLIOTECAS NAS UNIVERSIDADES FEDERAIS BRASILEIRAS..... 41**

Priscila Ramos Carvalho

Marcos Gonçalves Ramos

Ricardo Medeiros Pimenta

 <https://doi.org/10.22533/at.ed.4452330054>**GESTÃO DA INFORMAÇÃO: DESAFIOS E COMPETÊNCIAS PARA PRESERVAÇÃO DIGITAL..... 61**

Angélica Cintra Fermann

Francisco Carlos Paletta

 <https://doi.org/10.22533/at.ed.4452330055>**SPRESERVAÇÃO DIGITAL: CONSIDERAÇÕES SOBRE ARQUIVAMENTO DE CONTEÚDO DA WEB .....85**

Antonio Paulo Carretta

 <https://doi.org/10.22533/at.ed.4452330056>**SOBRE OS ORGANIZADORES .....97**

# TECNOLOGIA E ORGANIZAÇÃO DA INFORMAÇÃO

O TOI – VI Congresso Internacional em Tecnologia e Organização da Informação – Contribuições para a Ciência da Informação - é iniciativa do Grupo de Pesquisa “Observatório do Mercado de Trabalho em Informação e Documentação (OMTID) - CNPq” da Escola de Comunicações e Artes ECA\_USP.

## Programação VI TOI 2020 Online – Library and Information Science

PROGRAMAÇÃO	17 de junho	18 de junho	07 de julho	08 de julho	15 de julho	04 de agosto
09h00 às 12h30	II Simpósio Inteligência Artificial e Ciência de Dados	II Simpósio Inteligência Artificial e Ciência de Dados	I Simpósio Economia Comportamental e Tomada de Decisão	I Simpósio Economia Comportamental e Tomada de Decisão	V Simpósio Gestão do Conhecimento e Inteligência Competitiva	VI Simpósio Preservação e Conservação de Arquivo
PROGRAMAÇÃO	11 de agosto	13 de agosto	18 de agosto	23 de setembro	08 de outubro	20 de outubro
09h00 às 12h30	I Simpósio Políticas e Estratégias de Preservação de Conteúdo na Web	I Simpósio userExperience e Arquitetura da Informação	III Simpósio Curadoria Digital		I Simpósio de Humanidades Digitais	I Simpósio Sustentabilidade Meio Ambiente e Políticas Públicas
16h00 às 18h00			Submissão de Trabalhos Curadoria Digital		Submissão de Trabalhos Humanidades Digitais	
19h00 às 21h00				V Colóquio de Informação e Documentação Jurídica		
PROGRAMAÇÃO	03 de novembro	02 de dezembro	Evento Satélite	Evento Satélite	Evento Satélite	Evento Satélite
09h00 às 13h00	I Simpósio Informação, Ciência, Tecnologia e Inovação		Círculo Iberoamericano de Ciência da Informação Documental (23/10) (CIBERICID)	Recursos Emerald 03/11/2020	SophA Painel: Protocolos de Segurança Pós-COVID-19 nas Bibliotecas 24/11/2020	Ischool Day South America Pré-TOI
18h00 às 21h00	Submissão de Trabalho I Simpósio Informação, Ciência, Tecnologia e Inovação	II Encontro Observatório Mercado de Trabalho Profissional da Informação		-	-	-
PROGRAMAÇÃO	04 de março 2021	11 de março 2021	15 de março 2021	18 de março 2021	25 de março 2021	15 de abril 2021
09h00 às 12h30	Apresentação de Trabalhos I Simpósio userXperience	Apresentação de Trabalhos III Simpósio Curadoria Digital	I Simpósio Organização e Representação da Informação e do Conhecimento	Apresentação de Trabalhos I Simpósio Informação, Ciência, Tecnologia e Inovação	Apresentação de Trabalhos III Simpósio Humanidades Digitais	II Simpósio Cidades Inteligentes e IOT
16h00 às 18h00			Submissão I Simpósio Organização e Representação da Informação e do Conhecimento			



**OMTID**

OBSERVATÓRIO DO MERCADO DE TRABALHO DO PROFISSIONAL DA INFORMAÇÃO E DOCUMENTAÇÃO

Evento Online Gratuito com Certificado  
 LOCAL: TOI 2020 ONLINE – Canal YouTube  
<http://toi.eca.usp.br/>



Destacamos que o 6º Congresso Internacional em Tecnologia e Organização da Informação em sua missão apresenta aderência aos objetivos de ensino, pesquisa e extensão universitária nos seguintes eixos temáticos:

- Ciência de Dados e Inteligência Artificial
- Internet das Coisas e Cidades Inteligentes
- Curadoria Digital
- Tecnologia e Inovação em Serviços
- Informação, Ciência e Tecnologia
- Humanidades Digitais
- Gestão do Conhecimento e Inteligência Competitiva
- Preservação e Conservação de Acervo

O VI Congresso Internacional em Tecnologia e Organização da Informação tem como principal foco reunir pesquisadores, estudantes e profissionais da área da Ciência da Informação: Biblioteconomia, Arquivologia, Museologia, e afins, com objetivo de promover a reflexão e o diálogo em torno de temas relevantes para a sociedade da informação: Gestão da Informação e do Conhecimento, Tecnologia e Sistemas de Informação, Biblioteca Digital, Conservação e Preservação de Acervo, Curadoria Digital, Ética da Informação, Humanidades Digitais, Recuperação da Informação na Web de Dados, Organização da Informação e do Conhecimento, Ciência de Dados, Inteligência Artificial, Internet das Coisas IoT, Cultura Computacional, Mercado de Trabalho do Profissional da Informação.

Entre os principais resultados espera-se promover a discussão sobre os temas centrais do Congresso e contribuir para a integração entre o meio acadêmico e as comunidades profissionais, ampliando o interesse pela pesquisa, pelo compartilhamento de informação, e sobre as práticas mais inovadoras no campo da Ciência da Informação.

No contexto da Ciência da Informação, Biblioteconomia, Museologia, Arquivologia e áreas afins, os avanços impostos pela tecnologia da informação e das comunicações têm proporcionado avanços significativos nas melhores práticas de gestão, otimizar o processo de tomada de decisão e focar nas demandas por busca, acesso, recuperação, apropriação e uso da informação na Era Digital. A gestão e o suporte destes ambientes heterogêneos e complexos, repletos de recursos e dispositivos computacionais, comprovadamente têm se revelado de gestão complexa e fortemente impactado pelo fenômeno do Big Data, pela Ciência de Dados e pela Inteligência Artificial.

Neste cenário torna-se relevante avaliar a complexidade os principais desafios da Era Digital e das Transformações impostas aos processos de organização e gestão da informação e do conhecimento. A crescente disponibilização das tecnologias e o fluxo cada vez maior e mais rápido do acesso à informação têm desafiado os Profissionais da Informação no entendimento de novas práticas de apropriação e uso da informação na produção de conhecimento e criação de vantagem competitiva organizacional.

Aumentar a produtividade e construir ambientes ágeis que permitam às Organizações responder às demandas da gestão da informação é fator fundamental para a construção de

um portfólio adequado de Serviços de Informação que proporcione atender as demandas do Usuário cada vez mais conectado e independente no uso de ferramentas Web no entorno digital.

No âmbito do VI Congresso Internacional em Tecnologia e Organização da Informação - TOI, ocorre a sexta edição do SBEI - Simpósio Brasileiro de Ética da Informação, organizado pela Profa. Isa Maria Freire, UFPB e Prof. Francisco Carlos Paletta, USP.

O VI Congresso Internacional em Tecnologia e Organização da Informação colabora acadêmica e cientificamente com o CIC.DIGITAL PORTO Center for Research in Communication Information and Digital Culture Universidade do Porto, com o CITGEM – Centro de Investigação Transdisciplinar Cultura, Espaço e Memória - unidade de I&D sediada na Faculdade de Letras da Universidade do Porto e o Laboratório de Tecnologias Intelectuais LT<sup>i</sup> da UFPB; e tem por missão promover e desenvolver a troca de experiências profissionais, apresentar melhores práticas na gestão da informação e apresentar resultados de pesquisas.

Nesta sexta edição o TOI consolida-se como um Fórum permanente na Universidade de São Paulo em promover e estabelecer instrumentos para a reflexão e compreensão de temas induzidos pela rápida transformação tecnológica no ambiente digital e as novas formas de busca, acesso, organização, apropriação, uso e preservação da informação na Era Digital.

**Prof. Dr. Francisco Carlos Paletta**

*Escola de Comunicações e Artes*

*Universidade de São Paulo*

**Conference Chair - VI TOI 2020**

**Junho 2020 - Abril 2021**

# A CONSTRUÇÃO DE REPOSITÓRIOS DIGITAIS COMO FERRAMENTAS DE PRESERVAÇÃO E APROPRIAÇÃO DAS INFORMAÇÕES DE SAÚDE

Hélio Salomão Silva Cordoeira

ECA-USP

**RESUMO:** A pandemia do covid-19 demonstrou a capacidade catalisadora e disruptiva que a oferta de dados, informações e conhecimento em saúde tem tanto para o sistema de saúde em si quanto para a academia e o mercado. Construir políticas públicas cada vez mais engajadas em preservar informações em saúde através da criação e manutenção de repositórios digitais de informação em saúde se mostra uma solução cada vez mais tecnicamente viável e eficiente.

**PALAVRAS-CHAVE:** Informações em saúde, Repositórios digitais, preservação.

## THE CONSTRUCTION OF DIGITAL REPOSITORIES AS TOOLS FOR THE PRESERVATION AND APPROPRIATION OF HEALTH INFORMATION

**ABSTRACT:** The covid-19 pandemic demonstrated the catalytic and disruptive capacity that the provision of health data, information and knowledge has both for the health system itself and for the academy and the market. Building public policies increasingly engaged in preserving health information through the creation and maintenance of digital health information repositories proves to be an increasingly technically feasible and efficient solution.

**KEY WORDS:** Health information, Digital repositories, preservation.

## INTRODUÇÃO

A preservação de documentos digitais se mostra desafiadora tanto pela velocidade de produção, quanto pelas inseguranças que este suporte traz consigo. Hoje bilhões de terabytes de informação são produzidos em um ritmo nunca experimentado pela humanidade. Produzimos e armazenamos informação em escala e velocidades muito superiores a qualquer outro tempo, contudo o fazemos ainda sem garantias de que futuras gerações terão acesso a essas informações. É possível que uma parte considerável da história do início do século XXI se perca nos próximos anos.

Não estamos falando só de tweets, blogs e sites genéricos. Estamos falando de registros médicos, parte da cultura, registros de ações do estado. Tudo corre um grande risco de não existir ou existir sem uma forma viável de acesso. Um grande marco na preservação das informações digitais foi a perda dos registros das sondas Viking I e Viking II, esta perda fora o estopim para a construção do modelo OAIS e para o crescimento sobre

o Tema em todo mundo.

No campo da saúde a informação desempenha papel fundamental, médicos e profissionais da saúde estão cotidianamente tentando coletar informações processáveis e tomar decisões. Assim, no final do processo, quando da cura do paciente, aquele atendimento produziu e coletou um largo conjunto de informações e conhecimentos novos que podem ser reutilizados no atendimento de outros pacientes. O registro desse processo é útil não somente como história e memória, mas é capaz de alavancar o próprio processo de tomada de decisão em nível tático (Diagnostico) ou estratégico (Gestão, tanto da unidade quanto do sistema de saúde nos três níveis federativos).

Este artigo assim busca demonstrar como a construção de repositórios digitais como parte de uma política de preservação que garanta a comunicação através do tempo (Ferreira 2006), o aumento dos níveis de transparência e accountability, a apropriação das informações de saúde bem como possibilita a produção de produtos de informação que constituam melhorias de processos tornando-os mais eficientes, rápidos e seguros.

Durante a pandemia da Covid-19 foram construídas algumas soluções informacionais para a comunicação do andamento da pandemia para população, busca de criar ou melhorar os tratamentos para doença e no subsídio da pesquisa científica. Uma dessas soluções é o repositório DataSharing/BR ainda, fora profícua a produção de programas que utilizando inteligência artificial buscavam auxiliar os esforços do sistema de saúde na mitigação da pandemia.

Este trabalho busca juntar-se a outros na defesa de uma política de preservação de documentos de saúde que mais do que criar arquivos em unidades de saúde constitua um sistema capaz de dar conta das necessidades informacionais do Sistema de saúde Brasileiro. A construção de repositórios digitais com dados ou informações sobre saúde poderiam auxiliar tanto a tomada de decisões, quanto novas investigações. Impactar tanto a produção científica, quanto a gestão das unidades de saúde em nível local ou estratégico.

## **DESENVOLVIMENTO**

Algumas definições passam a ser importantes para o prosseguimento desse trabalho. Primeiramente o dado. Esta escolha, de iniciar pelo dado, não é trivial. É possível que ontologicamente falando a menor parte que poderíamos decompor uma mensagem ou documento seria o dado. Para Semidão figura como “fato” ou “átomo”, um elemento bruto, desprovido de significação imediata. Algo perceptível e de baixo teor semântico. (SEMIDÃO, p. 71 2004)

Semidão faz uma análise etimológica do termo “dado”. Para ele é possível extrair algumas noções centrais acerca do uso da palavra. A partir da sua pesquisa o autor expressa que dado é utilizado como ideia de algo que aconteceu, feito ou realizado, algo que é atribuído a um objeto cognoscitivo uma característica que expressa a essência constitutiva dele. Estatísticas e fatos coletados que juntas são capazes de criar referências e análises, coisas conhecidas e ou assumidas como fatos e que constituem base para

o raciocínio ou cálculo, suposição ou premissa de onde se pode iniciar inferências e por fim um pedaço de informação. (SEMIDÃO, p. 71-72 2004) Estas características, segundo Semidão, corroboram a noção depreendida a partir delas: dados como ponto base de um processo cognitivo. (SEMIDÃO, p. 72 2004)

Uma coisa dada ou concedida; algo conhecido ou suposto como fato, e feito base para raciocínio ou para cálculo; uma suposição ou premissa da qual são elaboradas inferências. (SEMIDÃO, p. 72 2004)

Silva e Gomes (2015) em seu artigo buscaram organizar diversas manifestações conceituais de informação na CI. A ideia de dado como base ou parte da informação é recorrente nas definições sobre informação. Para Hayes (1986) informação é uma propriedade dos dados resultante de ou produzida por um processo realizado sobre os dados (SEMIDÃO, p. 72 2004) Saracevic e Wood (1986) informação consolidada seria um conjunto de mensagens ou um sentido atribuído aos dados. A ideia de dados organizados como informação também está presente no Harrold's Librarian's Glossary (1989). (SILVA E GOMES, p. 146-147 2015)

Para Luciano Floridi (2005),

Informação semântica definida em quatro etapas: D.1. A Informação ( $\lambda$ ) é constituída por  $n$  dados ( $d$ ), sendo  $n \geq 1$ ; D.2. Os dados são bem formados ( $wfd$ ); D.3. Os  $wfd$  são significativos, ou seja, possuem um significado ( $mwfd = \delta$ ); F.4. Os  $\delta$  são verdadeiros. (SILVA E GOMES, p. 147 2015)

Silva e Gomes 2015 Relações de significado quantitativo (metadados) e qualitativo (conteúdos histórica e cognitivamente potenciais dos sujeitos da informação ((desenvolver))) p.148

Semidão (2004) em sua análise etimológica do vocábulo informação conclui que as definições o termo informação está invariavelmente relacionado ao conhecimento e à comunicação, independentemente da motivação que suscita o ato de informar (SEMIDÃO, p. 77 2004). Assim, todas as definições remeteriam em maior ou menor grau a noção processo cognitivo (processo de conhecer) em outras palavras, uma transposição ou mudança de um estado de coisas para outro mais apto e valorado, em termos cognoscitivos. (SEMIDÃO, p. 78 2004)

Para Silva e Gomes (2015) a pluralidade conceitual de informação na ciência da informação seja considerando particularidades conceituais, seja considerando as vozes conceituais dos estudiosos reverberam de uma multiplicidade de locais e origens em todo o globo e constroem percepções universais ou locais sobre informação.

Para Capurro e Hjørland (2007) O conceito de informação como é usado em inglês cotidianamente, no sentido de um conhecimento comunicado desempenha papel central na sociedade atual. Eles citam um estudo de Schrader (1983) no qual o autor estudou aproximadamente 700 definições de CI e seus antecedentes, e conclui que:

A literatura de CI é caracterizada pelo caos conceitual. Este Caos conceitual

advém de uma variedade de problemas na literatura conceitual da CI: citação acrítica de definições anteriores, fusão de teoria e prática, afirmações obsessivas de status científico, uma visão estreita da tecnologia, analogias inadequadas, definições circulares e multiplicidade de noções vagas, contraditórias e, às vezes, bizarras quanto à natureza do termo informação (SCHRADER, 1983, p.99)

Para os autores não deveríamos apresentar o conceito de informação isoladamente, mas deveríamos considerá-los de junto com outros conceitos. Os autores seguem suas conclusões avaliando que por mais que uma definição conceitual possa beneficiar o status do profissional da área, tal condição não é absoluta e existe boas produções que não se valem de um conceito de informação. Assim, o termo deve ser usado para melhor elucidar as questões teóricas e não a ser.

Há muitos conceitos de informação, como já demonstrado nesse trabalho, estes estão espalhados e inseridos em estruturas teóricas que segundo os autores são ora mais explícitas ora menos. Contudo, Capurro e Hjørland (2007) consideram uma distinção mais importante: entre informação como objeto ou coisa (por exemplo, números de bits) e informação como conceito subjetivo, informação como signo; isto é, como dependente da interpretação de um agente cognitivo. (CAPURRO e HJORLAND p. 193 2007)

Segundo Innarelli (2015), a sociedade preserva e constrói conhecimento por meio do processo de registro e da ininterrupta transmissão de informação. Através da capacidade humana de apropriar-se do conhecimento gerado e armazenado o reproduz e o desenvolve. E assim, através desse ciclo é possível alcançar um novo estado de conhecimento e o desenvolvimento sociocultural e científico da humanidade.

Esta transmissão ininterrupta do conhecimento a apropriação pelo ser humano ao longo da história se realizou para a satisfação das crescentes necessidades humanas. A transformação das tecnologias de transmissão da informação que culminariam no que hoje chamamos de sociedade da informação se desenrolam de igual modo a partir da mesma razão. (INNARELLI p. 40, 2015)

Para Le Coadic (2004) um conhecimento é o ato de conhecer, neste ato o espírito apreende um objeto. Nesse sentido conhecer algo é ter a capacidade de construir a ideia sobre esta coisa. Para Nonaka e Takeuchi (1997) e Nonaka (2000) o conhecimento é fruto da espiral de conhecimento e diferencia-se em conhecimento tácito e explícito, sendo o conhecimento tácito o conhecimento pessoal e desenvolvido a partir de experiências individuais e o explícito o formal, sintetizado e registrado. A produção do conhecimento nesse modelo depende da transmissão. Este acontece em duas redes distintas, a rede soft é em geral formada por relações informais e sustenta a criação de conhecimentos tácitos já a rede hard se constitui de relacionamentos formais, institucionais, pelas estruturas administrativas e infraestruturas tecnológicas comunicacionais e sustenta a produção e conhecimento explícito. (INNARELLI p. 41, 2015)

Em concordância com Nonaka e Takeuchi sobre o processo de produção

e transmissão do conhecimento, Davenport e Prusak (1998) também fazem uma abordagem conceitual da comunicação da informação na instituição. Para os autores a informação é comunicada a partir das redes hard e soft, sendo que ambas permeiam a geração do conhecimento nas instituições. Rossetti e Morales (2007) também apresentam a mesma visão de Nonaka e Takeuchi no que diz respeito ao conhecimento tácito e explícito e às redes de comunicação apresentadas como soft e hard (INNARELLI p. 41, 2015)

Buckland (1991) aborda informação como o objeto de produção do conhecimento e subdivide três categorias distintas information-as-process, information-as-knowledge e information-as-thing. Nesta abordagem, o autor apresenta a information-as-process como processo de transmissão da informação, information-as-knowledge como informação assimilada (não registrado) e a information-as-thing como conhecimento tangível (registrado). (INNARELLI p. 42, 2015)

Para Innarelli (2015) é possível entender information-as-process e o information-as-knowledge correspondem à transmissão e assimilação da informação. E o conhecimento explícito registrado em documentos e arquivos refere-se ao information-as-thing.

Para Buckland (1997,1998) o documento tem como base a informação registrada, este registro de conhecimento acontece por meio do uso de recursos tecnológicos, as redes formais de comunicação, da infraestrutura da instituição e do registro da informação. Innarelli (2015) expressa essa definição por meio de uma expressão: Documento = Conhecimento explícito = Information-as-thing (INNARELLI, 2015, p.52)

O objeto da preservação da informação é o conhecimento explícito, ou registrado, a information-as-thing ou o documento. Sobre este objeto que as teorias, técnicas e estratégias de preservação se debruçaram.

Para Innarelli (2015) a diferença entre um documento analógico e um documento digital precede o suporte, sendo o fator preponderante a capacidade humana de ter acesso à informação registrada sem a utilização de uma tecnologia externa que decodifique a informação. Como forma de exemplo ele utiliza um cartão perfurado onde a informação ali registrada não está acessível a um usuário a menos que ele tenha acesso a uma tecnologia que o auxilie, o suporte físico ser o papel não garante que o objeto informacional seja analógico.

Devemos, contudo, salientar que os textos armazenados em computadores ou mesmo na nuvem estão em algum lugar armazenado em discos físicos, o objeto físico do documento digital constitui o que o hardware é capaz de interagir e interpretar. Assim, o hardware transformará os símbolos inscritos no suporte físico em um conjunto de dados inteligíveis para o software em alguma linguagem de programação. Este converterá as mesmas em uma linguagem inteligível para o usuário. (FERREIRA, 2006, p. 22)

Para Innarelli (2015) uma vez compreendemos a importância do documento frente ao processo de valorização da informação na sociedade da informação e ao abordar os documentos no âmbito da CI surgem duas questões importantes. A primeira quais são

estes documentos? E a segunda o fato de serem considerados documentos o tornam arquivísticos? (INNARELLI, p. 52. 2015)

Para responder essas questões, Innarelli inicia discutindo os objetivos das instituições responsáveis pela guarda de documentos uma vez que o significado de documento poderá variar a partir de cada instituição. (INNARELLI, p. 53. 2015)

Assim, segundo Paes:

ARQUIVO – É a acumulação ordenada dos documentos, em sua maioria textuais, criados por uma instituição ou pessoa, no curso de sua atividade, e preservados para consecução de seus objetivos, visando à utilidade que poderão oferecer no futuro.

BIBLIOTECA – É o conjunto de material, em sua maioria impresso, disposto ordenadamente para estudo, pesquisa e consulta.

MUSEU – É uma instituição de interesse público, criada com a finalidade de conservar, estudar e colocar à disposição do público conjuntos de peças e objetos de valor cultural (PAES, 2004, p. 16). (INNARELLI, p. 52. 2015)

Para Belloto,

A forma/função pela qual o documento é criado é que determina seu uso e seu destino de armazenamento futuro. É a razão de sua origem e de seu emprego, e não o suporte sobre o qual está constituído, que vai determinar sua condição de documento de arquivo, de biblioteca, de centro de documentação ou de museu (BELLOTTO, 2006, p. 36).

Innarelli (2015) ao definir os destinos preferenciais de cada tipo de documentos, descreve como sendo documentos de biblioteca qualquer informação fixada em suporte cuja função principal é didática, técnica, cultural e/ou científica. (INNARELLI, p. 54. 2015) Já ao descrever os documentos preferenciais dos museus, Innarelli argumenta que apesar de museus e Bibliotecas serem instituições colecionadoras e temáticas museus se distinguem das bibliotecas por atribuírem significados diferentes para cada objeto. (INNARELLI, p. 54. 2015)

No museu, de forma muito contundente, tem-se consciência de que o acervo é composto de obras ou peças aos quais se atribui – intencionalmente – um valor documental. [...] No entanto, as peças ou objetos, na condição de documentos intencionalmente incorporados ao acervo museológico, não significam per se: o significado lhes é atribuído em função dos objetivos do museu. Assim, por exemplo, uma faca de prata receberá significados e descrições diferenciadas caso integre o acervo de um museu do crime, um museu da gastronomia ou um museu da ourivesaria (SMIT, 2011, p. 33) (INNARELLI, p. 54. 2015)

Na visão arquivística, argumenta Innarelli (2015), o documento produzido no decorrer das atividades ou funções de uma entidade pública ou privada, tem como função primária o registro das atividades ou funções que viabilizam o funcionamento institucional. (INNARELLI, p. 54. 2015) Assim, para o autor, o documento de arquivo é visceral pois sua relação com o produtor é o fator que o define como arquivístico, e assim sendo, sem esta

relação o documento perde seu significado e seu valor arquivístico.

Os documentos de arquivo são produzidos por uma entidade pública ou privada ou por uma família ou pessoa no transcurso das funções que justificam sua existência como tal, guardando essas documentos relações orgânicas entre si. [...] Surgem, pois, por motivos funcionais administrativos e legais. Tratam sobretudo de provar, de testemunhar alguma coisa. (BELLOTTO, 2006, p. 37) (INNARELLI, p. 54. 2015)

Masson (2008) ao tentar responder o que é um repositório faz o que ela chama de “arqueologia da palavra”, esta ação se justificaria por causa dos múltiplos sentidos e usos que se tem atribuído a palavra repositório. (MASON p106. 2008) Assim, o primeiro sentido apresentado, dicionarizado, repositório seria aquilo que “que é próprio para guardar alguma coisa; depósito” (MASON p106. 2008)

O verbete, “repertório”, do latim “repertorium”, é usado em sinonímia a “catálogo seletivo” e a “inventário”, em terminologia Arquivística, e é terminologia biblioteconômica. O significado vernacular, no dicionário, é registrado como “coleção, compilação, conjunto” (Idem: 1218), mantendo, portanto, similaridade de sentido com o significado terminológico (MASON p106-107. 2008)

A palavra repositório aparece como um verbete no dicionário Terminologia Arquivística (Camargo e Bellotto, 1996: 67). “Repositório Ver: depósito (2).” No verbete “depósito”, temos o sentido 1: “Ação pela qual os documentos são colocados sob a custódia de uma instituição arquivística.” No sentido 2, indicado pelo verbete “repositório”, assim está definido: “Prédio ou área para armazenamento” (MASON p107. 2008) Contudo para Mason argumenta que o verbete que lhe parece mais próximo do objeto repositório digital seria: “depósito legal: obrigação legal de remeter a bibliotecas, e excepcionalmente a arquivos públicos, um ou mais exemplares de cada obra impressa no país, estado ou município.” (MASON p107. 2008)

“Dicionário do Livro: terminologia relativa ao suporte”, Dicionario enciclopédico de Ciencias de la Documentación de Lópes Yepes” e “Dicionário de terminologia arquivística” o verbete Repositório não foi encontrado. Contudo no dicionário do IBLN, encontramos o verbete deposito legal: “Depósito legal 1. Entrega obrigatória e gratuita de um ou vários exemplares de toda e qualquer publicação impressa num País a uma ou mais instituições públicas para tal designada na lei. 2. Entrega imediata, obrigatória e gratuita de originais de determinados documentos de arquivo a uma instituição pública designada na lei para sua depositária (ex: leis, acordos, etc.) v. tb Aquisição p. 31.” (MASON p107-108. 2008)

No “Diccionario del archivero bibliotecario: terminología de la elaboración, tratamiento y utilización de los materiales propios de los centros documentales” contém o verbete “repositorio archivístico: lugar donde se guarda un archivo” amplia: “repositorio bibliográfico: biblioteca” e “repositorio documental: archivo”. Nos conceitos assim definidos, faz sentido o paralelismo: “repositório digital: repositório de objetos digitais”, mas, sem dúvida, tautológico. (MASON p108. 2008)

Mason (2015) na tentativa de averiguar a possível função paralela do depósito legal e do repositório apresenta Lei Federal (brasileira) nº 10.994, de 14 de dezembro de 2004 que define a Biblioteca Nacional como depósito legal de todas as publicações do país. (MASON p108. 2008)

No site da Biblioteca Nacional, há a citação do Artigo 2º da Lei nº 10.094/2004, com a definição de “depósito legal” como sendo uma exigência, por força de lei, de remessa à Biblioteca Nacional de um exemplar de todas as publicações produzidas em território nacional, por qualquer meio ou processo. O objetivo principal do Depósito Legal é assegurar a coleta, a guarda e a difusão da produção intelectual brasileira, visando à preservação e formação da Coleção Memória Nacional (MASON p108. 2008)

O repositório digital, argumenta Mason (2015) instituições) parece ter como objetivo preservar as publicações criadas em meio digital, ou as que são digitalizadas, referentes a artigos em periódicos, ou na web, os próprios periódicos, as atas de congresso e outros eventos, trabalhos de pesquisas, teses e dissertações para que constituam uma memória institucional e que possam estar disponível e acessível a quem precisar consultá-la, ou, no caso dos repositórios digitais temáticos, a memória do conhecimento em uma área especializada do conhecimento. (MASON p.109, 2008)

Mason argumenta que de uma perspectiva epistemológica e teórico - metodológico os repositórios digitais não têm sido objetos de estudos na CI mesmo que estejam frequentemente presentes em artigos. Estes parecem estar mais dedicados a defesa da divulgação das implementações até agora realizadas do que preocupados em estudar os repositórios digitais na área da CI. (MASON p128. 2008)

Mason (2008) reúne definições de repositórios digitais, estas definições permitem averiguar a pluralidade de respostas para a pergunta: “O que são repositórios digitais”. Estas variam entre ser uma ferramenta, um serviço ou uma instituição, cujo objetivo principal é a preservação e disseminação da informação, em geral, restrita à informação científica e o seu uso na comunidade científica e nas universidades, numa modalidade alternativa e aperfeiçoada de comunicação científica, assim como são os periódicos e publicações eletrônicas (MASON p132. 2008)

repositório digital aparece definido ora como coleção, ora como arquivo. Em qual dos sentidos de arquivo? É no sentido percebido em Informática? Se fosse assim porque a designação original foram “open archives” e não “open files”? Tanto coleção como arquivo são vocábulos que aparecem nas definições ... seja na afirmação de que repositórios institucionais são constituídos por “coleção de arquivos digitais”, ou “arquivos digitais de coleções”, o que evidentemente não é a mesma coisa, pois a ordem, neste caso, altera o significado. Ou, ainda, como “coleções digitais que armazenam [...] a produção intelectual das comunidades universitárias”, o que resulta em circularidade e redundância. (MASON p134. 2008)

O repositórios digitais então funcionariam a partir das seguintes características: o auto arquivamento, como uma variação dos depósitos legais de obras; a interoperabilidade,

para a garantia da disseminação e acesso, através da adoção de protocolos de arquivos abertos e, idealmente de acesso aberto, e a preservação de informação produzida em meio digital, constituindo se, ao mesmo tempo, em garantia de visibilidade institucional e da memória institucional, porém em um conceito de memória diverso daquele previsto, tradicionalmente, pelos arquivos e bibliotecas. (MASON p139. 2008)

Para Ferreira (2006) os níveis de abstração do objeto digital iniciariam com o Objeto físico, este inscrito em uma linguagem de máquina seria interpretado pelo hardware que o elevaria a objeto lógico inscrito em alguma linguagem de programação seria então interpretado pelo Software convertido em um objeto conceitual que poderia então ser utilizado pelo ser humano e assim construiria um objeto experimentado. Essa sequência é uma via de mão dupla, é possível que este documento esteja sendo lido de forma digital e nesse sentido o sistema estaria descrito na ordem exata que está ocorrendo. Contudo, este documento também fora concebido de forma digital e assim sendo, este passou do conceito imaginado por mim até o seu armazenamento no suporte físico em uma linguagem e máquina, para só depois poder ser transportado, copiado ou lido.

Em uma situação ideal o objeto conceitual, seja o documento analógico ou digital, deverá se assemelhar muito ao objeto conceitual imaginado pelo emissor. Assim sendo, A preservação digital é a atividade responsável por garantir que a comunicação entre um emissor e um receptor é possível, não só através do espaço, mas também através do tempo. (FERREIRA, 2006, p. 24). Para Ferreira, a preservação digital só é possível se a totalidade da abstração já descrita acima se encontrarem descrita e interpretáveis. Se a cadeia de interpretação que permite elevar um objeto digital desde o seu nível físico até ao nível conceptual for rompida, a comunicação deixa de ser possível e o objeto perder-se-á para sempre. (FERREIRA, 2006, p. 24)

O número de estratégias disponíveis com o objetivo de solucionar a problemática da preservação dos objetos digitais tem crescido nos últimos anos segundo Ferreira (2006). Para Lee as estratégias podem ser agrupadas em três classe fundamentais emulação, migração e encapsulamento (FERREIRA, 2006, p. 31). Já Thibodeau organiza a estratégias em um mapa bidimensional colocando em extremos do eixo cartesiano a preservação do objeto físico/lógico e no outro a preservação do objeto conceitual. (FERREIRA, 2006, p. 31)

Para Ferreira (2006) em termos genéricos o conceito de autenticidade é utilizado para traduzir *elementos diplomáticos que permitem aferir se um dado objeto é autêntico* (FERREIRA 2006, p.50 ) Refere-se a identificação *do “porquê”, do “quando”, do “onde” e do “por quem” de um objeto digital* (FERREIRA 2006, p.50 ). Com essa finalidade, seria segundo Ferreira, imprescindível *contextualizar a sua existência, descrever a sua história de custódia e atestar que a sua integridade não foi comprometida, provar que existe um conjunto de propriedades, consideradas significativas, que foram corretamente preservadas ao longo do tempo.* (FERREIRA 2006, p.50 )

Se o problema da autenticidade não nasceu com o advento do objeto digital, uma vez

que esta preocupação é encontrada ao longo da história, é no contexto do documento de natureza digital que este problema se torne sensivelmente mais complexo (Ferreira 2006) A simplicidade com que alterações podem ser inseridas e a rapidez com que esta informação corrompida pode ser disseminada elevam as dificuldades de garantir a autenticidade de documentos digitais.

## **A preservação de documentos de Saúde**

No período de 1918 a 1919 a gripe espanhola se espalhou pelo mundo inteiro, matando de 20 a 80 milhões de pessoas. De origem viral, não havia tratamento conhecido. Como veio, se extinguiu. Com o intuito de pesquisar meios de evitar uma nova catástrofe, a comunidade internacional das áreas médica e de saúde pública procurou por décadas algum vestígio biológico do vírus causador dessa enfermidade. Só depois de muito tempo, foi encontrada uma amostra de tecido humano infectado pelo vírus num hospital militar da Inglaterra. A partir desses vestígios estão sendo desenvolvidas pesquisas para se descobrir vacinas e meios de tratamento da gripe espanhola. As pesquisas em torno da amostra só se tornaram possíveis graças à preservação dos arquivos científicos, datados de 1916, daquele hospital militar (SAYÃO, SALES, p. 179, 2012)

Para Moreno, Coeli e Munck (2009) desde a antiguidade é possível encontrar a necessidade de comunicar algo a alguém sobre a própria saúde, ou sobre a saúde de um terceiro ou mesmo de um grupo. E assim, informação de saúde poderia de forma inicial ser traduzido como um compósito de transmissão e/ou recepção de eventos relacionados ao cuidado em saúde. A partir do século XIX o crescimento dos estudos em epidemiologia e a necessidade crescente de se comunicar questões relacionadas à saúde de grandes populações aumenta consideravelmente as informações de saúde produzidas e transmitidas.

Quase que concomitantemente, a estatística do final desse século XIX e início do século XX, inspiradora de estudiosos como Benthan, Price, Laplace, Galton (Rosen, 1994) pode ser vista, também, como um ponto de partida importante para a geração de Informações em Saúde de forma agregada e preditiva. Daí, pode-se partir, sem muito pecado, para as primeiras peças da Informação em Saúde, compostas pelas Estatísticas Vitais, pelas Tábuas de Sobrevida, enfim, por instrumentos de predição e inferência de estados de saúde a partir do status atual de um grupo de pessoas em determinado contexto de saúde. (Moreno, Coeli e Munck 2009)

Moreno, Coeli e Munck (2009) listam três subáreas que se formaram a partir do desdobramento de numerosos caminhos que a informação de saúde percorreu. A primeira abarca o perfil da população (de que adoece e morre, dados demográficos e socioeconômicos), bem como os serviços prestados os materiais e medicamentos consumidos e a força de trabalho envolvida. A segunda subárea corresponde a conhecer as necessidades da população atendida, bem como o uso potencial e real da rede instalada e os possíveis investimentos necessários. A última subcategoria seria a fim de planejar, controlar e avaliar as ações e serviços de saúde (EPSJV, 2005).

Daí termos, de forma esquemática (Moraes, 2007), a possibilidade de

observar a Informação em Saúde como subsídio para o próprio setor saúde: na administração; na assistência; no controle e avaliação; no orçamento e finanças; no planejamento; nos recursos humanos; na regulação; na saúde suplementar; no geoprocessamento em saúde; e na vigilância (epidemiológica, sanitária, ambiental) (Moreno, Coeli e Munck 2009)

A preservação dos metadados, informações, documentos e conhecimentos produzidos no âmbito do atendimento e nas unidades de saúde não é de assim um objeto de importância exclusiva de entusiastas ou cientistas da informação. Historicamente e atualmente melhorias na gestão do conhecimento e informações de saúde constituíram poderosas ferramentas para os sistemas de saúde. Sobre a informação de saúde Piconi, Galvão e Ricarte (2013) pontuam:

a informação gerada durante o tratamento do paciente é de importante valor para o próprio paciente, para a população e para a ciência, pois é uma fonte de pesquisa, sendo considerada uma fonte de memória, patrimônio documental e cultural de uma sociedade, além de ser de extrema importância para garantir o atendimento médico e a eficiência no diagnóstico para tratamento do paciente (PICONI; GALVÃO; RICARTE, 2013, p.4)

Para Mello e Vianna (2019) esses elementos justificam a importância de se preservar as informações digitais na área da saúde, considerando que provavelmente foram os mesmos elementos que motivaram um número considerável de países a promoverem políticas de preservação digital de informações de saúde entre os quais os autores citam:

iniciativa italiana do Polo archivistico dell'Emilia-Romagna (ParER), a eBiblioCCE desenvolvida pelo Centro Nacional de Cirurgia, na Espanha, o Sistema de Preservação de Recursos Eletrônico (SPER), proposto pelo National Institutes Of Health, EUA, a metodologia de auditoria para sistemas de informação para avaliar a sua capacidade de preservar a informação digital de forma segura a médio e longo prazo –TRAC, desenvolvida na Espanha, a proposta para armazenar às informações em DNA, desenvolvida na Europa e entre outras. (Mello e Vianna p. 22 2019)

Para Pinto, Sales (2017) a preservação das informações de saúde é fundamental tanto as oriundas de pesquisas científicas quanto as produzidas pelo atendimento médico, como os prontuários eletrônicos (PEP). Para as autoras a preservação do PEP,

representante maior da chamada documentação sanitária – garantiria a realização de pesquisas e estudos comparativos embasados nessas fontes que visassem investigar o surgimento dessas doenças ou se elas se 'inovam' bem como as medidas terapêuticas postas em execução tendo em vista a sua cura ou, ainda, as orientações prescritas para se conviver com elas (Pinto, Sales p. 2017)

E assim, concluem, uma política de preservação desses documentos seria uma condição zine qua non tanto para medicina baseada em evidência quanto para pesquisas, ensino e para a constituição de provas jurídicas.

## Repositórios de Informações de saúde, DataSharing/BR

No dia 01/07/2020 entrava em operação o COVID-19 Data Sharing/BR, este é o primeiro repositório de dados abertos relativo a Covid-19 do Brasil com dados laboratoriais e clínicos de hospitais do estado de São Paulo. De acordo com a Fapesp *“o repositório abriga dados abertos e anonimizados de, inicialmente, mais de 177 mil pacientes, 9.634 dados de desfecho e um total de quase 5 milhões de resultados de exames clínicos e laboratoriais realizados em todo o país pelo Grupo Fleury e na cidade de São Paulo pelos hospitais Israelita Albert Einstein e Sírio-Libanês desde novembro de 2019.”*

A base de dados fora resultante de uma iniciativa da própria Fapesp em parceria com a USP e tem como objetivo compartilhar *“informações compartilhar informações clínicas anonimizados de pacientes para subsidiar pesquisas científicas sobre a doença nas diversas áreas de conhecimento”*

A plataforma disponibiliza 3 conjuntos de dados, um por cada hospital, cada um deles com um dicionário de dados, um .csv para pacientes anonimizados e um .csv para resultados de exames. Na planilha paciente é possível encontrar as seguintes informações: ID\_PACIENTE, IC\_SEXO, AA\_NASCIMENTO, CD\_PAI, ICD\_UF, CD\_MUNICIPIO, CD\_CEP. Para planilha exames é disponibilizada as seguintes informações: ID\_PACIENTE, DT\_COLETA, DE\_ORIGE, IDE\_EXAME, DE\_ANALITO, DE\_RESULTADO, CD\_UNIDADE, DE\_VALOR\_REFERÊNCIA.

O Covid-19 DataSharing/BR tem como objetivo acelerar os processos de pesquisa relacionadas ao COVID em todo o mundo por meio do compartilhamento de dados de pacientes brasileiros. Diferentes de outras plataformas a iniciativa Brasileira disponibiliza os dados de forma livre e sem pré-processamento para a homogeneização. Seus conjuntos de dados têm abrangem clínica, laboratorial e informações de diagnóstico, bem como alguns dados demográficos. Todos os registros são anonimizados atender aos padrões internacionais e às leis brasileiras. (Mello, Suman, Medeiros, Prado, Rizzatti, Nunes, Barnabé, Ferreira, Reis, Rizzo, Maciel, Cesar, Carvalho. p.2 2020)

O desenvolvimento da plataforma foi gerenciado por cientistas e divididos em duas camadas. Uma primeira estratégica, um grupo de seis cientistas para o gerenciamento administrativo e questões de pesquisa do Covid-19. Um segundo grupo de cientistas da computação e de dados foi nomeado para projetar e implementar o repositório e pré-processamento dos dados. A plataforma foi projetada para a atualização periódica e projetada para no futuro suportar conter fontes e tipos de dados adicionais, incluindo imagens médicas de COVID-19 pacientes e relatórios associados. (Mello, Suman, Medeiros, Prado, Rizzatti, Nunes, Barnabé, Ferreira, Reis, Rizzo, Maciel, Cesar, Carvalho. p.2 2020)

Para sua implementação houve algumas decisões de design da plataforma que foram tomadas. O repositório seria incorporado a Rede de Repositório de Dados de Pesquisa Aberta do Estado de São Paulo, os dados seriam abertos e anonimizados em

conformidade com leis internacionais e brasileiras de proteção de dados. Os tipos de registros disponibilizados seriam aqueles que as instituições já tinham quantidades significativas de dados, fornecendo assim uma amostra inicial considerável para os pesquisadores. (Mello, Suman, Medeiros, Prado, Rizzatti, Nunes, Barnabé, Ferreira, Reis, Rizzo, Maciel, Cesar, Carvalho. p.3 2020)

O registro deveria conter a respeito dos pacientes informações como idade, sexo, espaço local. Também deveria conter exames clínicos e laboratoriais, juntamente com informações sobre o primário Endpoints e transferências de pacientes. Esses dados seriam fornecidos acerca de indivíduos que haviam se submetidos a testes para o covid-19 desde o início da pandemia no Brasil. Assim, os registros dos exames iniciariam em 1º de novembro de 2019 para permitir análises de tendências. (Mello, Suman, Medeiros, Prado, Rizzatti, Nunes, Barnabé, Ferreira, Reis, Rizzo, Maciel, Cesar, Carvalho. p.3 2020)

A plataforma seria aberta a qualquer novo integrante. A instituição de saúde deveria depositar os dados no repositório seguindo as regras pré-estabelecidas descritas acima. O desenvolvimento do projeto levou em conta também padrões internacionais de interoperabilidade, os dados devem ser documentados por meio de metadados e dicionários de dados associados, fornecendo todos os campos obrigatórios estabelecidos pelo Rede de Repositórios de São Paulo (ou seja, Autor, Título, Descrição, Palavras-chave / Assunto, URI, Persistente Identificador, tipo de arquivo e informações de financiamento) para permitir a localização. (Mello, Suman, Medeiros, Prado, Rizzatti, Nunes, Barnabé, Ferreira, Reis, Rizzo, Maciel, Cesar, Carvalho. p.4 2020)

A infraestrutura computacional do COVID-19 DataSharing/BR é dependente da plataforma computacional da Rede Aberta de Repositórios de Dados de Pesquisa do Estado de São Paulo. A arquitetura é semelhante à de um sistema de banco de dados federado, sendo composto por dois elementos principais:

1. repositórios das instituições participantes, que são projetados e mantido de forma independente;
2. um mecanismo de busca de metadados, que coleta e expõe diariamente por meio de uma interface comum de informações sobre os metadados disponíveis em cada instituição (Mello, Suman, Medeiros, Prado, Rizzatti, Nunes, Barnabé, Ferreira, Reis, Rizzo, Maciel, Cesar, Carvalho. p.5 2020)

Todos os metadados da federação estão alocados no repositório de metadados. O motor de busca foi desenvolvido pela Universidade de São Paulo onde é mantido o repositório de metadados. Metadados são transferidos usando o protocolo padrão OAI-PMH para coleta de metadados. Assim o repositório FAPESP COVID-19 DataSharing/BR constituiu mais um membro da arquitetura federada da Rede Aberta de Repositórios de Dados de Pesquisa do Estado de São Paulo. (Mello, Suman, Medeiros, Prado, Rizzatti, Nunes, Barnabé, Ferreira, Reis, Rizzo, Maciel, Cesar, Carvalho. p.5 2020)

O design de implementação do COVID-19 DataSharing/BR foi desenvolvido e é

mantido pela Universidade de São Paulo. Para tanto foi considerado algumas questões de projeto na implementação. A seleção e o pré-processamento (anonimização) dos dados são realizados pelas instituições que fornecedoras. Assim, todos os dados da plataforma já ingressam despersonalizados. Os usuários da plataforma não têm acesso aos dados de origem (pessoalizados) que ficam sob a guarda das instituições de origem. O Repositório é particionado por instituição o que permite cada participante do repositório construir um produto diferente adicionando novos tipos de dados à sua partição quando estes forem disponíveis. (Mello, Suman, Medeiros, Prado, Rizzatti, Nunes, Barnabé, Ferreira, Reis, Rizzo, Maciel, Cesar, Carvalho. p.5 2020)

Com objetivo de assegurar a segurança da plataforma cada instituição possui acesso autorizado e seguro protocolo para fazer upload de seu conjunto de dados no repositório usando uma interface padrão onde metadados são fornecidos. (figura) Depois do upload dos metadados pela instituição cada conjunto de dados será verificado de acordo com verificado em relação a um conjunto de regras. Nesse momento busca-se garantir sua conformidade com a despersonalização e esquemas de nomenclatura do repositório. Caso os dados inseridos na plataforma não estejam em conformidade com as regras estabelecidas eles são devolvidos as instituições de origem. (Mello, Suman, Medeiros, Prado, Rizzatti, Nunes, Barnabé, Ferreira, Reis, Rizzo, Maciel, Cesar, Carvalho. p.6 2020)

Para segurança física dos metadados o repositório armazena os metadados em estruturas de nuvem observando a uma redundância fornecida por dois datacenters. Há um monitoramento constante do crescimento do volume de dados e uma equipe específica realiza periodicamente testes relacionados à escalabilidade e exatidão em a fim de garantir o funcionamento contínuo do repositório. Esta mesma equipe é responsável por registrar novas instituições e incluir novas funcionalidades. (Mello, Suman, Medeiros, Prado, Rizzatti, Nunes, Barnabé, Ferreira, Reis, Rizzo, Maciel, Cesar, Carvalho. p.6 2020)

Por fim, com o objetivo de garantir mais independência entre as instituições fornecedoras dos metadados, elas devem enviar um conjunto de dados contendo não só os próprios dados, mas também os dicionários que descrevem os pacotes. Esse ponto visa garantir que no futuro cada instituição possa adaptar seu produto além do pacote inicial e assim por exemplo, adicionar imagens ou informações de diagnóstico. (Mello, Suman, Medeiros, Prado, Rizzatti, Nunes, Barnabé, Ferreira, Reis, Rizzo, Maciel, Cesar, Carvalho. p.6 2020)

O repositório foi publicado em duas etapas. Na primeira etapa uma amostra de 300 pacientes foi disponibilizada para testes e feedback no dia 17 de junho de 2020. No período de duas semanas estes dados foram baixados 600 vezes. principalmente do Brasil, mas também da América do Norte, Europa e Ásia. Nesse ponto houve dois principais feedbacks recebidos:

1. solicitação de atributos adicionais, alguns dos quais impossíveis de fornecer sem violar o paciente privacidade (por exemplo, por meio de precisão espacial

refinada);

2. solicitação de mais detalhes metadados. (Mello, Suman, Medeiros, Prado, Rizzatti, Nunes, Barnabé, Ferreira, Reis, Rizzo, Maciel, Cesar, Carvalho. p.6 2020)

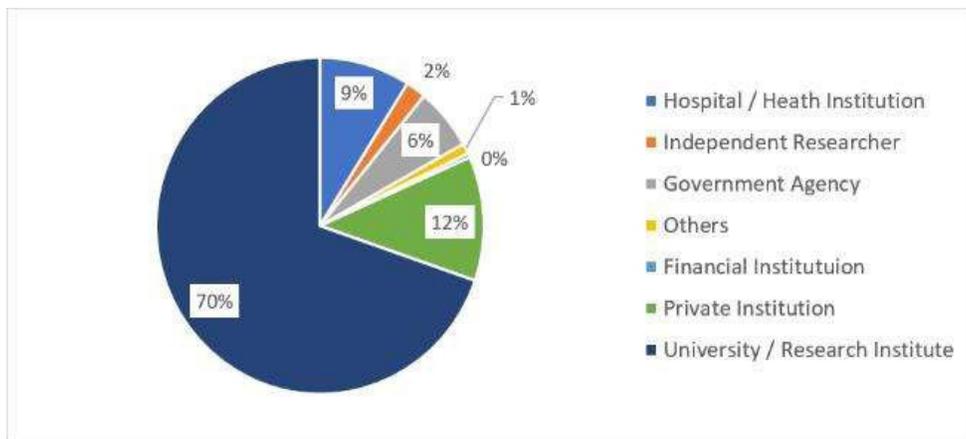
A partir do feedback obtido foi inaugurada a segunda etapa onde foi disponibilizado 1º de julho de 2020 do primeiro conjunto completo de registros de aproximadamente 180.000 indivíduos, para os quais os conteúdos de metadados. No período de duas semanas o repositório foi visto por mais de 1.500 vezes e com aproximadamente 800 downloads de dados de 15 países diferentes.

Fonte	Pacientes	Exames Clínicos	Primary endpoints
Fleury Group	129.597	2.496.592	0
Sírio Libanês Hospital	2.732	371,358	9.634
Albert Einstein Hospital	44.879	1.867.091	0
Total	177.208	4.735,041	9.634

Fonte (Mello, Suman, Medeiros, Prado, Rizzatti, Nunes, Barnabé, Ferreira, Reis, Rizzo, Maciel, Cesar, Carvalho. p.6 2020)

Ordem	Termo pesquisado	Visitas	% do Total
1	subject_keyword:pandemia	13116	80.24%
2	subject_keyword:sorologia	10482	64.13%
3	dateIssued_keyword:2020	10204	62.43%
4	subject_keyword:resultados exames	9796	59.93%
5	subject_keyword:coronavirus	9659	59.09%
6	subject_keyword:PCR	9511	58.19%
7	subject_keyword:covid-19	9157	56.02%
8	has_content_in_original_bundle_keyword: true	4180	25.57%
9	author_keyword:Fleury, Grupo	4108	25.13%
10	author_keyword: Albert Einstein, Hospital Israelita	2851	17.44%

Fonte (Mello, Suman, Medeiros, Prado, Rizzatti, Nunes, Barnabé, Ferreira, Reis, Rizzo, Maciel, Cesar, Carvalho. p.7 2020)



Fonte (Mello, Suman, Medeiros, Prado, Rizzatti, Nunes, Barnabé, Ferreira, Reis, Rizzo, Maciel, Cesar, Carvalho. p.7 2020)

Nogueira, Santos, Batista, Maia, Carvalho, Ferreira, Silva, Giunchetti, Paz, Galdino, (2020) desenvolveram um estudo que buscava ações e estratégias de universidades e institutos de pesquisa para o diagnóstico rápido de Covid-19 no Brasil foram identificados 88 iniciativas para o diagnóstico divididas entre molecular, imunológicos e métodos alternativos como o uso de aprendizado de máquina e inteligência artificial.

Atualmente, o diagnóstico da COVID-19 é realizado por meio de métodos moleculares, como a transcrição reversa do RNA viral seguida de reação em cadeia da polimerase (RT-PCR). Contudo, é uma técnica que exige equipamentos sofisticados e profissionais treinados para a sua realização (GUO et al., 2020). Para complementar os métodos moleculares, testes sorológicos estão sendo realizados para auxiliar no diagnóstico da COVID-19. Além de atuarem como indicadores de infecção podem fornecer informações sobre o tempo de exposição ao vírus (UDUGAMA et al., 2020). (Nogueira, Santos, Batista, Maia, Carvalho, Ferreira, Silva, Giunchetti, Paz, Galdino, p. 163, 2020)

As pesquisas de testes para o covid-19 baseadas em técnicas de inteligência artificial associada a imagem ou outros testes clínicos mostram-se não só promissoras mais capazes de associadas aos atuais testes levar a uma política de testagem mais barata e eficiente.

Origem	Objetivo	Nome
Laboratório de Imagem, Sinais e Acústica e do Laboratório de Tecnologia da Tomada de Decisão (UnB)	sistema que consegue determinar, em segundos, casos suspeitos da COVID-19 por meio da leitura de radiografias	
Universidade Estadual de Maringá (UEM)	Identificar a pneumonia causada pela COVID-19 utilizando imagens de radiografias de tórax.	RYDLS-20
Laboratório de Computação de Sistemas Inteligentes, juntamente com integrantes do Departamento de Informática (UFPR)	identificação de padrões causados pela COVID-19 através de imagens de raio-x.	
Pesquisadores da Faculdade de Ciências Médicas (UERJ)	utilização da Ultrassonografia de Tórax como alternativa para a avaliação dos casos graves que necessitam de acompanhamento mais efetivo *****	
PPG em Engenharia Biomédica da UFPE	um sistema inteligente cujo objetivo é otimizar o diagnóstico através de radiografias do tórax	IKONOS
pesquisadores da Universidade Federal de Santa Catarina (UFSC)	propõem a classificação de imagens de raio-X e tomografias da região torácica de pacientes suspeitos por meio da inteligência artificial	
Departamento de Ciência da Computação (UFRRJ)	Aplicativo avaliar se a imagem de radiografia possui o padrão associado à COVID-19 ou outras doenças respiratórias	XRayCovid-19
pesquisadores da USP juntamente com o Hospital das Clínicas	reúne dados sobre casos do Brasil e tem como referência imagens de radiografias e tomografias da região do tórax. A RadVid19 possui um algoritmo que avaliará as imagens em poucos segundos e fornecerá informações de diferentes estágios da evolução da doença	RadVid19
Hospital Universitário do Oeste do Paraná	estudo para a detecção da COVID-19, por meio de exames de rotina como o hemograma. O algoritmo do estudo reúne, padrões para identificar se o paciente tem pré-disposição positiva ou negativa para a doença	
Programa de Pós-graduação em Engenharia Biomédica da UFPE	propõem o desenvolvimento do Sistema Heg.IA que consiste em um protótipo de um sistema de apoio ao diagnóstico com base em exames de sangue	
Escola de Engenharia da UFMG juntamente com a Faculdade de Medicina, do Hospital das Clínicas e do Departamento de Computação da Universidade Federal de Ouro Preto (UFOP)	propõem a criação de um software baseado em inteligência artificial, visando ajudar na detecção automática de síndromes respiratórias agudas	Minds

Fonte Nogueira, Santos, Batista, Maia, Carvalho, Ferreira, Silva, Giunchetti, Paz, Galdino, p. 164- 165, 2020

Algumas dessas iniciativas já estão em funcionamento, outras ainda estão em fase de testes. Contudo cada uma dessas iniciativa mostra a reação do sistema de saúde e da ciência brasileira a um momento de grande estresse social, econômico e sanitário. Demonstra que existe a capacidade dentro do país de produção de produtos informacionais capazes de impactar diretamente o sistema de saúde brasileiro.

O Ravid-19, um dos testes listados, já está em operação em seu site<sup>1</sup> fornece alguns números atuais dos trabalhos desenvolvidos. Ao todo são 39 centros de diagnóstico e hospitais credenciados, 9 empresas parceiras e um total 10.270 imagens armazenadas. A plataforma lista 3 benefícios do uso de inteligência artificial na pandemia primeiramente é mais econômico e permite um uso mais eficiente dos recursos médicos tanto na otimização do uso dos exames de diagnostico quanto na gerência dos leitos e da ocupação dos hospitais. O segundo benefício listado é a melhoria do atendimento ao paciente, tornando mais rápido e eficiente, permitindo o início do tratamento o mais rápido possível. E por fim o impacto na qualidade do diagnostico, tornando-o mais assertivo na detecção da covid-19.

Contudo esses benefícios listados poderiam ser aplicados a qualquer uma das iniciativas listadas. Existe, contudo, alguns entraves para o desenvolvimento e mesmo para empregar soluções informacionais como essas em larga escala, faltam repositórios confiáveis disponíveis em sistemas abertos de informação.

No período de JAN/2008 até JUN/2020 foram realizadas no Brasil 743.938.541 radiografia, em média 9.853.490,61 por mês. Só durante o período da pandemia foram feitas 32.682.854 médias de 4.085.356,75 por mês. No mesmo período, foram realizadas no Brasil 189.817.284 ultrassonografias, 44.921.820 tomografias e 10.567.359 exames de diagnóstico de ressonância Magnética. somente no sistema público de saúde, todos os dados são Ministério da Saúde - Sistema de Informações Ambulatoriais do SUS (SIA/SUS).

Todos esses exames diagnósticos devem ser armazenados por até 20 anos no prontuário dos pacientes. De acordo com Artigo 6 da Lei nº 13.787 de 27 de dezembro de 2018 *“Decorrido o prazo mínimo de 20 (vinte) anos a partir do último registro, os prontuários em suporte de papel e os digitalizados poderão ser eliminados.”* Assim, a cada 20 anos uma parte da memória de saúde do Brasil se perde, mais do que isso a possibilidade de utilização dessa informação em produtos de informação de qualquer natureza também se perde.

No Brasil a saúde é um direito constitucional, este direito consome x% do PIB e enfrenta sérios problemas com oferta insuficiente, falta de médicos e especialistas em alguns locais. O uso dos documentos de saúde tem a capacidade de promover rápidos ganhos de eficiência tanto em nível tático como estratégico. A preservação desses documentos e a construção de Repositórios mais abrangentes se mostra desafiadora em um país de proporções continentais como o Brasil, mais igualmente recompensadora.

---

<sup>1</sup> <https://radvid19.com.br/>

## Conclusão

Os repositórios digitais constituem uma possível ferramenta para políticas de preservação de dados de saúde. A construção destes podem, como no caso DataSharing/BR se aproveitar de infraestruturas já existentes no Brasil e respeitando a legislação no que se refere a privacidade dos atendimentos médicos poderiam disponibilizar para pesquisadores das mais variadas áreas, universidades, gestores públicos e mercado dados, informações e documentos capazes de constituir insumos para suas pesquisas, decisões e algoritmos.

A opção pelo fomento desses repositórios pode também contornar dificuldades técnicas e de infraestrutura existente, assim como os investimentos em um sistema que seja nacional e unificado. O que seria o melhor dos mundos possíveis, contudo distante da realidade atual. A produção de repositórios em menor escala bem como os prováveis impactos positivos podem fomentar no futuro modelos mais abrangentes e institucionais.

## CONSIDERAÇÕES FINAIS

No Brasil o seja a academia, o mercado ou o estado tem desenvolvido, ainda que de modo insipiente, produtos informacionais que poderiam de forma isolada ou conjunta impactar a saúde dos brasileiros e o sistema de saúde brasileiro de forma positiva. Se é verdade que falta de investimento, desarticulação entre público e privado, desinteresse de certos setores da sociedade, resistência interna cultural a mudanças também é verdade que as informações em saúde muita das vezes estão indisponíveis ou em locais de difícil acesso. A construção de repositórios públicos e acessíveis de informações em saúde se não resolve a situação age como forma de catalizadora na medida que viabiliza ideias, prova e constrói racionais, algoritmos funcionais e programas.

A falta de uma política pública de informação eficiente e perene capaz de preservar e tornar acessível a informação em saúde afasta o Brasil dos possíveis avanços que o século XXI oferta postergando o sonho de uma saúde mais eficiente e justa.

É imprescindível seja da perspectiva da democracia, accountability e transparência, seja pelo potencial catalizador e disruptivo a criação, manutenção e gerenciamento de sistemas de repositórios de informações de saúde. Mesmo que esses repositórios recebam sejam inicialmente incompletas, como acontece hoje, uma política de preservação e acessibilidade da informação pública deve ser direcionada em duas direções uma atual, disruptiva e de criação de um novo paradigma e uma segunda futurista e perene. Devemos ter sempre em mente esses dois momentos ou passaremos anos lutando por uma utopia que pode nunca se realizar perdendo muito crescimento na atualidade.

## REFERÊNCIAS

ARELLANO, Miguel Angel. Preservação de documentos digitais. **Ci. Inf., Brasília**, v. 33, n. 2, p. 15-27, maio/ago. 2004

BRASIL. **Lei nº 13.787** de 27 de dezembro de 2018.

CAPURRO, Rafael; HJORLAND, Birger O conceito de informação. **Perspectivas em Ciência da Informação**, v. 12, n.1, p.148-207, jan/abr. 2007

INNARELLI, Humberto Celeste. **Gestão da preservação de documentos arquivísticos digitais**: proposta de um modelo conceitual. Tese de doutorado - Programa de Pós- graduação em ciência da informação – Escola de comunicação e Artes / Universidade de São Paulo

MASSON, Sílvia Mendes. Os Repositórios digitais no âmbito da Sociedade Informacional. **PRISMA. COM.** n. ° 7 2008 ISSN: 1646 – 3153

MELLO, Josiane; VIANNA, William Barbosa. Preservação Digital Da Informação Em Saúde: Panorama Quali-Quantitativo Da Produção Científica Internacional. **Pesq. Bras. em Ci. da Inf. e Bib.**, João Pessoa, v. 14, n. 2, p. 15-352019

MELLO, Luiz Eugenio; SUMAN, Andrea; MEDEIROS, Claudia Bauzer; PRADO, Claudia; RIZZATTI, Edgar; NUNES, Fatima L. S.; BARNABÉ, Gabriela FERREIRA, João Eduardo; SÁ, José de; REIS, Luiz Fernando RIZZO, Luiz Vicente; SARNO, Luzia; LAMONICA, Raphael de; MACIEL, Rui M. B.; CESAR JR, Roberto Marcondes; CARVALHO, Rodrigo. **Opening Brazilian COVID-19 patient data to support world research on pandemics.**

MORENO, A. B.; COELI, C. M.; MUNCK, S. **Informação em saúde. Dicionário da educação profissional em saúde.** Disponível em <http://www.sites.epsjv.fiocruz.br/dicionario/verbetes/infsau.html>

MORENO, Fernanda Passini; LEITE, Fernando César Lima; ARELLANO, Miguel Ángel Márdero. Acesso livre a publicações e repositórios digitais em ciência da informação no Brasil **Perspect. ciênc. inf.**, Belo Horizonte, v.11 n.1, p. 82-94, jan./abr. 2006

NOGUEIRA, Laís Moreira; SANTOS, Michelli dos; BATISTA, Amanda Aparecida; MAIA, Ana Clara Gontijo; CARVALHO, Sílvio Elisei Carvalho; FERREIRA, André Vinicius Fernandes SILVA, Jonatas Oliveira GIUNCHETTI, Rodolfo Cordeiro; PAZ, Mariana Campos da; GALDINO, Alexsandro Sobreira. O envolvimento da Ciência Brasileira no diagnóstico da COVID-19. **Ciência em Movimento - Reabilitação e Saúde**, n. 43, v. 22, de junho de 2020

PINTO, Virginia Bentes; SALES; Odete Máyra Mesquita. Proposta De Aplicabilidade Da Preservação Digital Ao Prontuário Eletrônico Do Paciente. **RDBCI: Revista Digital Biblioteconomia e Ciência da Informação.** DOI 10.20396/rdbci.v15i2.8646311

SANTOS, Koenigkam Santos; FERREIRA, Raniery Ferreira Júnior; WADA, Danilo Tadao TENÓRIO, Ariane Priscilla Magalhães; BARBOSA, Marcello Henrique Nogueira MARQUES, Paulo Mazzoncini de Azevedo. **Inteligência artificial, aprendizado de máquina, diagnóstico auxiliado por computador e radiômica**: avanços da imagem rumo à medicina de precisão. **Radiol Bras.** 2019 Nov/Dez;52(6):387–396

SAYÃO, Luis Fernando; SALES, Luana Farias Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**; v. 22, n. 3 (2012)

SEMIDÃO, Rafael Aparecido Moron. **Dados, Informação E Conhecimento Enquanto Elementos De Compreensão Do Universo Conceitual Da Ciência Da Informação**: Contribuições Teóricas Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2014.

SILVA, Jonathas Luiz Carvalho GOMES Henriette Ferreira Conceitos De Informação Na Ciência Da Informação: percepções analíticas, proposições e categorizações. **Inf. & Soc.:Est.**, João Pessoa, v.25, n.1, p. 145-157, jan./abr. 2015

# A CURADORIA DIGITAL NAS REDES SOCIAIS: O TWITTER DO PRESIDENTE AMERICANO

**Rosália Maria Silva Oliveira,**

Comunicóloga, pós-graduada em Mídia, Informação e Cultura (USP). Atualmente é pesquisadora e mestranda na linha de pesquisa de Organização do Conhecimento no Departamento de Informação e Cultura (ECA)

Trabalho final apresentado como Requisito para Aprovação na Disciplina de Curadoria Digital sob Coordenação e Orientação do Professor Dr. Francisco Carlos Paletta no Programa de Pós-graduação em Ciência da Informação Escola de Comunicações e Artes – USP em agosto 2021.

**RESUMO:** O presente artigo propõe apresentar uma revisão preliminar sobre o vasto universo de conceitos, aplicações e desafios relacionados à curadoria digital nas redes sociais. Por meio de pesquisa bibliográfica exploratória inicial e de exemplificação buscamos contribuir com o debate sobre tais temas na perspectiva da informação e da comunicação, no contexto digital. Além de apresentar um estudo de caso que tangibiliza a aplicação do conceito.

**PALAVRAS-CHAVE:** Curadoria Digital. Comunicação. Redes Digitais.

## DIGITAL CURATION IN SOCIAL NETWORKS: THE TWITTER OF THE AMERICAN PRESIDENT

**ABSTRACT:** This article proposes to present a preliminary review on the vast universe of concepts, applications and challenges related to digital curation on social networks. Through initial exploratory bibliographic research and exemplification, we seek to contribute to the debate on such topics from the perspective of information and communication, in the digital context. In addition to presenting a case study that makes the application of the concept tangible.

**KEYWORDS:** Content Curation. Communication. Digital networks.

## INTRODUÇÃO

Curadoria pode ser definido como “ato ou efeito de curar, função, atributo, cargo, poder de curador, curatela” (HOUAISS; VILLAR, 2009, p. 586).

Tal definição aponta um caminho a ser seguido, mas se perde na falta de substância equalificação para o termo não se torne um mero verbete jogado no meio de tantos outros que buscam dar ares de entendimento em discussões sobre possíveis tendências e assuntos da moda. Para evitar tal situação, buscou-se amparo na definição de acordo com o Dicionário de Biblioteconomia e Arquivologia, que conceitua curador, ou seja, aquele

que exerce curadoria como (CUNHA; CAVALCANTI, 2008, p.112).

Especialista que administra museu, zoológico, aquário, jardim botânico, centro de recursos naturais e sítios históricos. Faz a supervisão direta da aquisição, armazenamento e exibição das coleções, incluindo a negociação e autorização de compra ou venda, troca ou empréstimo de coleções. Também pode ser responsável pela autenticação, avaliação e categorização das espécies de uma coleção, bem como colabora na condução dos projetos de pesquisa e dos programas educacionais da instituição. Atualmente, também pode estar envolvido com as ações de levantamento de recursos e promoção, o que pode incluir a redação e revisão do projeto de pedido de verbas, artigos periódicos e material publicitário, além de participar de reuniões, convenções e eventos cívicos; dirigente de um museu ou coleção de arte.

Mesmo com a definição de uma fonte especializada em Ciência da Informação, é possível compreender que o termo possui uma complexidade que não cabe na definição sucinta do Houaiss, e ainda assim, mais abrangente que o recortado pelo Dicionário de Biblioteconomia e Arquivologia, sendo esse segundo, possível vítima das limitações locais, uma vez que foi publicado em 2008 e talvez pelas redes sociais terem um modelo muito simplificado e diferente do que existe hoje, além de outras limitações da web, o contexto digital não foi abordado na definição.

Entretanto, é possível fazer recortes nessa definição como *“armazenamento e exibição das coleções”* ou ainda *“autenticação, avaliação e categorização”* e com o mínimo de abstração é possível encaixar tais tarefas no ambiente digital, sendo assim, possível extrapolar e em caráter complementar, é possível afirmar que o curador digital é aquele que avalia, autentica, categoriza, armazena, exhibe arquivos e ou informações digitais.

Em definição mais acurada e com olhar prático, o DCC (Digital Curation Centre) que foi fundado em 2004 no Reino Unido (por isso a indicação local junto com o conceito), possui um recorte específico e aqui traduzido diretamente. De acordo com o Digital Curation Centre ([2004?]).

A curadoria digital é o gerenciamento e preservação de dados / informações digitais em longo prazo. A curadoria digital envolve manter, preservar e agregar valor aos dados de pesquisa digital em todo o seu ciclo de vida.

O gerenciamento ativo de dados de pesquisa reduz as ameaças ao seu valor de pesquisa de longo prazo e mitiga o risco de obsolescência digital. Enquanto isso, dados com curadoria em repositórios digitais confiáveis podem ser compartilhados entre a comunidade de pesquisa mais ampla do Reino Unido.

Além de reduzir a duplicação de esforços na criação de dados de pesquisa, a curadoria aumenta o valor de longo prazo dos dados existentes, tornando-os disponíveis para futuras pesquisas de alta qualidade.

Portanto, para fins didáticos, o conceito de curadoria digital que melhor compreende as possibilidades do tema exposto é do DCC, sendo utilizado até o fim da proposição desse artigo, a menos que novas situações evoquem a necessidade de nova conceituação.

Essa complexidade em definir o conceito só reafirma que Curadores Digitais

precisam ser “profissionais versáteis como um canivete suíço” (Osswald, 2013), conforme Poole (2016).

## **METODOLOGIA**

O presente artigo é resultado de pesquisa bibliográfica exploratória, preliminar, que busca discorrer sobre a definição, contexto, limitação e a aplicação da curadoria digital nas redes sociais.

O tema que nos propomos a investigar é vasto, diverso, e sua complexidade assim como outros diversos temas, ainda está sendo desenvolvida enquanto os estudos sobre estes ainda evoluem. Especialmente se considerada a perspectiva da abordagem aqui proposta, que avalia tais temáticas não só de um ponto de vista técnico, mas também de sua aplicabilidade na sociedade.

O aumento exponencial na produção de conteúdo dentro das redes sociais, seja por parte de indivíduos ou instituições, mostra que hoje temos uma fonte inesgotável, mas todo esse excesso pode gerar uma ausência de compressão, justamente pelo seu excesso.

Um claro exemplo das consequências pelo excesso de informação produzida é o paradoxo que se cria, com “fake news” sendo cada vez mais comuns e, uma sensação de incompreensão real dos fatos por parte do usuário que não sabe qual informação de fato é relevante e confiável.

Obedecer a todas as etapas do ciclo de vida da curadoria digital é um desafio e reconhecer que nem sempre é possível encontrar no mundo real, um exemplo completo e ideal como o que foi desenvolvido pela academia, até porque em diversos momentos, etapas que são claramente separadas no esquema, aparecem de forma “inseparável” durante a pesquisa.

O recorte aqui apresentado, busca contribuir para a reflexão de como a aplicação da Curadoria Digital pode beneficiar o público de forma geral, fornecendo conteúdo maior relevância, sem prejuízos de armazenamento e preservação.

Além do que já foi dito, e compreendendo que a Curadoria Digital é um instrumento que evolui com o seu tempo, temos ainda a evolução das redes sociais que não necessariamente padrões de desenvolvimento técnico, e com funcionamentos cada vez mais complexos e restrições a dados antigos e / ou em grande escala da atualidade, adicionam uma camada extra de complexidade ao tema.

## **CURADORIA DIGITAL**

Após a compreensão do conceito, é necessário ainda elencar os pontos fundamentais que compõem a curadoria digital, que segundo (Poole, 2016) são:

**Ciclo de vida da informação, Dados e Metadados e Compartilhamento e Reuso** visão possivelmente inspirada na visão de Higgins (2008).

A apresentação do modelo do “Ciclo de Vida da Informação” do DCC que possui aspectos semelhantes na compreensão das etapas, como por exemplo: **O Ciclo de Vida da Curadoria, Dados (Objetivos digitais ou Bancos de dados), Ações de Ciclo de Vida Completo, Ações Sequenciais e Ações Ocasionais.**

## 1 | CURADORIA DIGITAL: CICLO DE VIDA DA INFORMAÇÃO

Observando os pontos fundamentais de Poole, e ainda com base nos conceitos do DCC, o modelo de Ciclo de Vida da Informação desenvolvido por Sarah Higgins e exposto abaixo, compreende nas palavras de Higgins.

O modelo de ciclo de vida de curadoria do DCC fornece uma visão geral gráfica de alto nível dos estágios necessários para a curadoria e preservação bem-sucedida dos dados da conceituação inicial ou do recebimento. O modelo pode ser usado para planejar atividades dentro de uma organização ou consórcio para garantir que todas as etapas necessárias sejam realizadas, cada uma na sequência correta. O modelo permite que a funcionalidade granular seja mapeada em relação a ele; definir funções e responsabilidades; e construir uma estrutura de padrões e tecnologias para implementar. Pode ajudar no processo de identificação de etapas adicionais que podem ser necessárias, ou ações que não são exigidas por certas situações ou disciplinas, e garantindo que os processos e políticas sejam devidamente documentados. (Higgins, 2008, p. 137)

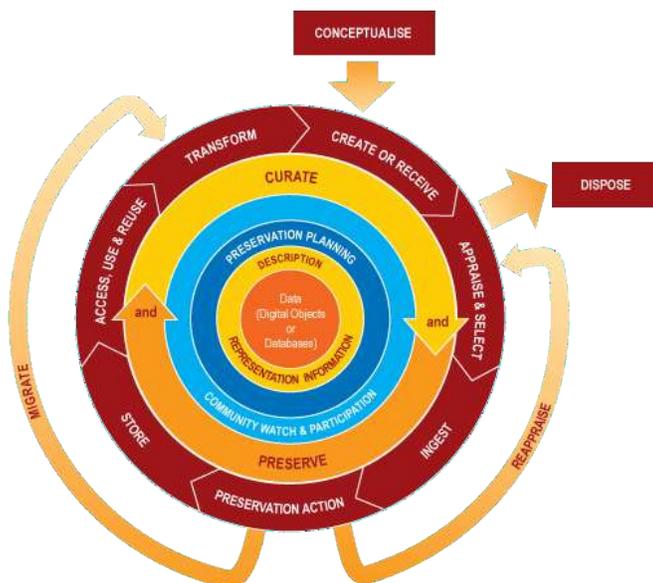


Figura 1 – Ciclo de Vida da Curadoria Digital

Fonte: Digital Curation Centre

O Ciclo de vida da Curadoria Digital pode ser dividido em duas fases. A primeira das ações sequenciais e a segunda das ações ocasionais.

De acordo com Higgins:

### **Ações Sequenciais**

**Conceituar:** conceber e planejar a criação de dados, incluindo método de captura e opções de armazenamento.

**Criar e receber:** Criar dados incluindo metadados administrativos, descritivos, estruturais e técnicos. Os metadados de preservação também podem ser adicionados no momento da criação; receber dados, de acordo com políticas de coleta documentadas, de criadores de dados, outros arquivos, repositórios ou centros de dados e, se necessário, atribuir metadados apropriados.

**Avaliar e selecionar:** Avalia os dados e seleciona para curadoria e preservação de longo prazo. Siga as orientações, políticas ou requisitos legais documentados.

**Ingerir:** Transferir dados para um arquivo, repositório, centro de dados ou outro custodiante. Siga as orientações, políticas ou requisitos legais documentados.

**Ação de Preservação:** Realizar ações para garantir a preservação e retenção a longo prazo da natureza autorizada dos dados. As ações de preservação devem garantir que os dados permaneçam autênticos, confiáveis e utilizáveis, mantendo sua integridade. As ações incluem limpeza de dados, validação, atribuição de metadados de preservação, atribuição de informações de representação e garantia de estruturas de dados ou formatos de arquivo aceitáveis.

**Armazenar:** Armazene os dados de maneira segura, aderindo aos padrões relevantes.

**Acessar, usar e reutilizar:** Certifique-se de que os dados estejam acessíveis tanto para usuários designados quanto para reutilizadores, no dia a dia. Isso pode ser na forma de informações publicadas publicamente disponíveis. Controles de acesso robustos e procedimentos de autenticação podem ser aplicáveis.

**Transformar:** Crie novos dados do original, por exemplo: Pela migração para um formato diferente; criando um subconjunto, por seleção ou consulta, para criar resultados recém-derivados, talvez para publicação.

### **Ações ocasionais**

**Eliminar:** Elimine os dados que não foram selecionados para a cura e preservação a longo prazo de acordo com as políticas documentadas, orientações ou requisitos legais. Normalmente, os dados podem ser transferidos para outro arquivo, repositório, centro de dados ou outro custodiante. Em alguns casos, os dados são destruídos. A natureza dos dados pode, por razões legais, exigir a destruição segura.

**Reavaliação:** Retorna dados que falham nos procedimentos de validação para posterior avaliação e nova seleção.

**Migrar:** migra os dados para um formato diferente. Isso pode ser feito de acordo com o ambiente de armazenamento ou para garantir a imunidade dos dados contra a obsolescência do hardware ou software. (Higgins, 2008, p. 138)

## 2 | CURADORIA DIGITAL: DADOS E METADADOS

Apontado por Poole, como um dos pontos fundamentais da curadoria digital, os dados e metadados aparecem no segundo item listado no ciclo de vida de Higgins: **Criar e Receber**. O tópico afirma que dados e metadados devem ser atribuídos ao criar e / ou receber arquivos.

Usando ainda uma compreensão simplificada de dados cunhada no (NATIONAL SCIENCE BOARD, 2005, p.9) os dados são “qualquer informação que possa ser armazenada em formato digital”.

Nesse contexto, os metadados não existem de forma independente, ao contrário do que pode ocorrer com os dados e isso implica que para o processo de curadoria digital ser praticado de forma plena em suas possibilidades técnicas e intelectuais, todos os dados criados e / ou recebidos, devem possuir metadados, que em instâncias diferentes, contribuirão ainda para a interoperabilidade, indexação e recuperação da informação no ambiente digital.

## 3 | CURADORIA DIGITAL: COMPARTILHAMENTO E REUSO

Ainda com base na premissa dos pontos fundamentais da curadoria digital de Poole, apesar de termos diferentes, os conceitos Compartilhamento e Reuso também estão presentes no esquema de ciclo de vida de Higgins. Reuso pode ser compreendido como reutilização e o compartilhamento seria a intersecção entre os conceitos de uso e transformação.

## A REDE SOCIAL COMO FERRAMENTA DE COMUNICAÇÃO INSTITUCIONAL

De forma autoexplicativa, a rede social é um componente da sociedade e não é surpreendente que em maior ou menor grau, essas ferramentas foram incluídas nas instituições públicas e privadas, de forma oficial ou às margens da falta de políticas voltadas para tais instrumentos.

O fato é que ao longo da última década podemos observar de forma privilegiada o avanço da comunicação política e eleitoral mundial nas redes sociais, principalmente no Twitter, microblogue sediado nos Estados Unidos, que permite o compartilhamento de textos com limite de até 280 caracteres, imagens, vídeos e links, além das possibilidades de responder (comentar) e retuitar (compartilhar).

A simplicidade nas formas de comunicar dentro do twitter, concede uma singular dinâmica ágil, descentralizada e não linear para as discussões dentro da ferramenta, que ainda possui um painel que mensura em tempo real os termos, ou assuntos mais comentados do momento, os chamados trends, ou em tradução direta, tendências.

Apenas para contexto histórico, em 2008, a campanha da primeira eleição presidencial de Barack Obama, que seria eleito e reeleito presidente dos EUA entre 2009

e 2017, fez uso até então inédito das redes sociais e principalmente do Twitter, na época recém fundado.

Desde então, mundo afora, as redes sociais se tornaram a principal ferramenta de marketing eleitoral, devido sua proximidade com eleitores e mesmo potencialidades publicitárias.

A parte disso, e com recorte no uso institucional das redes sociais, em 2015, foi criado o perfil @POTUS no twitter. POTUS é um acrônimo para “President of the United States”, ou Presidente dos Estados Unidos em tradução direta. E essa foi a primeira iniciativa, ou pelo menos a mais significativa, para transformar um perfil dentro de uma rede social em um instrumento institucional, que assim como o cargo de presidente, que ao fim do mandato é transferido para o próximo eleito, o perfil então, seria transferido também.

Para caráter explicativo, vale destacar que por terminologia, será usado @POTUS todas as vezes em que se fizer necessário falar do perfil no twitter, evitando assim a repetição desnecessária de twitter, já compreendido como @ aqui.

## @POTUS E A CURADORIA DIGITAL

Como já citado anteriormente, o @POTUS se tornou uma ferramenta institucional e atualmente o perfil ocupado por Joe Biden, o 46° presidente dos EUA, conforme print do perfil e link: <https://twitter.com/POTUS>



Figura 2 – Print do perfil @POTUS no Twitter

Fonte: Perfil @POTUS no Twitter

Por se tratar de um perfil institucional para um cargo eletivo, quando o mandato de

um presidente termina, é adicionado ao @POTUS o número ordinal correspondente a sua eleição. Ex: @POTUS45 é o perfil de Donald Trump preservado para registro histórico pela **Administração Nacional de Arquivos e Registros** ou Nara (acrônimo em inglês), assim como o @POTUS44 preserva a comunicação produzida por Barack Obama.



Figura 3 – Print dos perfis Potus dos mandatos anteriores

Fonte: Compilação da autora.

Compreendida a dinâmica do @POTUS e os elementos da curadoria digital e apesar de não ser possível identificar de forma clara, direta e documentada, não é absurdo ponderar sobre a possibilidade que de fato exista um processo de curadoria digital na gestão do @POTUS, mesmo que em maior ou menor grau.

No exercício dessa compreensão, é possível abstrair ainda que se um presidente possui assessores que escrevem ou ajudam com seus discursos, assim como David Litt, prestava auxílio para Barack Obama (Forbes, 2018).

É completamente racional que exista ajuda equivalente para a comunicação no Twitter, e é nesse momento que o profissional ou equipe responsável, pode e deve **“Conceituar e Criar”**.

No caso de **“Acesso e Uso”**, especificamente em rede sociais, o curador digital tempo pouco ou até mesmo nenhum poder de atuação, já que tal prerrogativa é da administração da rede social. O **“Uso”** pode ser compreendido ainda como a “entrega” desse conteúdo, mas isso também é feito via algoritmo, ficando mais uma vez limitada a ação do curador digital nessas etapas.

Com o recorte de uma rede social institucional que quase, senão diariamente produz novos conteúdos, e aqui para exercício teórico, estamos atribuindo maior ou menor “interferência” do processo de curadoria digital, são nas etapas após conceituação e criação de acordo com ciclo de vida da curadoria digital que é possível encontrar de forma concreta e prática a curadoria digital.

Para tanto, faremos um novo recorte dentro do nosso objeto @POTUS, uma vez

que os tweets do atual presidente ainda não foram arquivados pelo NARA e a conta do @POTUS45 apesar de ter sido arquivada, não possui conteúdo porque o ex-presidente Donald Trump foi banido da rede social por violação das regras, vamos analisar o @POTUS44 de Barack Obama.



Figura 4 – Print do perfil @Potus44 no Twitter

Fonte: Perfil @POTUS no Twitter

Ao clicar no link: <https://www.obamalibrary.gov/research/archived-white-house-websites-and-social-media> disponível no perfil @POTUS44, somos direcionados para o portal da biblioteca presidencial, na subpágina de arquivos de web sites e mídias sociais.

No tópico voltado às mídias sociais arquivadas, existe uma breve informação sobre o processo preservação e transferência do perfil, além dos demais perfis arquivados de um link com a lista de todos os perfis da gestão em questão que foram armazenados.

Existem ainda outros dois links direcionando para uma página que detalha as iniciativas de preservação de um projeto até então inédito de curadoria digital do conteúdo produzido nas redes sociais por um presidente e sua equipe: <https://obamawhitehouse.archives.gov/blog/2016/10/31/digital-transition-how-presidential-transition-works-social-media-age> e outra que resume as ações e lista alguns ativos digitais que serão deixados como “legado” e disponíveis para consultas futuras dos quais um em especial será traduzido diretamente aqui.

Segundo Miller (2017, <https://obamawhitehouse.archives.gov/blog/2017/01/05/new-lenses-first-social-media-presidency>)

**ArchiveSocial:** uma plataforma de arquivamento de mídia social, está hospedando um arquivo aberto consolidando mais de um quarto de milhão de postagens de mídia social da Casa Branca que são facilmente pesquisáveis

por data, plataforma e palavra-chave. O arquivo aberto agora está disponível ao público em <http://ObamaWhiteHouseArchive.social> (MILLER, 2017)

Ao clicar no link do “Archive.Social” somos direcionados para a seguinte página.

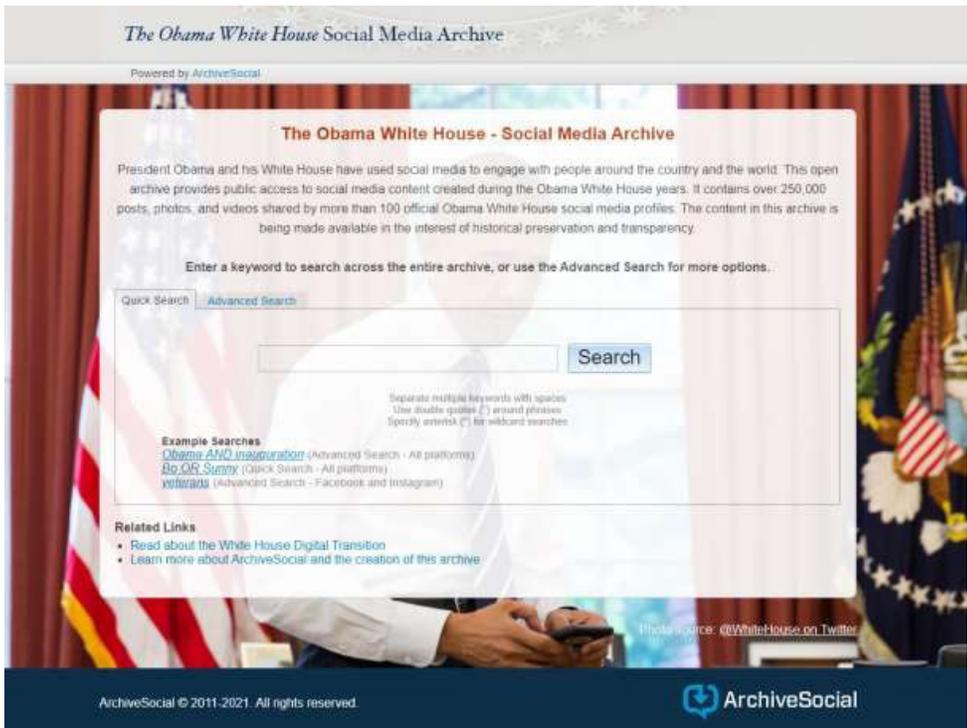


Figura 5 – Print da página <http://obamawhitehouse.gov.archivesocial.com/>

Fonte: <http://obamawhitehouse.gov.archivesocial.com/>

A interface é simples, com apenas uma barra de busca e as opções de “Pesquisa Rápida” ou “Pesquisa Avançada”, conforme destacado na imagem. Ao clicar em “Pesquisa Avançada” somos direcionados para a seguinte visão.



Figura 6 – Print da página <http://obamawhitehouse.gov.archivesocial.com/>

Fonte: <http://obamawhitehouse.gov.archivesocial.com/>

As possibilidades de pesquisa se multiplicam demonstrando a complexidade do projeto pela primeira vez. Além de 7 redes sociais possíveis de pesquisar, é possível orientar a busca por **“texto, perfil em rede social, de um perfil específico ou para um perfil específico”**. Soma-se a essa gama de opções o **“contém ou não contém”**. As pesquisas podem ser feitas em 130 perfis que foram preservados na gestão @POTUS44.



Figura 7 – Print da página <http://obamawhitehouse.gov.archivesocial.com/>  
 Fonte: <http://obamawhitehouse.gov.archivesocial.com/>

Além da busca avançada é possível fazer buscas cruzadas, combinando diversas opções disponíveis. Numa pesquisa simples pelo perfil específico @POTUS44 temos o seguinte resultado.



Figura 8 – Print da página <http://obamawhitehouse.gov.archivesocial.com/> com resultados de pesquisa  
 Fonte: <http://obamawhitehouse.gov.archivesocial.com/>

A busca trouxe resultados apenas do perfil selecionado, e como esse foi o único filtro aplicado, retornou com o histórico completo, que pode ser observado no lado esquerdo da imagem: 352 tweets e logo abaixo, mais a separação por mês e ano com as respectivas quantidades.

The Obama White House Social Media Archive

Powered by ArchiveSocial

### The Obama White House - Search Results

These results include content created by official White House social media accounts.

1 - 10 of 352 results for **Advanced Search** Sort by **Relevance** Descending

**Twitter**  
All (352)  
Tweets (352)

**Matching Dates**  
July 2015 (44)  
January 2016 (34)  
May 2015 (27)  
June 2015 (23)  
August 2015 (23)  
September 2015 (23)  
January 2017 (19)  
October 2016 (18)  
October 2015 (17)  
June 2016 (17)  
more

**Record Details**

Content type  
Twitter - Tweets

Account  
@POTUS

Version history  
outubro 6, 2016 at 8:17:43 PM GMT-03:00 (Current)

Archive status  
Digitally signed; Indexed

Figura 9 – Print da página <http://obamawhitehouse.gov.archivesocial.com/> com resultados de pesquisa metadados visíveis.

Fonte: <http://obamawhitehouse.gov.archivesocial.com/>

Ao clicar em qualquer tweet aparece ao lado a etiqueta com os detalhes do registro, ou metadados que nesse caso são em tradução direta.

**Tipo de conteúdo:** Twitter – Tweets

**Conta:** @POTUS

**Versão histórica:** outubro 6, 2016 at 8:17:43 PM GMT-03:00 (Current)

**Status do arquivo:** assinado digitalmente; indexado

Há ainda a possibilidade mais simplificada de ter acesso apenas aos arquivos do então presidente e primeira-dama no Twitter via download, e Facebook, Twitter e Vine da Casa Branca, também via download.

Exemplificando o conteúdo disponibilizado nos arquivos do @POTUS44 que no computador foi baixado como: C:\Users\Rosalia\Downloads\POTUS111716 e exibido na imagem abaixo.

Nome	Tipo	Tamanho
css	Pasta de arquivos	
data	Pasta de arquivos	
img	Pasta de arquivos	
js	Pasta de arquivos	
lib	Pasta de arquivos	
index	Microsoft Edge HTML D...	4 KB
README	Documento de Texto	2 KB
tweets	Arquivo de Valores Sepa...	99 KB

Figura 10 – Print da pasta dos arquivos disponibilizados via download

Fonte: Print da autora.

De todos os arquivos disponibilizados, 3 merecem atenção especial. Tweets trata-se de uma planilha em .csv (extensão do arquivo em planilha) com todos os tweets do @POTUS44. O index é uma versão com aprimoramento visual que pode ser aberto direto no navegador de internet e simula a interface gráfica do twitter, conforme imagem abaixo.

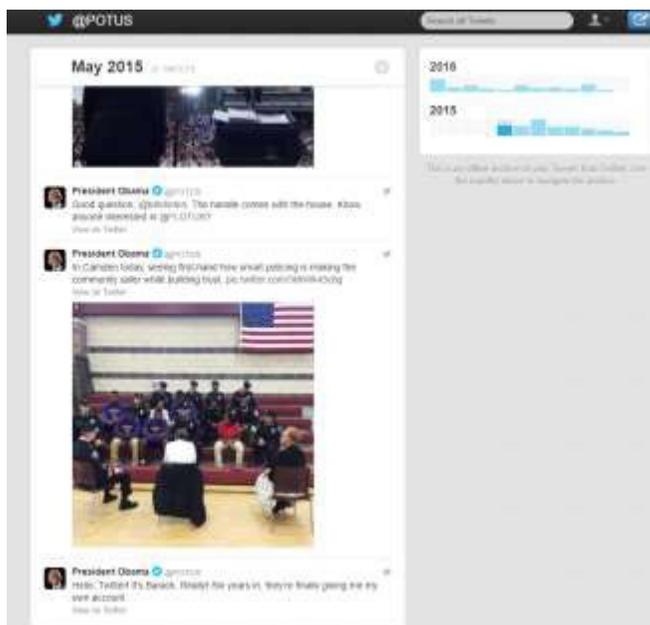


Figura 11 – Print do arquivo index que simula a interface do twitter disponibilizado via download

Fonte: Print da autora.

E o último arquivo “ReadMe” ou Leia-me em tradução direta é citado integralmente abaixo para que não ocorra perda de informação e seja exemplificado de forma prática

como os procedimentos de curadoria digital foram executados nesse recorte.

De acordo com a Biblioteca Presidencial Barack Obama, hospedada no National Archives (2017? <https://www.obamalibrary.gov/research/archived-white-house-websites-and-social-media#socialmedia>).

Como usar seu arquivo do Twitter

A maneira mais simples de usar seu arquivo do Twitter é por meio da interface do navegador de arquivo fornecida neste arquivo. Basta clicar duas vezes em `index.html` na pasta raiz e você pode navegar por todo o seu histórico de Tweets de dentro do seu navegador.

Na pasta `data`, seu arquivo do Twitter está presente em dois formatos: exportações JSON e CSV por mês e ano.

\* CSV é um formato genérico que pode ser importado para várias ferramentas de dados, aplicativos de planilha ou consumido simplesmente usando uma linguagem de programação.

JSON para desenvolvedores

\* A exportação JSON contém uma representação completa de seus Tweets conforme retornado pela v1.1 da API do Twitter. Consulte <https://dev.twitter.com/docs/api/1.1> para obter mais informações.

\* A exportação JSON também é usada para alimentar a interface do navegador de arquivos (index.html).

\* Para consumir a exportação em um analisador JSON genérico em qualquer idioma, retire a primeira e a última linha de cada arquivo. Para fornecer feedback, fazer perguntas ou compartilhar ideias com outros desenvolvedores do Twitter, participe dos fóruns de discussão em <https://dev.twitter.com>. (NATIONAL ARCHIVES, [2017?])

Com a amplitude de material exposto anteriormente conseguimos identificar de forma clara as ações sequenciais previstas no ciclo de vida da Curadoria Digital. Conforme abordado ainda por Higgins, existem etapas sequenciais que podemos compreender até como etapas de recorrência, num fluxo de retroalimentação e etapas ocasionais, que devem ser executadas, caso exista necessidade.

## CONSIDERAÇÕES FINAIS

Apesar da grande disponibilidade de dados e informações por parte do National Archives, não foi possível encontrar um documento estruturado do processo de curadoria digital, ou qualquer outro arquivo com regras e especificações.

A parte disso, através das informações colhidas e aqui expostas, é improvável que não exista uma política de curadoria digital.

Durante a investigação exploratória foi possível observar mesmo que de forma não mapeada completamente, a maior parte das etapas do ciclo de vida da curadoria digital (Higgins, 2008). Em consonância a isso, os pontos fundamentais destacados por (Poole,

2016) também são compreendidos dentro do recorte do @POTUS44.

Por mais que a preservação seja uma das etapas do ciclo de vida da curadoria digital, em alguns momentos é possível que haja confusão, e por mais que ambas as disciplinas tenham por objetivo preservação dos arquivos ao longo do tempo, a Curadoria Digital “pode ser interpretada como uma evolução do entendimento e das questões de preservação” (Santos, 2016, p. 461).

O recorte utilizado aborda uma instituição governamental que disponibilizou como acervo histórico pesquisável, todo o conteúdo produzido nas redes sociais de gestão presidencial, incluindo os integrantes de sua equipe, do momento em que foi instituída como regra até o último dia do mandato. Com estrutura e provavelmente recursos financeiros que podem ser considerados fora da realidade de muitas outras instituições, é possível que em certa medida, exista a incompreensão e o sentimento de impossibilidade de ações mais simples em ambientes com menos recursos e infraestrutura. Mas destaca-se ainda que apesar de possuir o back-up de todo material produzido em ambiente externo indexado e compatível com pesquisas, foi disponibilizado de forma muito mais “modesta”, apenas a planilha para download com o conteúdo produzido somente no Twitter, ou seja, o processo de curadoria digital deve passar também por uma análise, sobre como oferecer soluções que se adequem dentro da realidade e dos recursos da instituição em que será aplicada.

Sempre que possível, devemos preferir algum nível de curadoria digital, do que nenhum, sendo que mesmo de forma limitada, promover o acesso nas melhores condições possíveis, já é melhor que não oferecer acesso algum.

## REFERÊNCIAS

CURADORIA. In: HOUAISS, Antônio; VILLAR, Mauro de Salles. **DICIONÁRIO HOUAISS**: da língua portuguesa. Rio de Janeiro: Objetiva, 2009. p. 586.

CUNHA, Murilo Bastos da; CAVALCANTI, Cordélia Robalinho de Oliveira. Curador. In: **DICIONÁRIO de Biblioteconomia e Arquivologia**. Brasília: BRIQUET DE LEMOS Livros, 2008. p. 112.

DIGITAL CURATION CENTRE (UK). **Digital Curation**: What is digital curation? [S. l.], [2004?]. Disponível em: <https://www.dcc.ac.uk/about/digital-curation>. Acesso em: 20 fev. 2021.

OSSWALD, Achim. Skills for the Future: Educational Opportunities for Digital Curation Professionals. **DigCurV 2013**, Florença, 2013. Disponível em: <http://eprints.rclis.org/27897/>. Acesso em: 20 fev. 2021.

POOLE, Alex H. The Conceptual Landscape of Digital Curation. **Journal of Documentation**, [s. l.], v. 72, 2016. Disponível em: [https://www.researchgate.net/publication/301200534\\_The\\_Conceptual\\_Landscape\\_of\\_Digital\\_Curation](https://www.researchgate.net/publication/301200534_The_Conceptual_Landscape_of_Digital_Curation). Acesso em: 20 fev. 2021.

HIGGINS, Sarah. The DCC Curation Lifecycle Model. **The International Journal of Digital Curation**, [s. l.], v. 3, ed. 1, p. 134 - 140, 2008. Disponível em: [https://www.researchgate.net/publication/220924444\\_The\\_DCC\\_curation\\_lifecycle\\_model](https://www.researchgate.net/publication/220924444_The_DCC_curation_lifecycle_model). Acesso em: 20 fev. 2021.

HIGGINS, Sarah. The DCC Curation Lifecycle Model. **The International Journal of Digital Curation**, [s. l.], v. 3, ed. 1, p. 137, 2008. Disponível em: [https://www.researchgate.net/publication/220924444\\_The\\_DCC\\_curation\\_lifecycle\\_model](https://www.researchgate.net/publication/220924444_The_DCC_curation_lifecycle_model). Acesso em: 20 fev. 2021.

HIGGINS, Sarah. **The DCC Curation Lifecycle Model**. 2008. desenho técnico. Disponível em: [https://www.researchgate.net/publication/220924444\\_The\\_DCC\\_curation\\_lifecycle\\_model](https://www.researchgate.net/publication/220924444_The_DCC_curation_lifecycle_model). Acesso em: 20 fev. 2021.

HIGGINS, Sarah. The DCC Curation Lifecycle Model. **The International Journal of Digital Curation**, [s. l.], v. 3, ed. 1, p. 138, 2008. Disponível em: [https://www.researchgate.net/publication/220924444\\_The\\_DCC\\_curation\\_lifecycle\\_model](https://www.researchgate.net/publication/220924444_The_DCC_curation_lifecycle_model). Acesso em: 20 fev. 2021.

NATIONAL SCIENCE BOARD. Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century. **National Science Board**, [s. l.], 2005. Disponível em: <https://www.nsf.gov/geo/geo-data-policies/nsb-0540-1.pdf>. Acesso em: 20 fev. 2021.

FORBES. Como escrever um discurso presidencial. **Forbes**, [s. l.], 21 set. 2018. Disponível em: <https://forbes.com.br/colunas/2018/09/como-escrever-um-discurso-presidencial/>. Acesso em: 20 fev. 2021.

MILLER, Joshua. **New Lenses on the First Social Media Presidency**. [S. l.], 5 jan. 2017. Disponível em: <https://obamawhitehouse.archives.gov/blog/2017/01/05/new-lenses-first-social-media-presidency>. Acesso em: 20 fev. 2021.

NATIONAL ARCHIVES. Barack Obama Presidential Library. **Archived White House Websites and Social Media**. [S. l.], [2017?]. Disponível em: <https://www.obamalibrary.gov/research/archived-white-house-websites-and-social-media#socialmedia>. Acesso em: 20 fev. 2021.

SANTOS, Thayse Natália Cantanhede. Curadoria digital e preservação digital: cruzamentos conceituais. **RDBC: REVISTA DIGITAL DE BIBLIOTECONOMIA E CIÊNCIA DA INFORMAÇÃO**, Campinas, v. 14, n. 3, 2016. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8646336>. Acesso em: 20 fev. 2021.

# CURADORIA DIGITAL: MAPA DE BIBLIOTECAS NAS UNIVERSIDADES FEDERAIS BRASILEIRAS

**Priscila Ramos Carvalho,**

Doutoranda em Ciência da Informação no Programa de Pós-Graduação em Ciência da Informação do IBICT.

**Marcos Gonçalves Ramos,**

Doutorando em Ciência da Informação no Programa de Pós-Graduação em Ciência da Informação do IBICT.

**Ricardo Medeiros Pimenta,**

Pesquisador 2 do CNPq (bolsa de produtividade) e Jovem Cientista do Nosso Estado FAPERJ (2018 - 2020). Pesquisador Titular do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) do Ministério da Ciência, Tecnologia, Inovação e Comunicações (MCTIC). Professor Permanente do Programa de Pós-graduação em Ciência da Informação (PPGCI/IBICT-UFRJ).

**RESUMO:** No contexto digitalização do conhecimento, o artigo apresenta o resultado do mapeamento das bibliotecas e os serviços destinados às atividades acadêmicas de ensino, pesquisa e extensão nas 60 universidades federais brasileiras, abordando as seguintes questões: Qual o papel da biblioteca? Possui alguma política de informação? Disponibiliza serviços de incentivo à pesquisa científica? Possui o serviço de curadoria digital de dados de pesquisa? A partir da perspectiva da Ciência da Informação, escolheu-se a Informetria como base metodológica para realizar a pesquisa empírica nos websites das bibliotecas. O resultado apontou que as bibliotecas não possuem curadoria digital, porém foram identificados vários serviços de estímulo e suporte a pesquisa, demonstrando a mudança do papel da biblioteca para um centro de conhecimento e apoio à pesquisa científica.

**PALAVRAS-CHAVE:** Curadoria Digital; Biblioteca; Universidade; Dados de Pesquisa.

## DIGITAL CURATOR: MAP OF LIBRARIES IN BRAZILIAN FEDERAL UNIVERSITIES

**ABSTRACT:** In the context of digitalization of knowledge, the article presents the result of the mapping of libraries and services for academic teaching, research, and extension activities in 60 Brazilian federal universities, addressing the following questions: What is the role of the library? Do you have an information policy? Does it provide incentive services for scientific research? Do you have a digital research data curation service? From the perspective of Information Science, Informetrics was chosen as the methodological basis for conducting empirical research on the websites of libraries. The result showed that libraries do not have digital curatorship, however, several services to stimulate and support research have been identified, demonstrating the change in the role of the library to a knowledge center and support for scientific research.

**KEYWORDS:** Digital Curation; Library; University; Research data.

## INTRODUÇÃO

As universidades públicas são o principal local de produção de pesquisa de científicas, ou seja, centros de geração de conhecimento e tecnologias essenciais para o desenvolvimento econômico e competitividade de um país. A partir da expansão das Tecnologias de Informação e Comunicação (TICs), a pesquisa científica passou a produzir e disponibilizar uma quantidade extraordinária de dados, o que motivou a demanda por profissionais com competências multidisciplinares para implementar padrões e boas práticas na gestão de dados de pesquisa.

Os dados de pesquisa são itens informacionais complexos, pois abrangem dados originados por diversas fontes e atividades durante o ciclo de vida da pesquisa, desde dados coletados por instrumentos em experimentos de laboratórios, artefatos arqueológicos, genoma humano até dados astronômicos de um telescópio espacial.

A curadoria digital é um serviço que compreende o planejamento e gerenciamento de ativos digitais durante o ciclo de vida útil dos dados, envolvendo a seleção, a descrição e a preservação em longo prazo, para a reutilização dos dados em consonância com a política de guarda e descarte de arquivos digitais de cada instituição.

De acordo com Sayão e Sales (2012),

[...] a curadoria digital emerge como uma nova área de práticas e de pesquisa de espectro amplo que dialoga com várias disciplinas e muitos gêneros de profissionais. Ela une as tecnologias e boas práticas do arquivamento e da preservação digital e dos repositórios digitais confiáveis com a gestão dos dados científicos, criando uma área de pesquisa cujos desdobramentos, de amplo espectro, ainda são imprevisíveis. Isto porque, como se trata de uma área que só recentemente despontou como crítica para a pesquisa, ainda restam muitas lacunas práticas e teóricas a serem equacionadas, orientadas, preferencialmente, por uma abordagem multidisciplinar. (SAYÃO; SALES, 2012, p.185).

A curadoria digital pode ser considerada um dos resultados da digitalização de fontes de informação que começou com a reprodução do documento impresso para formatos digitais, nos anos 80 do século XX, tendo como objetivo a conversão para *bytes* (conjunto de oito bits) de maneira que a fonte de informação pudesse ser interpretada por uma máquina, transformando o documento original em um objeto digital. A digitalização acontece através do processo que envolve equipamentos (scanner planetário), programas (OCR) e formatos (PDF/A) capazes de realizar a transformação digital.

Os objetos digitais carregam dados estruturados e não-estruturados que se integram ao conceito de *big data* em termos de relações de volume, variedade, velocidade e variabilidade. O crescimento das fontes de informação tem levado a administradores públicos e privados a definir estratégias de preservação digital que por sua vez adotadas por instituições de pesquisa, devem influenciar a infraestrutura tecnológica e o custo final.

Na era digital, a governança dos dados tornou-se um tema central para o

estabelecimento de políticas de acesso aberto aos dados de pesquisa, garantindo a disponibilização de forma segura e ética de todas as fontes de dados para novos estudos, sendo assim, uma condição primordial para a realização dos objetivos da Ciência Aberta.

No cenário de globalização do conhecimento, o presente trabalho tem como objetivo analisar as bibliotecas universitárias e os serviços destinados às atividades acadêmicas de ensino, pesquisa e extensão, abordando as seguintes questões: Qual o papel da biblioteca? Existe alguma política de informação? Disponibiliza serviços em benefício da pesquisa? Possui curadoria digital de dados de pesquisa?

Em virtude da pandemia, o estudo é baseado na pesquisa realizada nos websites das 60 universidades federais brasileiras, além de revisão de literatura e trabalhos anteriores sobre bibliotecas universitárias de Cunha e Diógenes (2010) como também de Viana (2016).

## **REVISÃO DE LITERATURA**

No Brasil, as iniciativas de incentivo ao acesso aberto surgiram no início da década de 2000, como caso da SCIELO que passou a disponibilizar periódicos eletrônicos pautados nessa filosofia, bem como o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) por meio de diretrizes e ações como a plataforma de acesso aberto para a editoração de periódicos científicos e a Biblioteca Digital de Teses e Dissertações (BDTD), em cooperação com as bibliotecas universitárias, assim tornando-se uma referência de sucesso no desenvolvimento de repositórios de acesso aberto no país (CUNHA; DIÓGENES, 2016, p..118)

No contexto da Ciência Aberta, às instituições acadêmicas e agências de fomento à pesquisa passaram a valorizar e dar importância aos dados como fontes de recursos informacionais que devem ser preservados (SAYÃO, SALES, 2013). Em 31 de outubro de 2017, a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) anunciou que os pedidos de financiamento para projetos temáticos, aqueles com duração de cinco anos que se destacam por seus objetivos ousados, deveriam apresentar o plano de gestão de dados, ou seja, um documento complementar de no máximo duas páginas, descrevendo os dados que seriam produzidos pelos projetos, como seriam disponibilizados e preservados.

Nesse sentido, o plano de gestão de dados e o acesso aberto aos dados de pesquisa produzidos nas universidades brasileiras contribuem para objetivo da Ciência Aberta, isto é, tornar a ciência mais acessível, confiável e eficiente por meio da validação de métodos e a reprodutibilidade dos resultados da pesquisa, propiciando a formação de redes e trocas de conhecimentos entre pesquisadores.

No ambiente da digitalização, as bibliotecas digitais tornam-se uma importante ferramenta no centro da atividade intelectual, ampliando a expectativa de armazenamento estático e recuperação de informações para serem facilitadoras da geração do conhecimento

por meio da comunicação, da colaboração e da interação entre cientistas, pesquisadores e o público em geral, sobre temas pertinentes às informações armazenadas (CANDELA *et al.*, 2007).

A automação das bibliotecas universitárias brasileiras evoluiu a começar pela adoção de sistemas baseados em mainframes, a experiência com os sistemas operativo ISIS, sigla em inglês de *Intel System Implementation Supervisor*, passando para softwares livres e o uso tecnologias como RFID, sigla em inglês de *Radio-Frequency IDentification*, até as ferramentas de automação baseadas em nuvem, possibilitando a integração de registros bibliográficos do acervo físico com os recursos eletrônicos acessíveis na internet (VIANA, 2010, p.80).

As bibliotecas universitárias têm um papel importante na preservação da memória e do patrimônio científico do país, portanto, cada vez mais precisam adequar seus serviços, infraestrutura e equipe (perfil do bibliotecário) aos desafios da era digital e dos grandes volumes de dados de pesquisa. Nessa linha, a curadoria digital seria um serviço essencial, uma vez que é o conjunto de atividades para o gerenciamento e preservação de dados digitais em longo prazo, abrangendo o planejamento, a digitalização de documentos físicos e a garantia da disponibilidade dos dados em formatos adequados para a reutilização no futuro. (ABBOTT, 2008).

O *Digital Curation Center* (DCC) propõe um modelo de ciclo de vida no âmbito da curadoria digital como forma de assegurar a continuidade em longo prazo dos objetos digitais, retratado na figura 1:

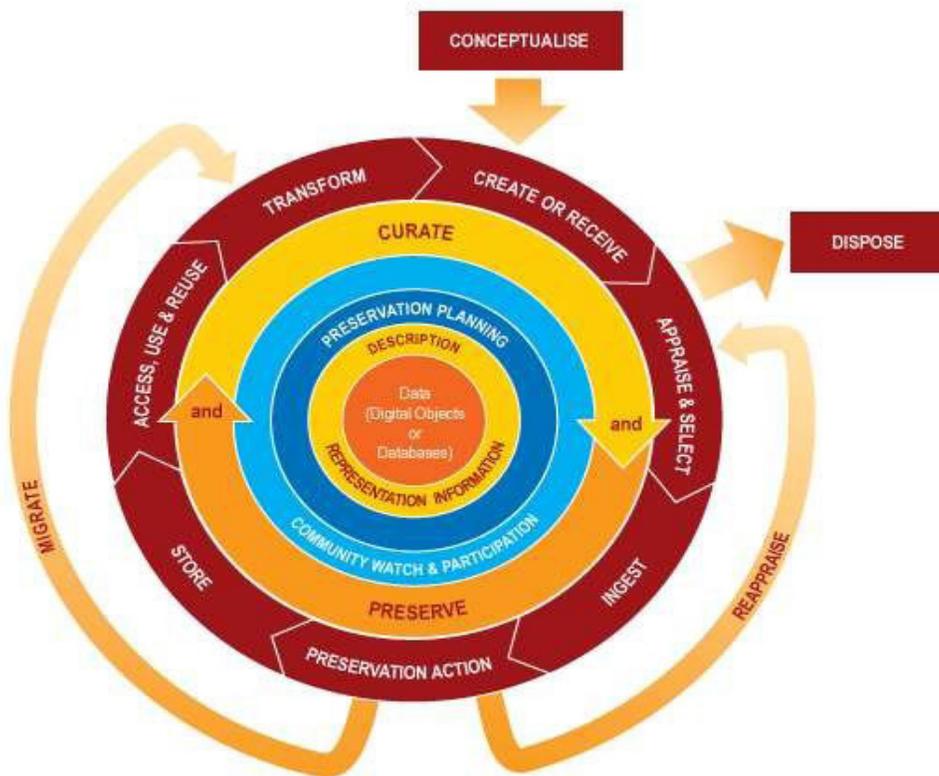


Figura 1: Modelo do Ciclo de Vida da Curadoria

Fonte: *The DCC Curation Lifecycle Model*<sup>1</sup>

Higgins (2008, pp.136) explica que o modelo de ciclo de vida proposto no DCC não é definitivo e precisa evoluir, ou seja, adaptar-se às mudanças e a contextos específicos. O ciclo de vida ajuda os curadores a entender os processos envolvidos e as melhores práticas de curadoria digital, visando o desenvolvimento da curadoria, da preservação e de metodologias para organizações.

Cabe pontuar sobre as principais ações da curadoria digital, as quais estão próximas ao centro do ciclo de vida proposto pelo DCC e permeiam quatro atuações: 1) descrição e representação da informação: seria o momento de atribuição dos metadados descritivos, estruturais, administrativos e de preservação segundo padrões pré-estabelecidos; 2) planejamento da preservação: ocorre durante todo ciclo de vida; 3) monitoramento e participação da comunidade: seriam as contribuições e adequações às expectativas e necessidades da sociedade seja na parte técnica de acesso ou inserção de novos metadados; 4) curadoria e preservação: seria a realização das atividades contínuas necessárias durante todo ciclo de vida com objetivo de preservar os dados.

<sup>1</sup> Lifecycle Model - DCC. Disponível em: < <https://www.dcc.ac.uk/guidance/curation-lifecycle-model> > Acesso em: 20 out. 2020

Diversas questões são pertinentes na implementação da curadoria digital, em particular a conscientização de instituições e pesquisadores sobre o acompanhamento de todo o fluxo de dados durante o ciclo de vida, ou seja, a necessidade de monitoramento constante e avaliações periódicas para adição de valor aos dados e divulgação, a fim de favorecer o uso e reuso para criação de novas funções (PALETTA; GONÇALVES, 2016, p.56).

Siebra e Silva (2017) ressaltam que além do modelo do DCC, criado pela *University of Edinburgh*, descrito por Higgins (2008), existem outros exemplos de curadoria digital, tais como: *Digital Curation Unit* (DCU), *UK Data Archive* (2010), *DataOne* (2012), Ciclo de Vida de Dados (CVD-CI, 2013) de Sant'Ana e *Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval* (CASPAR, 2014).

Na esfera da padronização das atividades de preservação digital, o modelo *Open Archival Information System* (OAIS) tem sido equiparado à norma ISO 14.721 de 2003 (SOUZA *et al.*, 2012). Entre as proposições sobre o tema convém destacar as seguintes instituições: o consórcio internacional chamado *Inter-University Consortium for Political and Social Research* (ICPSR) em companhia de *Data Observation Network for Earth*, patrocinado pela *National Science Foundation* (NSF), e o *UK Office for Library and Information Science Networking* (UKON), que posteriormente originou o projeto *Research 360*, o qual pretende cobrir todo o ciclo de vida da pesquisa em qualquer área do conhecimento.

Vale acrescentar que o ciclo de vida da curadoria digital acompanha o ciclo de vida da pesquisa e seus itens informacionais contribuindo para os Sistemas de Referenciamento de Pesquisas Correntes (CRIS), sigla em inglês de *Current Research Information System*, cujo objetivo é permitir o acesso às informações relevantes, possibilitando oportunidades de financiamento e a divulgação dos resultados da pesquisa. Nesse sentido, a curadoria de dados de pesquisa e os Sistemas de Referenciamento de Pesquisa Corrente são processos que se complementam a partir da biblioteca como um centro de produção de conhecimento, visto que universidades públicas brasileiras são historicamente as instituições de pesquisa mais amplas e reconhecidas no país.

## **METODOLOGIA**

Diante da importância do compartilhamento de dados de pesquisa, o estudo de caráter empírico-descritivo teve o objetivo analisar as bibliotecas e seus serviços destinados às atividades acadêmicas de ensino, pesquisa e extensão nas 60 universidades federais brasileiras, abordando as seguintes questões: Qual o papel da biblioteca? Possui alguma política de informação? Disponibiliza serviços de incentivo à pesquisa científica? Possui o serviço de curadoria digital de dados de pesquisa?

Tomou-se como base metodológica a Informetria, que é o estudo dos aspectos quantitativos da informação em qualquer formato, e não apenas em registros catalográficos

ou bibliografias, referente a qualquer grupo social, e não apenas aos cientistas. (MACIAS-CHAPULA, 1998, p. 135).

A pesquisa buscou compreender o papel da biblioteca no processo de acesso à informação, a existência de uma política voltada para preservação e a disponibilização de dados de pesquisa, bem como a comunicação e a divulgação científica da produção acadêmica da universidade, diante dos desafios de *big data* e da virada digital das universidades brasileiras.

Nessa linha, delineamos os critérios para analisar as bibliotecas: qual o número de bibliotecas do Sistema de Bibliotecas e Informação (SIBI); quais os serviços disponibilizados pelo SIBI; qual o número de revistas científicas das instituições e se estão ligadas às bibliotecas; qual o perfil do repositório da instituição e se está ligado a biblioteca (institucional e/ou dados de pesquisa); e se existe algum serviço em auxílio a pesquisa científica, em especial a curadoria digital de dados de pesquisa.

Em virtude de a curadoria digital ser um serviço fundamental para biblioteca se posicionar como um centro de conhecimento e de suporte à pesquisa científica nas universidades, buscamos verificar se existem grupos de pesquisa já envolvidos em nome dessa mudança na perspectiva acadêmica no Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), tendo o resultado retratado no quadro 1.

Instituição	Grupo	Data de Criação	Líder	2º Líder	Área Predominante
Instituto Brasileiro de Informação em Ciência e Tecnologia	Base Bibliográfica Referencial da Ciência Brasileira (BReCiBr)	04/12/2019	Washington Luís Ribeiro de	-	Ciências Sociais Aplicadas/ Ciência da Informação
Universidade Tecnológica Federal do Paraná	Inovação, Desenvolvimento e Aplicação de Tecnologias Digitais na Educação	20/12/2019	Iolanda Bueno de Camargo Cortelazzo	Marcus Vinicius Santos Kucharski	Ciências Humanas/ Educação
Instituto Brasileiro de Informação em Ciência e Tecnologia	Estudos e Práticas de Preservação Digital	04/02/2020	Miguel Ángel Márdero Arellano	-	Ciências Sociais Aplicadas/ Ciência da Informação
Instituto Brasileiro de Informação em Ciência e Tecnologia	BRIET: Biblioteconomia, Recuperação de Informação, E-Science e suas teorias	03/04/2020	Luana Farias Sales Marques	Luis Fernando Sayão	Ciências Sociais Aplicadas/ Ciência da Informação

Fundação Casade Rui Barbosa	Grupo de Pesquisade Tecnologias e Comunicação em Instituições de Memória (GPTICIM)	19/05/2020	Ana Ligia Silva Medeiros	-	Ciências Sociais Aplicadas/ Ciência da Informação
Universidade Federal de Pernambuco	Núcleo de Curadoria Digital	20/07/2020	Sandra de Albuquerque	Thais Helen do Nascimento Santos	Ciências Sociais Aplicadas/ Ciência da Informação
Universidade Federal do Tocantins	Informação, Comunicação e Memória	25/08/2020	Francisco Gilson Rebouças Pôrto Junior	-	Ciências Sociais Aplicadas/ Ciência da Informação
Universidade Federal de São Carlos	Dados e Metadados	28/08/2020	Ana Carolina Simionato Arakaki	-	Ciências Sociais Aplicadas/ Ciência da Informação
Universidade Federal do Espírito Santo	Tecnologia e Organização da Informação e do Conhecimento	07/09/2020	Daniela Lucas da Silva Lemos	-	Ciências Sociais Aplicadas/ Ciência da Informação
Universidade de São Paulo	Observatório do Mercado de Trabalho do Profissional da Informação na Era Digital	17/10/2020	Francisco Carlos Paletta	Waldomiro de Castro Santos Vergueiro	Ciências Sociais Aplicadas/Ciência da Informação
Universidade Federal Fluminense	Grupo CNPq UFF Ged/A - Documentos Digitais: Gestão, Curadoria Digital, Preservação, Acesso e Transparência Ativa em Cadeia de Curadoria Digital Arquivística (CCDA)	24/10/2020	Daniel Flores	Sérgio Ricardo da Silva Rodrigues	Ciências Sociais Aplicadas/ Ciência da Informação
<b>Total de registros: 11</b>					

Quadro 1: Resultado da Busca de Grupos de Pesquisa sobre Curadoria Digital

Fonte: Resultado de busca no Diretório de Grupos de Pesquisa do CNPq em 27/10/2020

O resultado da busca apresentou 11 grupos criados a partir de dezembro de 2019,

demonstrando assim que o tema curadoria digital é novo e emergente no campo da Ciência da Informação com 10 grupos de pesquisa no CNPq.

## RESULTADO E DISCUSSÃO

Apresentamos a seguir o resultado e as considerações sobre a pesquisa realizada nos websites das 60 universidades federais brasileiras divididas por região com intuito de permitir a visualização das possíveis diferenças no Brasil.

### 1 | UNIVERSIDADES DA REGIÃO NORTE

O resultado mostrou o total de 69 bibliotecas na região Norte, distribuídas em nove universidades federais, com destaques para UFAM com 13 (18,84% do total) e UFPA com 14 (20,29% do total), retratado no gráfico 1.

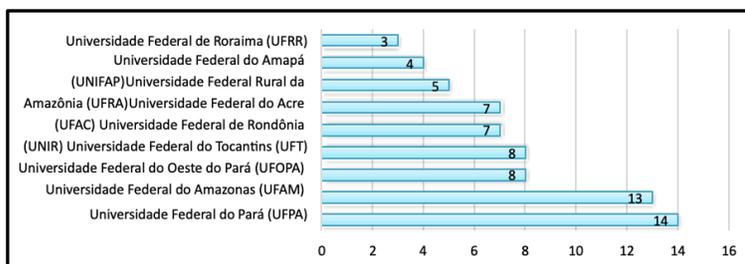


Gráfico 1: N° de Bibliotecas do SIBI por Universidade – Região Norte

Fonte: criação nossa com base na pesquisa de websites das universidades

Já em relação às revistas científicas o total foi de 147 unidades, com destaque para UFAM com 36 (24,49% do total), conforme o gráfico 2. É interessante perceber que os periódicos científicos produzidos pelas universidades nem sempre possuem destaque nos websites das bibliotecas, bem como não há uma aparente relação entre o número de bibliotecas e número de revistas científicas, apesar da divulgação científica e metrias da informação serem abordadas nos cursos de Biblioteconomia e Ciência da Informação.

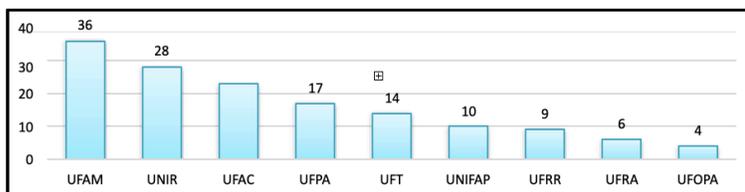


Gráfico 2: N° de Revistas Científicas por Universidade – Região Norte

Fonte: criação nossa com base na pesquisa de websites das universidades

O resultado dos serviços apontou que as bibliotecas disponibilizam políticas diferentes para biblioteca e repositório. Nem todas as universidades possuem repositório, como é o caso da UFOPA onde o repositório está em construção.

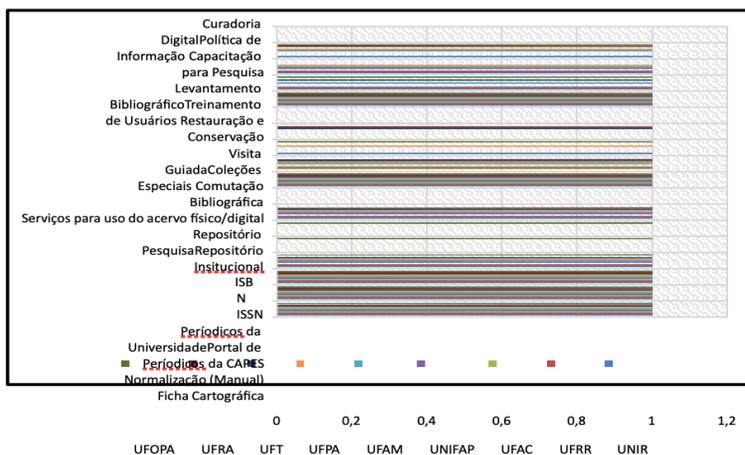


Gráfico 3: Serviços das Bibliotecas por Universidade – Região Norte

Fonte: criação nossa com base na pesquisa de websites das universidades

Os pontos negativos identificados foram: nenhuma das universidades faz restauração; nenhuma possui repositório de dados de pesquisa e nem serviço de curadoria digital de dados de pesquisa. Apesar das universidades disponibilizarem serviços em favor da pesquisa como, por exemplo, a UFAM que realiza treinamentos do Mendeley, pesquisa em bases de dados (Scopus, Web of Science, outros) e visualização de informações científicas com *software* de mineração de dados (VosViewer), bem como análise de grafos e gráficos.

Já os pontos positivos foram os seguintes: o uso do recurso de banners e botões na home permitindo o acesso direto aos periódicos, repositórios e outros serviços do SIBI, assim como a capacitação de usuários para pesquisa. Compete mencionar que apenas a UFOPA deixou claro no website o serviço de ISBN e ISSN.

## 2 | UNIVERSIDADES DA REGIÃO NORDESTE

O resultado apresentou o total de 154 bibliotecas na região Nordeste, distribuídas em quinze universidades federais, com maior relevância para UFBA com 22 (14,29% do total) e UFRN com 24 (15,58% do total), conforme o gráfico 4.

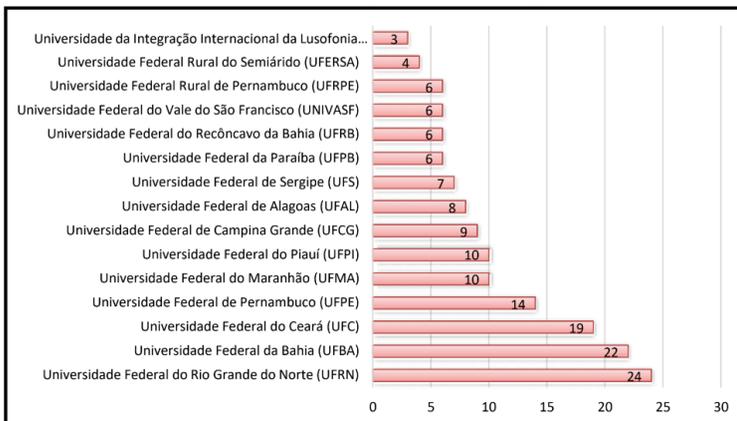


Gráfico 4: N° de Bibliotecas do SIBI por Universidade – Região Nordeste

Fonte: criação nossa com base na pesquisa de websites das universidades

O total de revistas científicas foi 430, com ênfase para UFPB com 63 (14,65% do total), UFBA com 59 (13,72% do total) e UFPI com 56 (13,02% do total), retratado no gráfico 5.

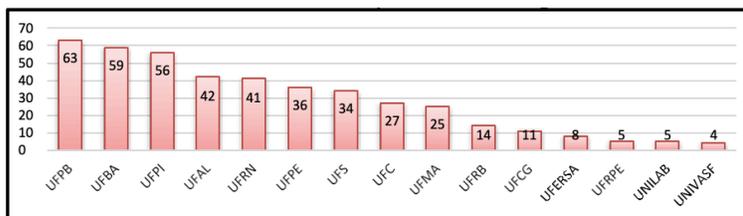


Gráfico 5: N° de Revistas Científicas por Universidade – Região Nordeste

Fonte: criação nossa com base na pesquisa de websites das universidades

O resultado dos serviços apontou que a maioria das bibliotecas disponibiliza pelo menos uma política relacionada às coleções, ao acervo, à biblioteca e ao repositório. Porém, não há uma política de informação que englobe todas as questões envolvendo a informação produzida, armazenada e disponibilizada por meio do acesso aberto nas universidades.

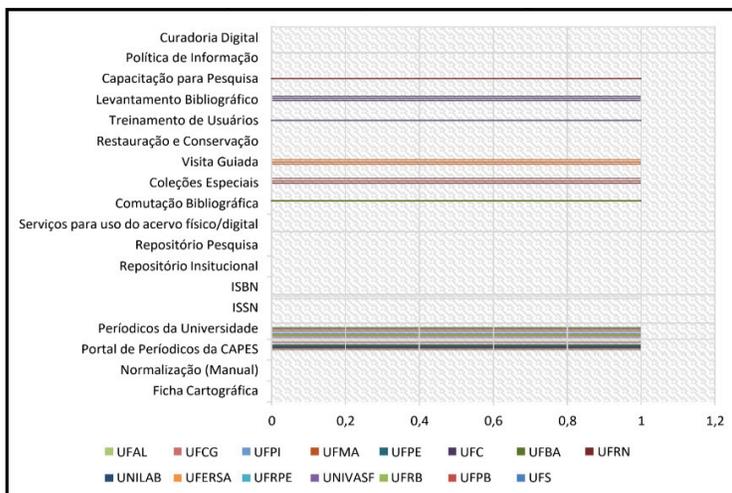


Gráfico 6: Serviços das Bibliotecas por Universidade – Região Nordeste

Fonte: criação nossa com base na pesquisa de websites das universidades

Os pontos negativos detectados foram: a dificuldade em achar informações nos websites de bibliotecas; nenhuma possui repositório de dados de pesquisa ou curadoria digital; nem o serviço de restauração ou conservação explícito no website.

Os pontos positivos foram: os serviços a favor da acessibilidade de deficientes como caso do núcleo de acessibilidade (UFBA), o repositório de informação acessível (UFRN) e o serviço de levantamento bibliográfico para pessoas com deficiência visual (UFC), assim como o centro de inclusão digital com cursos de capacitação (UFAL).

Notou-se que muitas universidades disponibilizam aplicativos para celular com objetivos diferentes, mas demonstrando a capacidade de adaptação às necessidades da sociedade (UFC, UFPE, outros). Além disso, a maioria das bibliotecas utilizam as redes sociais como canal de comunicação com seus usuários, como também o uso de novas mídias como caso do podcast (UFC) e do canal no YouTube com treinamentos (UNIVASF, UFERSA, outros).

### 3 | UNIVERSIDADES DA REGIÃO CENTRO-OESTE

O resultado revelou o total de 27 bibliotecas na região Centro-Oeste, distribuídas em cinco universidades federais, com notoriedade para a UFMS com 10 (37,04% do total), representado no gráfico 7.

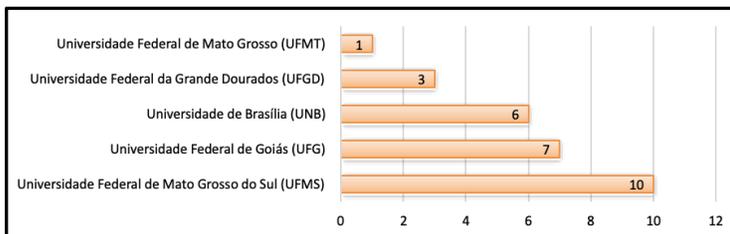


Gráfico 7: N° de Bibliotecas do SIBI por Universidade – Região Centro-Oeste

Fonte: criação nossa com base na pesquisa de websites das universidades

O resultado das revistas científicas foi igual a 187, com destaque para UNB com 76 (40,64% do total), conforme o gráfico 8.

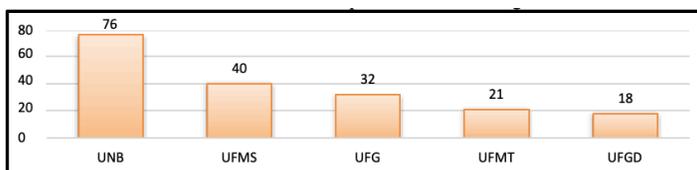


Gráfico 8: N° de Revistas Científicas por Universidade – Região Centro-Oeste

Fonte: criação nossa com base na pesquisa de websites das universidades

O resultado dos serviços sinalizou que os websites das bibliotecas não expõem toda gama de serviços prestados. O acesso à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) é mais visível do que às revistas científicas das instituições, bem como a divulgação de bases de dados do acesso ao repositório institucional.

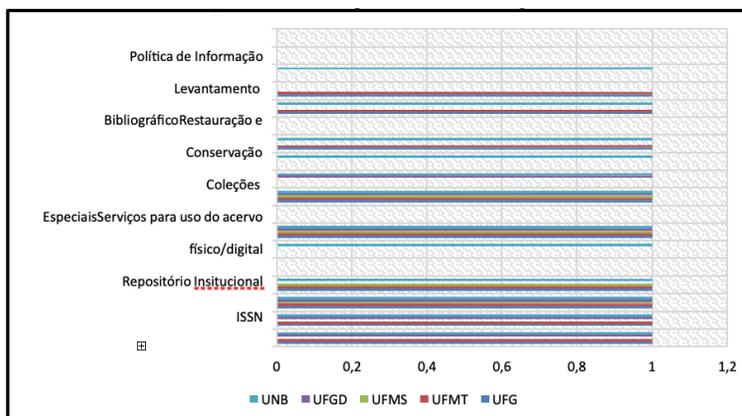


Gráfico 9: Serviços das Bibliotecas por Universidade – Região Centro-Oeste

Fonte: criação nossa com base na pesquisa de websites das universidades

Os pontos negativos percebidos foram: nenhuma possui repositório de dados de pesquisa e nem curadoria digital, bem como poucas mencionam uma política ou serviços em defesa da pesquisa científica. Já os pontos positivos foram: o uso de redes sociais pelas bibliotecas como canal de comunicação com usuários, assim como para divulgação dos serviços e acervo.

## 4 | UNIVERSIDADES DA REGIÃO SUDESTE

O resultado indicou o total de 238 bibliotecas na região Sudeste, distribuídas em dezenove universidades federais, com relevâncias para a UNIFESP com 44 (18,49% do total) e a UFRJ com 45 (18,91% do total), conforme o gráfico 10.

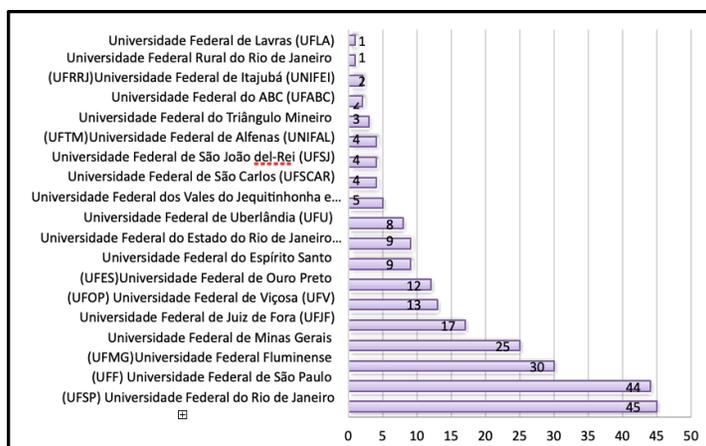


Gráfico 10: N° de Bibliotecas do SIBI por Universidade – Região Sudeste

Fonte: criação nossa com base na pesquisa de websites das universidades

O resultado total das revistas científicas foi igual a 625, com destaque para UNIFESP com 193 (30,88% do total), representado no gráfico 11.

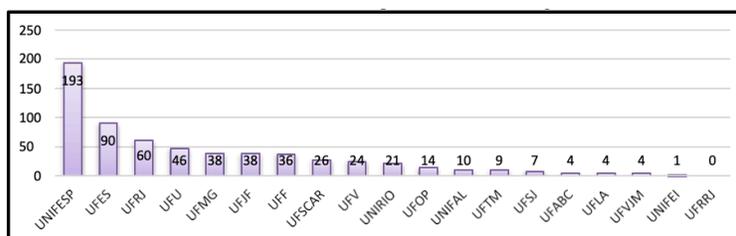


Gráfico 11: N° de Revistas Científicas por Universidade – Região Sudeste

Fonte: criação nossa com base na pesquisa de websites das universidades

O resultado dos serviços demonstrou que bibliotecas destacam a questão da digitalização por meio da divulgação do scanner planetário (UFU, UFJF, UFRRJ, UNIFAL, outros), como também foram identificados dois repositórios de dados de pesquisa, a saber: a UNIRIO sinalizou o repositório DATA Hórus na política de acesso aberto à informação de 2018, mas o mesmo ainda está em construção; e a UFSCAR disponibiliza manual de auto depósito de dados de pesquisa e cita a importância do plano de gestão de dados de pesquisa para garantir a gestão e o compartilhamento dos dados, em virtude de exigências da Fundo de Amparo à Pesquisa do Estado de São Paulo (FAPESP), que desde 2017 exige de seus pesquisadores o plano de gestão de dados e o depósito dos dados de pesquisa em um repositório de acesso aberto.

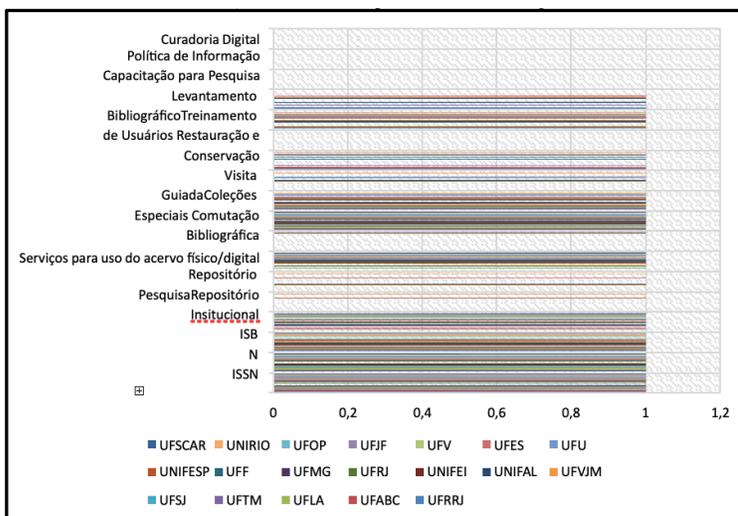


Gráfico 12: Serviços das Bibliotecas por Universidade – Região Sudeste

Fonte: criação nossa com base na pesquisa de websites das universidades

Foram observados alguns pontos negativos: a dificuldade de encontrar informações sobre os serviços, periódicos e repositórios das instituições nos websites das bibliotecas, bem como não possuir serviço de curadoria digital. Os pontos positivos foram os seguintes: disponibilidade de cursos de capacitação do Mendeley, Web of Science, Endnote, Scopus, Orcid e Journal Finder (UFJF), inclusive muitos treinamentos no YouTube (UFES), assim como serviços para deficientes como o caso das instruções de uso da biblioteca em libras (UFPEL).

É pertinente citar a adaptação das bibliotecas ao contexto de pandemia por meio de serviços e atendimento remoto, como também a disponibilização de informação qualificada sobre a covid-19, com destaque para o repositório institucional da biblioteca da UFLA que criou uma coleção de artigos científicos para auxiliar a pesquisa sobre a questão.

## 5 I UNIVERSIDADES DA REGIÃO SUL

O resultado sinalizou o total de 129 bibliotecas na região Sul, distribuídas em onze universidades federais, com destaque para UFRGS com 30 (23,26% do total), retratado no gráfico 13.

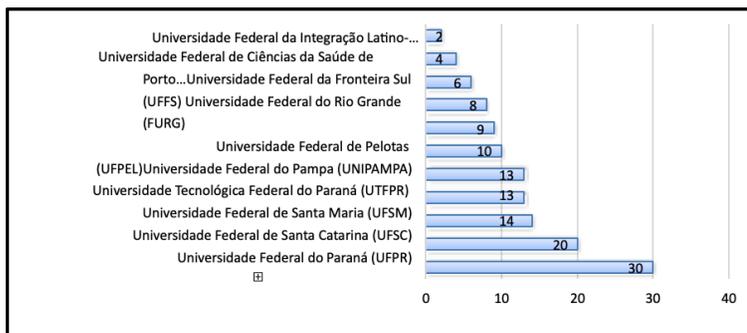


Gráfico 13: N° de Bibliotecas do SIBI por Universidade – Região Sul

Fonte: criação nossa com base na pesquisa de websites das universidades

Já o número total de revistas científicas foi de 340, com destaque para UFRGS com 85 (25% do total), conforme o gráfico 14.

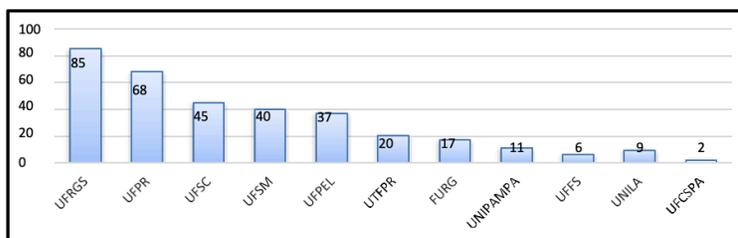


Gráfico 14: N° de Revistas Científicas por Universidade – Região Sul

Fonte: criação nossa com base na pesquisa de websites das universidades

O resultado dos serviços mostrou que nem todas as bibliotecas disponibilizam uma política de informação, seja da biblioteca ou do repositório.

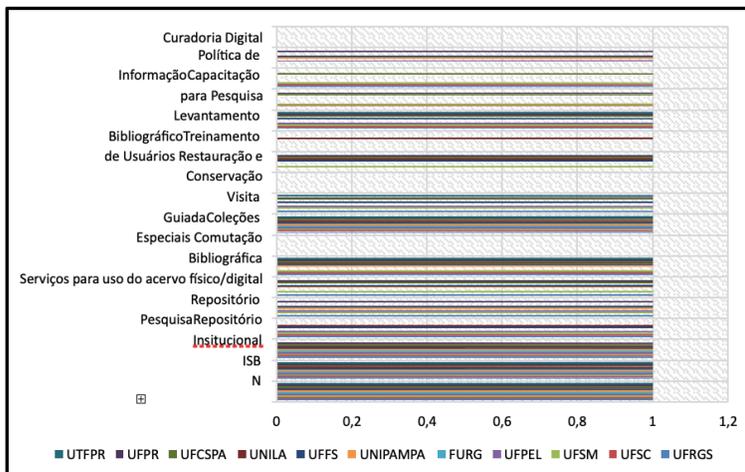


Gráfico 15: Serviços das Bibliotecas por Universidade – Região Sul

Fonte: criação nossa com base na pesquisa de websites das universidades

Alguns pontos negativos foram reconhecidos, a saber: nem sempre o repositório institucional está vinculado à biblioteca (UFPEL, FURG, outros), nem às revistas científicas das universidades (FURG, UNIPAMPA, outros). Os pontos positivos foram os seguintes: a adoção das redes sociais e o uso de canais no YouTube para cursos de capacitação e/ou treinamentos como, por exemplo, o Mendeley, ABNT, APA (UFSC). Vale ressaltar que a UNIPAMPA disponibiliza política de informação e de metadados do repositório.

## CONSIDERAÇÕES FINAIS

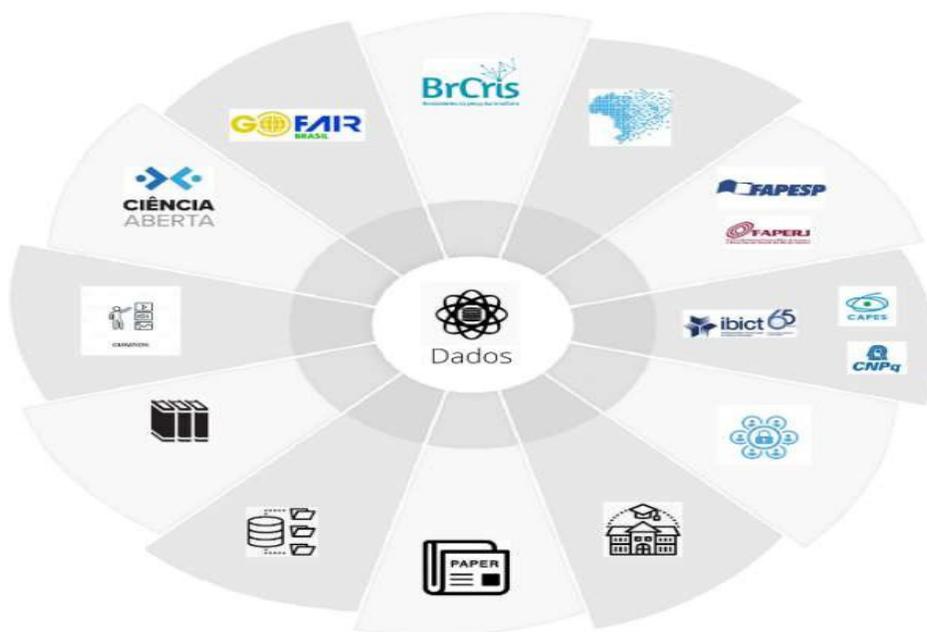
O resultado da pesquisa empírica-descritiva deixou claro que não há um padrão dos websites e nem dos serviços prestados pelas bibliotecas universitárias, portanto cada uma dá destaque ao que acha mais importante.

Convém pontuar a discrepância entre o número de universidades, bibliotecas, revistas científicas e infraestrutura de repositórios nas regiões brasileiras, o que possivelmente reflete no nível educacional, o acesso à informação e a competência crítica em informação da população.

Muitas bibliotecas perceberam a mudança do seu papel em prol da virada digital da pesquisa científica por meio de serviços de apoio a pesquisa (cursos, treinamentos, outros), a sinalização da implantação de repositórios de dados de pesquisa, porém não disponibilizam ainda o serviço de curadoria digital de dados de pesquisa e falta a definição de uma política de informação da universidade que abarque todas as questões relacionadas à informação e aos dados de pesquisa como a preservação, armazenamento, divulgação, acesso, reutilização, direitos de propriedade intelectual, proteção de dados pessoais, prazos de embargo, outros.

É relevante mencionar que as bibliotecas estão cada vez mais se adaptando ao mundo digital por meio de redes sociais, canais no YouTube, aplicativos de celular, digitalização do acervo, capacitação digital de usuários e inclusive na pandemia disponibilizaram informações qualificadas sobre covid-19 e atendimento remoto.

No contexto da produção científica nacional, o serviço de curadoria digital de dados de pesquisa e o bibliotecário de dados são fundamentais para lidar com a amplitude dos recursos e serviços que compõem o novo ecossistema de dados científicos em benefício da pesquisa brasileira, retratados no diagrama na figura 2, a saber: a) Universidades, b) Bibliotecas, c) Repositórios, d) Agências de fomento (FAPESP, FAPERJ, outros), e) Instituições de apoio à pesquisa (CAPES, CNPq, IBICT, Fiocruz, outros), f) Lei de Proteção de Dados Pessoais, g) Lei de Acesso à Informação, h) Ciência Aberta, i) BRCRIS, j) GoFAIR Brasil, l) Curadoria Digital, m) Comunicação e Divulgação Científicas.



**Figura 2:** Diagrama do Ecossistema de Dados Científicos

**Fonte:** criação dos autores

A complexidade do ecossistema de dados científicos provenientes de pesquisas acadêmicas gera novos serviços como a curadoria digital de dados de pesquisa e a adaptação dos bibliotecários para um modelo voltado à dados (bibliotecários de dados), isto é, um perfil profissional multidisciplinar que envolve habilidades com linguagens de programação, desde a coleta, o tratamento, a análise até a divulgação científica, bem como conhecimentos sobre armazenamento de dados, aplicação de metadados visando

a interoperabilidade, assim possibilitando a recuperação eficiente de dados por diversos sistemas, propiciando o reuso de dados de pesquisa e até mesmo a reprodutibilidade de pesquisas (KENNAN, 2016).

Nesse sentido, é possível inferir que os primeiros passos em prol da mudança do papel da biblioteca para um centro de conhecimento e suporte a pesquisa científica já foi realizado por instituições como a FAPESP por meio do reconhecimento da importância da gestão adequada dos dados e a necessidade do plano de gestão de dados de pesquisa, assim como o IBICT a partir da divulgação de diretrizes para repositórios de dados e publicações científicas.

## REFERÊNCIAS

ABBOTT, D. *What is digital curation?* Edinburgh: **Digital Curation Centre**, 2008.

CANDELA, L. *et al.* Setting the Foundation of Digital Libraries. **D-Lib Magazine**, v. 13, n. 3/4, Mar. Apr. 2007.

CUNHA, M. B.; DIÓGENES, F.C.B. A trajetória da biblioteca universitária no Brasil no período de 1910 a 2010. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, vol. 21, núm. 47, sep./dic., pp. 100-123, 2016.

CNPQ. Diretórios de Grupos de Pesquisa. Disponível em: <http://lattes.cnpq.br/web/dgp>. Acesso em: 27 out. 2020.

DIGITAL CURATION CENTRE (DCC). Disponível em: <http://www.dcc.ac.uk/>. Acesso em: 20 out. 2020.

FAPESP. Gestão de dados. Disponível em: <https://fapesp.br/gestaodedados>. Acesso em: 20 out. 2020.

HIGGINS, S. The DCC Curation Lifecycle Model. **International Journal of Digital Curation**, Edinburgh (UK), v. 1, n. 3, p.134-140, jan. 2008.

IBICT. Ibict disponibiliza documentos relacionados a diretrizes para repositórios de dados e de publicações científicas. Disponível em: <https://ibict.br/sala-de-imprensa/noticias/item/2490-ibict-disponibiliza-documentos-relacionados-a-diretrizes-para-repositorios-de-dados-e-de-publicacoes-cientificas>. Acesso em: 27 out. 2020.

KENNAN, M. A. Data Management: Knowledge and skills required in research, scientific and technical organisations. In: IFLA General Conference and Assembly: **IFLA WLIC 2016** (pp. 1-10).

MACIAS-CHAPULA, C. O papel da infometria e da cienciometria e sua perspectiva nacional e internacional. **Ci. Inf.**, Brasília, v. 27, n. 2, p. 134-140, maio/ago. 1998.

PALETTA, F. C.; GONÇALVES, V. J. da S. Curadoria Digital: o papel das bibliotecas na sociedade em rede. **Pesq. Bras. em Ci. da Inf. e Bib.**, João Pessoa, v. 11, n. 2, p. 047-058, 2016.

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para a preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, v. 22, n. 3, p. 179-191, set./dez. 2012.

SAYÃO, L. F.; SALES, L. F. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Tendências da Pesquisa Brasileira em Ciência da Informação**. v. 6, n.1, 2013.

SIEBRA, S. de A.; SILVA, F. de M. O. Análise de modelos de ciclos de vida para curadoria de objetos digitais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017, Recife. **Anais eletrônicos**...Recife: ENANCIB, 2017.

SOUZA, A. H. L. R. de; OLIVEIRA, A. F. de; D'AVILA, R. T; CHAVES, E. S. S. O modelo

de referência OAIS e a preservação digital distribuída. **Ci. Inf.**, Brasília, DF, v. 41 n. 1, p.65- 73, jan./abr., 2012

VIANNA, M. M. M. Uma breve história da automação de bibliotecas universitárias no Brasil e algumas perspectivas futuras. **RICI: R.lbero-amer. Ci. Inf.**, ISSN 1983-5213, Brasília, v. 9, n. 1, p. 43- 86, jan./jun.2016.

# GESTÃO DA INFORMAÇÃO: DESAFIOS E COMPETÊNCIAS PARA PRESERVAÇÃO DIGITAL

**Angélica Cintra Fermann,**

Programa de Pós-Graduação em Ciência da Informação ECA/USP

**Francisco Carlos Paletta,**

PPGCI ECA/USP

**RESUMO:** O objetivo deste estudo é realizar análises e reflexões sobre gestão da informação (GI), preservação digital e as competências do profissional da informação, especificamente o Bibliotecário. A pesquisa realizada é de natureza qualitativa e de caráter exploratório, com o propósito de propiciar às pesquisadoras uma maior familiaridade com o problema apontado no estudo e construir reflexões a partir de um conhecimento ampliado quanto aos desafios impostos na gestão da informação e no âmbito da preservação digital, buscando evidências nos temas estudados. O estudo inicialmente aborda o conceito de gestão da informação (GI), os desafios do *Big Data* e o ciclo da informação. Depois explicita e analisa políticas, procedimentos e estratégias de preservação de documentos digitais. Mais adiante explicita que perante um cenário tão desafiador, onde o objeto informação dialoga preferencialmente com as Tecnologias Digitais, a escolha do profissional com as competências e habilidades desejadas passa a ser uma estratégia para o sucesso do modelo adotado. Ao se compreender a necessidade de integrar vários saberes, desenvolvendo um conjunto de conhecimentos e habilidades (competências), o bibliotecário, profissional escolhido como objeto da pesquisa deve estar preparado para atender ao mercado de preservação digital. Nesta etapa da pesquisa ainda não há achados que permitam traçar uma conclusão segura, mas uma questão é evidente – formar profissionais da informação competentes para os enfrentamentos de uma economia digital é um intento a ser considerado pelas universidades que pretendem se engajar na transformação digital.

**PALAVRAS-CHAVE:** Gestão da informação; Curadoria Digital; Preservação Digital; Profissionais de Informação Digital.

## INFORMATION MANAGEMENT: CHALLENGES AND SKILLS FOR DIGITAL PRESERVATION

**ABSTRACT:** The objective of this paper is to make an analysis and reflection about the Information Management (IM), digital preservation and the skills of the information professional, especially, the Librarian. This research is from qualitative nature and exploratory character, with the purpose of providing to researchers a better familiarity with the problem pointed in this study and build reflections based on an expanded knowledge about the imposed challenge by the information management and by the scope of the digital preservation, pursuing evidence of the studied subjects. The paper, initially, addresses the concept of the Information

Management (IM), the challenges of Big Data and the information cycle. Secondly, explicit and analyzes politics, proceedings, and strategies of preservation of digital documents. Further, explicit that beyond a so challenging scenario, where the information object dialogues preferentially with the Digital Technologies, the chosen of the professional with the desire skill and abilities becomes a strategy for the success of the adopted model. When realized the need of the integration of multiples knowledge, developing a set of skills, the librarian, professional chosen as the studied object, must be prepared to answer the market of digital preservation. On this stage of the research there are no findings yet that could trace a safe conclusion. However, the question is evident - educate information professionals with the capacity to face the challenges of a digital economy is an attitude to be consider for the universities that intend to engage in the digital transformation.

**KEYWORDS:** Information resources management; Digital curation; Digital preservation; Digital information professional.

## INTRODUÇÃO

Este artigo nasceu no âmbito da disciplina de Curadoria Digital, a partir do desejo das pesquisadoras de buscar fundamentação teórica para seus projetos de pesquisa. Assim, seu objetivo é realizar análises e reflexões sobre gestão da informação (GI), preservação digital e as competências do profissional da informação, especificamente o Bibliotecário.

Conceituar a GI não é uma tarefa fácil e com o propósito de esclarecer o significado do termo gestão da informação, a seção 3, Gestão da Informação Digital analisa a pesquisa exploratória realizada com base em diversos autores.

Mas não é suficiente fazer a gestão da informação digital, sem uma maior compreensão das estratégias possíveis. Objetos são criados, transformados e distribuídos em uma multiplicidade de formatos e mídias, e o ciclo da obsolescência de *hardware*, *software* e formatos de documentos digitais é cada vez mais rápido.

Assim, na seção 4, Preservação de Documentos Digitais, serão abordadas algumas das estratégias possíveis de serem adotadas pelas instituições e organizações para viabilizarem a preservação de seus documentos digitais.

Um pilar estratégico para a condução do processo de gestão da informação está centrado no indivíduo e, portanto, as competências dos profissionais da informação no entorno da gestão da informação digital serão objeto da seção 5, Profissional da Informação na Curadoria Digital.

As competências deste profissional da informação ainda estão em processo de construção, demandadas pelo mercado de trabalho, neste momento em que a sociedade do sec. XXI está apenas se lançando em sua jornada de transformação digital e cujas habilidades técnicas e pessoais serão determinadas a partir dos resultados colhidos. Percorrer experiências de sucesso e identificar modelos que ainda se encontram em processo de construção, nos permite ganhar uma certa vantagem no que diz respeito a

assegurar que o indivíduo ainda em formação, na universidade, seja lapidado com saberes que estejam conectados com a demanda dos setores de uma Sociedade ávida por tais competências.

Este projeto é ainda incipiente no que diz respeito a conclusões e reformulações de modelos de apoio ao desenvolvimento de um plano organizacional, estratégias ou diretrizes, mas abre a porta para um novo olhar sobre a perspectiva de líderes que atuem na gestão da informação e com processos de preservação digital. Descortina o campo, ainda a ser muito explorado, da empregabilidade do profissional de informação, enquanto bibliotecário, quando se pensa em preservação digital documental.

## **METODOLOGIA**

Baseada nos objetivos propostos, esta pesquisa é de natureza qualitativa e de caráter exploratório, com o objetivo de propiciar maior familiaridade das pesquisadoras com o problema apontado anteriormente e construir reflexões a partir de um conhecimento ampliado quanto aos desafios impostos na gestão da informação e no âmbito da preservação digital, buscando evidências nos temas da gestão da informação, preservação do documento digital e competências da profissão do profissional da informação digital.

Desta forma, para a elaboração deste artigo foi realizado o levantamento bibliográfico e documental e analisados alguns estudos de caso que pudessem fornecer maior compreensão do objeto da pesquisa, como forma de analisar os fatos do ponto de vista empírico, com o propósito de confrontar a visão teórica com os dados da realidade no Brasil.

Inicialmente foram identificadas na bibliografia disponibilizada pela disciplina, os estudos e artigos que melhor dialogavam com o tema escolhido. A partir destes foi realizado um primeiro estudo exploratório, com a finalidade de proporcionar uma familiaridade com a área de pesquisa, bem como delimitar seus campos. Na sequência, já com o tema mais refinado, foi realizada a pesquisa e identificação de obras complementares junto ao Sistema Integrado de Bibliotecas da Universidade de São Paulo (USP), à BRAPCI (Base de Dados em Ciência da Informação) e ao *Google Academic*.

## **GESTÃO DA INFORMAÇÃO DIGITAL**

Esta seção propõe-se a abordar o conceito de gestão da informação (GI), os desafios do *Big Data*, o ciclo da informação e destacar questões dos desafios para preservação digital, que serão aprofundadas na seção 4.

A literatura contempla uma variedade de abordagens sobre GI, o que muitas vezes, provoca certa confusão e dificulta o seu entendimento (COLUMBIÉ, 2007; MIDDLETON, 2002), principalmente por estar também associada a outros tipos de gestão que também

estão relacionados com informação, como por exemplo gestão de documentos, gestão de sistemas de informação, gestão de dados entre outros.

Para Middleton (2002), as diferenças de abordagens ocorrem em função de diferentes situações e aplicações práticas, tais como: ambiente de negócios; ambiente computacional; ambiente de biblioteca; e ambiente dos arquivistas e gerentes de registros.

Assim, com o propósito de esclarecer o significado do termo gestão da informação, Detlor (2010), apresenta três perspectivas para GI: organizacional, biblioteca e pessoal. Na perspectiva organizacional, que é a mais difundida e popular, apresenta-se um foco voltado ao alcance dos objetivos competitivos e estratégicos da organização, por meio do gerenciamento de todos os processos de informação e seu ciclo de vida. Segundo Detlor (2010) são muitos os recursos informacionais geridos na organização, tais como: transações armazenadas em banco de dados; informações resumidas em *data warehouses*; e informações não estruturadas. Nesta perspectiva da organização, a GI apresenta-se frequentemente associada aos seguintes termos: gestão de sistemas de informação; gestão da tecnologia de informação; gestão de dados; inteligência de negócios; inteligência competitiva; gestão de registros; e gestão de conteúdo.

Na perspectiva da biblioteca, a GI está centrada no fornecimento de acesso aos recursos e serviços informacionais, principalmente o gerenciamento de gestão de coleções como livros e periódicos. Os termos associados a esta perspectiva são: classificação; organização do conhecimento; biblioteca digital; catalogação; índices e sistemas de recuperação.

A GI na perspectiva pessoal refere-se a gestão do ciclo de vida da informação no nível individual, onde pessoas criam, adquirem, organizam, armazenam e usam a informação para os propósitos pessoais do dia a dia, tais como: coisas a fazer; projetos; calendários; compromissos e outros.

Na visão de Detlor é possível estabelecer um conceito de GI comum às três perspectivas, onde os processos e sistemas necessários em cada etapa do ciclo de vida da informação são gerenciados de forma alinhada ao objetivo da GI que seria “ajudar as pessoas e organizações no acesso, processo e uso da informação de forma eficiente e efetiva” (DETLOR, 2010, p.103).

Por sua vez, Ponjuán Dante, propõe o conceito de GI como sendo um processo estratégico aplicado a qualquer tipo de organização, incluindo bibliotecas, de forma a englobar todos os seus processos, atividades e componentes, relacionados ao sistema que os governa (PONJUÁN DANTE, 2011), como a obtenção e utilização de recursos básicos, onde a gestão de ciclo de vida destes recursos apresenta-se como elemento fundamental (PONJUÁN DANTE, 2007) para um ganho na performance organizacional.

A origem da GI está diretamente ligada à área da Documentação, pois segundo Monteiro e Duarte, foi esse campo que introduziu os primeiros instrumentos e técnicas para lidar com o excesso de informação, fenômeno decorrente da popularização do livro

e periódicos do século XIX: “a Documentação envolvia as atividades de organização, armazenamento, preservação, recuperação e acesso à informação, o que condiciona a percepção dessa disciplina como fundadora da moderna gestão da informação” (MONTEIRO; DUARTE, 2018, p.91).

Desde o final do século XX o desafio da documentação digital traz à tona a necessidade de discussão em torno da problemática da preservação da informação no campo da GI, sob uma perspectiva sistêmica de informações registradas tanto em suportes físicos como digitais, englobando todo o ciclo de gestão do fluxo infocomunicacional e a produção informacional (PINTO, 2018). Os fluxos de informação que percorrem todo sistema organizacional, devem ser objeto de gestão, pois tais processos podem apresentar particularidades em diferentes tipos de organização. Entretanto, em geral é possível identificar as seguintes dimensões: o ambiente; os processos; as pessoas; a tecnologia; a infraestrutura e produtos ou serviços (PONJUÁN DANTE, 2007).

A preocupação com a preservação da informação torna-se mais complexa com o fortalecimento do fenômeno *Big Data* no século XXI. O termo *Big Data* é frequentemente associado a ideia de grande volume de dados (DAVENPORT, 2014; KAISLER *et al.*, 2014) “...tão grande que não cabem em contêiner usuais, tão desestruturados que não funcionam num banco de dados de linhas e colunas, ou o fluxo é tão contínuo que não serve num data warehouse estático” (DAVENPORT, 2014, p.1). Verifica-se a necessidade de software especial, além de computadores e *hardwares* dedicados para conseguir extrair, manipular e analisar o conjunto de dados grande demais para ser processado rotineiramente (SFETCU, 2019). Para um pesquisador, baixar um conjunto de dados do *Big Data* para seu computador seria inviável. A execução, *accountability*, preservação, reuso e reprodutibilidade dos dados são os novos desafios que requerem novas ferramentas, análise em nuvem e muitas vezes o trabalho colaborativo em equipe (CROSAS *et al.*, 2015).

Alguns dos desafios relacionados a aquisição e armazenamento do *Big Data* foram apontados por Kaisler *et. al* (2014), tais como o descompasso entre o aumento da velocidade de comunicação de rede e largura da banda que não acompanharam a capacidade de desempenho do processador e capacidade do disco de armazenamento, o problema do rastreamento da procedência da informação, a dificuldade de realização da limpeza dos dados. Estes são desafios que necessitam de um alto nível de gestão para solucioná-los.

Floridi (2010) destaca a importância de uma gestão eficiente do ciclo de vida da informação, exibido na Figura 1, pois “...apenas muito recentemente o progresso e o bem-estar humanos começaram a depender principalmente do gerenciamento bem-sucedido e eficiente do ciclo de vida das informações.”(FLORIDI, 2010, p.4).



Figura 1: Ciclo de vida da informação – Modelo de Floridi

Fonte: Adaptado de Floridi (2010)



Figura 2: Ciclo de vida da informação – Modelo de Ponjuán Dante

Fonte: Adaptado de Ponjuán Dante (1998)

Nota-se que o ciclo de vida da informação mencionado anteriormente por Detlor

(2010), finaliza com a etapa de uso da informação. O mesmo acontece com o modelo de Ponjuán Dante (1998), exibido na Figura 2. Porém, o modelo de Floridi (2010) apresenta a etapa de reciclar/eliminar, ou seja, uma decisão de destinação final da informação, a partir da qual podemos inferir que a preservação precede a reciclagem.

Middleton, já havia observado certa confusão dos profissionais da informação com relação ao ciclo de vida da informação, entre a informação que pode ser transmitida indefinidamente e os meios de comunicação que podem chegar a um fim. Todavia, o autor lembra que o documento que contém a informação possui uma destinação final, se destruídos ou mantidos indefinidamente. Decidida a manutenção, para o documento físico tradicionalmente pensamos nas bibliotecas e arquivos como locais para o seu armazenamento (MIDDLETON, 2002). Mas e para o documento digital? Como e em que local realizar a manutenção por tempo indefinido?

## **PRESERVAÇÃO DE DOCUMENTOS DIGITAIS**

Como observou Arellano, os “documentos digitais são considerados, atualmente, registros oficiais e são gerenciados segundo leis e padrões que compreendem todo o ciclo de vida desses materiais” (ARELLANO, 2004, p.25). Mas não é suficiente fazer a gestão da informação digital. Criados, transformados e distribuídos em uma multiplicidade de formatos e mídias, a perda de informação e de documentos devido a danos ao suporte físico (temperatura, umidade, poluição, ameaças biológicas), uso indevido, catástrofes, acabam revelando que mídias e suportes são um meio frágil e instável de armazenamento (ARELLANO, 2004). O ciclo da obsolescência de hardware, software e formatos de documentos digitais é cada vez mais rápido, o que torna mandatório “garantir o acesso contínuo a seus conteúdos e funcionalidades, por meio de recursos tecnológicos disponíveis à época em que ocorrer a sua utilização” (CONSELHO NACIONAL DE ARQUIVOS, 2003, p.2).

Entretanto, para garantir esse acesso contínuo é necessário haver o estabelecimento de políticas, procedimentos e estratégias de preservação de documentos digitais “que minimizem os efeitos da fragilidade e da obsolescência de hardware, software e formatos e que assegurem, ao longo do tempo, a autenticidade, a integridade, o acesso contínuo e o uso pleno da informação” como salientado como salientado na Carta para a Preservação do Patrimônio Arquivístico Digital (CONSELHO NACIONAL DE ARQUIVOS, 2003), documento publicado pelo Conselho Nacional de Arquivos baseado na Carta para a Preservação do Patrimônio Digital (*UNITED NATIONS EDUCATIONAL SCIENTIFIC AND CULTURAL ORGANIZATION*, 2003).

Nesta seção serão abordadas algumas das estratégias possíveis de serem adotadas pelas instituições e organizações para viabilizarem a preservação de seus documentos digitais. Esta é uma tarefa complexa, devido às características do meio digital e à natureza

dos objetos a serem preservados. Partindo da definição de que objeto digital é “aquele que foi criado em computador, podendo ser original ou uma versão depois de ter sido convertido (ou digitalizado)” (ARELLANO, 2004, p.16), a preservação de documentos digitais será aqui considerada a partir do conceito de Margaret Hedstrom, que define preservação como “o planejamento, a alocação de recursos e, a aplicação de métodos de preservação e tecnologias necessárias para assegurar que a informação digital de valor contínuo permaneça acessível e utilizável” (ARELLANO, 2004, p.17).

## 1 | ESTRATÉGIAS E MÉTODOS DE PRESERVAÇÃO

Iniciando pela perspectiva dos métodos de preservação, a elaboração de normas para armazenamento e compartilhamento adequado de objetos digitais é defendida por Arellano (2004, 2012) em seus estudos. Ele também recomenda a adoção de ferramentas que protejam os documentos digitais e garantam sua manutenção, as quais servem para

reparar e restaurar os registros protegidos (preservação retrospectiva), antecipando danos e reduzindo os riscos dos efeitos naturais (preservação prospectiva). (...). [Portanto,] as condições básicas à preservação digital seriam, então, a adoção desses métodos e tecnologias que integrariam a preservação física, lógica e intelectual dos objetos digitais. (ARELLANO, 2004, p. 17).

Preservação	Alvo
Física	Mídia de armazenamento
Lógica	Capacidade de leitura
Intelectual	Integridade e autenticidade da informação, leiaute

Tabela 1: Preservação física, lógica e intelectual

Fonte: Autoras

A preservação física dos objetos digitais foca na questão das mídias de armazenamento, abrangendo desde mídias magnéticas (como fitas cassete e DAT) até discos rígidos, passando pelos discos óticos (CD-ROM, por exemplo). A preservação lógica “procura na tecnologia formatos atualizados para inserção dos dados (correio eletrônico, material de áudio e audiovisual, material em rede etc.), novos *software* e *hardware*” (ARELLANO, 2004, p.17) que conservem a capacidade de leitura dos bits dos objetos digitais. A preservação intelectual objetiva garantir a integridade e autenticidade da informação, envolvendo também a preservação de leiaute.

As estratégias de preservação digital também devem se preocupar com a definição

de requisitos, que

especificam os elementos que deverão ser efetivamente mantidos, a série de bits que deverá ser recuperada; referem-se à capacidade de acesso aos conteúdos dinâmicos independentemente da sua apresentação, componentes de multimídia, hipertextualidade e interatividade. Observar essas condições significa identificar o objeto digital na sua origem e pelas suas dependências de *hardware e software* (ARELLANO, 2004, p.18).

Segundo Formenton, “este conjunto de requisitos (...) objetivam principalmente proteger e preservar por longo prazo a capacidade de acesso de forma utilizável do conteúdo presente nos objetos digitais, refletindo assim os próprios objetivos da preservação digital” (FORMENTON; GRACIOSO; CASTRO, 2015, p177).

De acordo Arellano (2004), os principais métodos recomendados para a preservação dos objetos digitais podem ser agrupados em dois tipos: os estruturais e os operacionais. Os métodos estruturais tratam dos investimentos iniciais por parte das instituições que estão se preparando para implementar algum processo de preservação e que adotam ou adaptam um dos modelos de metadados existentes ou seu próprio esquema. Os métodos operacionais são as medidas concretas aplicadas aos objetos digitais. Os métodos mais frequentemente utilizados estão elencados na Figura 3, abaixo.

Métodos Estruturais	Métodos Operacionais
Adoção de padrões	Conservação de <i>software/ hardware</i>
Elaboração de normas	Migração de suporte
Metadados de preservação digital	Conversão de formatos
Montagem de infra-estrutura	Emulação
Formação de consórcios	Preservação do conteúdo

Figura 3: Métodos de preservação digital

Fonte: Arellano, 2004, p.18

A adoção de padrões é baseada em quatro pilares:

1. Definição de um conjunto limitado de formatos para armazenar os dados e informações;
2. Utilização de padrões correntes para criar objetos digitais;
3. Monitoramento dos padrões adotados conforme se modificam;

#### 4. Migração para novos padrões uma vez estabelecidos.

A elaboração de normas, bem como de manuais ou guias permite solucionar ou amenizar os diferentes problemas que aparecerão durante o processo de preservação digital, pois oferecem soluções gerais em relação ao tratamento de objetos digitais e a gerência dos riscos envolvidos (FORMENTON; GRACIOSO; CASTRO, 2015). Alguns dos mais adotados são *Recomendações para a gestão de documentos de arquivos eletrônicos* (Portugal); *Curl Exemplars in Digital Archives Project* (CEDARS - Inglaterra); *Preserving and Accessing Networked Documentary Resources of Australia* (PANDORA- Austrália); e a Carta para a Preservação do Patrimônio Arquivístico Digital (Conselho Nacional de Arquivos – CONARQ - Brasil). (FORMENTON; GRACIOSO; CASTRO, 2015, p. 179).

A criação e uso de metadados é parte importante das estratégias operacionais de preservação digital, sendo um meio para garantir a autenticidade, registrar o gerenciamento de direitos e coleções de dados e para interagir com os recursos de busca. Eles são meio para armazenar a informação técnica que suporta a preservação dos objetos digitais e indicar a sua localização. Entretanto, “a implementação de um esquema de metadados de preservação requer no mínimo a limitação do escopo dos metadados para preservação, a fim de maximizar o potencial para sua automação e garantir o controle da mudança nos metadados” (ARELLANO, 2004, p.19). Além disso, deve-se implementar um padrão, criar repositórios de preservação e definir o script que extrairá os metadados de preservação.

Dentre as estratégias operacionais, a conservação de *software/hardware* pode ser importante para casos particulares ou históricos de *hardware ou software*, pois preserva a tecnologia envolvida na criação do objeto. Entretanto, possui várias desvantagens como custos elevados, necessidade de espaço e disponibilidade de suporte técnico, com mão de obra especializada e de alto custo.

A migração é uma das estratégias mais empregada pelas instituições. Consiste na transferência periódica dos dados e informações de uma determinada configuração de *hardware* e *software* para outra. Seu propósito é preservar a integridade dos objetos e assegurar sua recuperação, visualização e utilização apesar da mudança de tecnologia. Assim, objetiva preservar o conteúdo de forma utilizável, mas pode não preservar apresentação, funcionalidade e contexto do objeto digital. É uma estratégia eficiente para informações simples e homogêneas, mas ainda precisa evoluir com relação a objetos complexos. Os metadados são essenciais na migração, pois registram a história da migração do objeto digital. É uma estratégia que permite a recuperação de dados por meio da extração dos dados digitais de uma mídia obsoleta para outra acessível atualmente.

A estratégia de emulação implica na criação de novo *software* que imita o funcionamento do antigo *hardware* ou *software*. Esta estratégia preserva o físico e conteúdo, além das características originais e funcionalidades por isso é útil quando existe a relevância da aparência dos recursos digitais. Porém, tem custo elevado porque precisa

do desenvolvimento de técnicas de encapsulamento de documentos, seus metadados, *software* e especificações do emulador para assegurar a coesão e prevenir sua corrupção, e depende da habilidade de pessoas de Tecnologia da Informação para fazer a emulação, rodar em futuros computadores e recriar o comportamento do documento digital. Esta é uma estratégia que permite a reconstrução dos documentos por meio da interpretação dos documentos digitais originais usando as informações de formato de arquivo.

## 5 | MODELO DE PRESERVAÇÃO OAIS

O modelo *Open Archival Information System* (OAIS), publicado pelo *Consultive Committee for Space Data Systems* (CCSDS) é um modelo de referência para a preservação digital e uma iniciativa ISO (*International Organization for Standardization*), desde junho 2003, a ISO 14721. Apesar de não ser uma implementação específica, o OAIS fornece uma lista de condições do que deve ser considerado no estabelecimento de um projeto de preservação de qualquer documento, físico ou digital.

O modelo OAIS, sugere a criação de um “pacote de informação” formado de “informação de conteúdo” e “informação de descrição de preservação”, unindo assim os metadados ao objeto digital. Nele prevê-se a atuação de quatro entidades: produtores de informação, consumidores de informação, a administração do processo e o arquivo propriamente dito. E quatro categorias de objetos de informação, que podem ser visualizadas na Figura 4:

1. informação de conteúdo, que contém o objeto digital e a representação da informação necessária para transformar este objeto em informação com significado;
2. informação de descrição de preservação, descrição esta necessária para preservar adequadamente a informação do conteúdo a ela associada. Pode incluir informação de referência, de contexto, de proveniência ou de integridade;
3. informação de pacote, que vincula e relaciona os componentes de um pacote;
4. informação descritiva, que permite localizar, analisar, recuperar ou ordenar a informação de pacote.

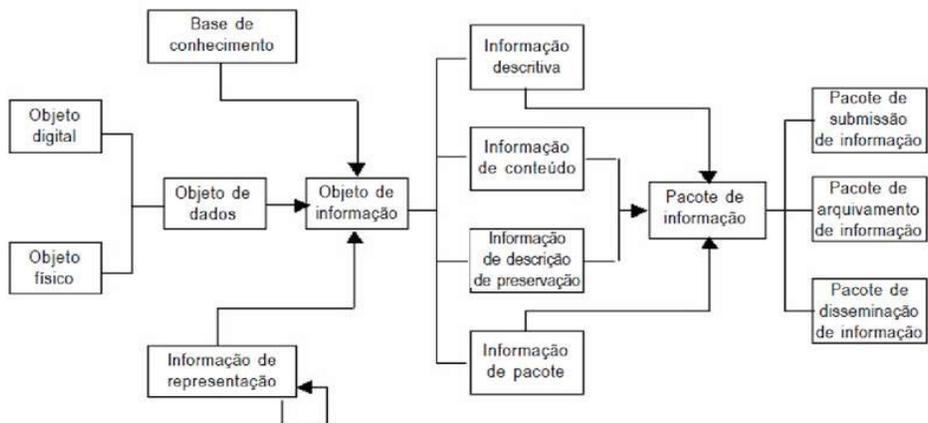


Figura 4: o modelo de referência OAIS

Fonte: (ARELLANO, 2004, p.20).

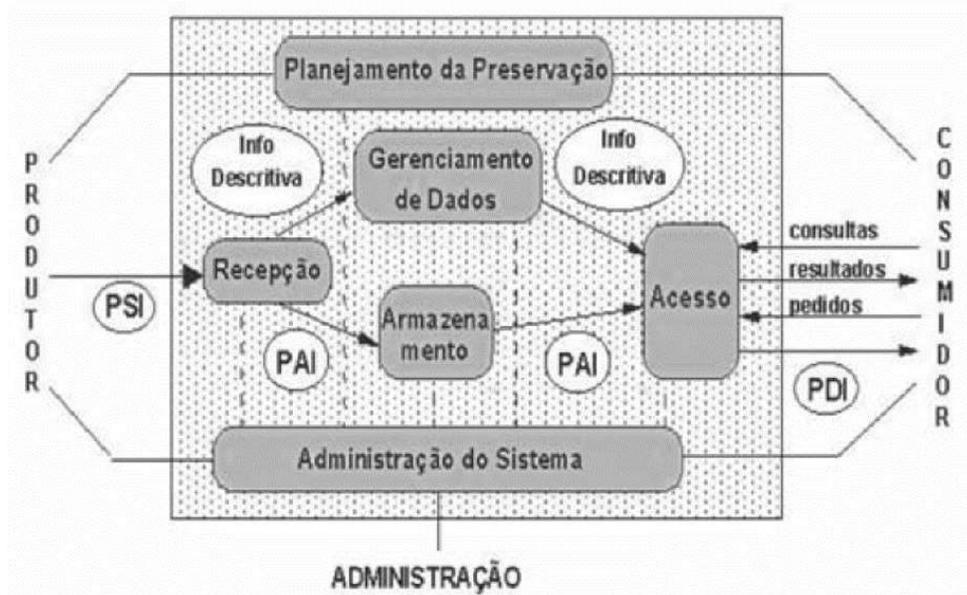


Figura 5: o modelo funcional OAIS

Fonte: Lavoie (2000) citado por Arellano, 2004, p.20.

No OAIS (Figura 5) atuam seis entidades funcionais que gerenciam o fluxo de informação entre si e identificam os componentes funcionais dos arquivos relacionados com a preservação dos arquivos relacionados com a preservação dos objetos digitais. São eles: recepção, armazenamento, gerenciamento de dados, administração de sistema,

planejamento de preservação e acesso.

## 5 | PLANEJAMENTO E ALOCAÇÃO DE RECURSOS

A preservação de longo prazo de objetos digitais abarca não apenas métodos e estratégias relativos à tecnologia, mas também questões administrativas, legais, políticas, econômico-financeiras e de descrição da informação utilizando metadados que permitam o gerenciamento da preservação digital e o acesso futuro. Como requer procedimentos específicos e técnicas apropriadas para cada tipo de formato e mídia, a preservação digital exige investimento elevado e contínuo em infraestrutura de tecnologia, pesquisa e capacitação de recursos humanos, notadamente a necessidade de contratação e capacitação de pessoal. Portanto, no planejamento da preservação digital, as metas e os objetivos pretendidos pelas instituições, a disponibilidade de recursos, as necessidades dos usuários, a preservação do formato de apresentação, das funcionalidades presentes e demais características e propriedades originais dos objetos no qual sustentam a validação de sua integridade e autenticidade, necessitam também ser analisados na elaboração do planejamento para a implantação de um projeto ou política de preservação digital (FORMENTON; GRACIOSO; CASTRO, 2015), tornando imperativo que a formulação de políticas institucionais de preservação seja realizada por especialistas da área de arquivística e Ciência da Informação (CI).

### PROFISSIONAL DA INFORMAÇÃO NA CURADORIA E PRESERVAÇÃO DIGITAL

Vimos na primeira parte deste artigo as variáveis que precisam ser levadas em consideração nas estratégias de gestão da informação, principalmente no que diz respeito ao fenômeno *Big Data*, que requer conhecimentos que vão além dos oferecidos por uma única ciência.

No que diz respeito a Preservação Digital, entre outros fatores relevantes, deve-se dar atenção especial para o item Tecnologia da Informação (TI), pela enormidade de técnicas quando se pensa em encapsulamento de documentos, seus metadados, *software* e especificações do emulador para assegurar a coesão e prevenir sua corrupção. Mais uma vez evidencia-se a necessidade de competências profissionais de indivíduos que vão além das fronteiras da CI já que para fazer a emulação, rodar em futuros computadores e recriar o comportamento do documento digital, tais saberes estão afetos a Ciência da Computação e correlatas.

Para Boeres e Cunha (2016) “a Ciência da Informação, campo multi e interdisciplinar, em função do seu objeto de estudo, que é a informação está cada vez mais ligado às tecnologias da informação...” e na última década, as tecnologias digitais (TD) assumem a cena central deste debate. De um lado estão as estratégias para sua gestão, questões

como quantidade de dados gerados, velocidades de processamento, interpretação e recuperação, e do outro lado está a preservação digital pois temos a cada dia os mais variados suportes e formatos, gerando a preocupação de como esta informação estará disponível e plenamente passível de uso daqui a 100 ou 200 anos, ou mais, como apontado na seção anterior.

Como assegurar que as escolhas de hoje possam dar a certeza do amanhã quanto aos mecanismos de gestão e preservação de dados digitais? Este é um dos desafios para profissionais da Ciência da Informação: preparar-se com as competências que se fazem necessárias ao processo de preservação e curadoria digital.

A partir dos anos 2010, com o rápido crescimento das empresas de tecnologia, de startups, do empreendedorismo, do uso de tecnologias com inteligência artificial e da Internet das Coisas (IOT), o mercado de trabalho passou a ditar uma busca por pessoas eficazes e dinâmicas, capazes de aumentar a perspectiva de vida das empresas e, conseqüentemente, alcançar o sucesso destas (CASE, 2017). O recrutamento nunca teve importância tão significativa nos resultados de uma organização como na Era do Conhecimento (TARAPANOFF, 2001).

Tendo em mente que o conhecimento e a inteligência das pessoas são os agregadores de valor aos produtos e aos serviços, conforme bem nos lembra Sita (2013) Sita (2013) quando afirma que o conhecimento passou a ser 'O' recurso e não 'Um' recurso, o profissional da informação, no campo da Ciência da Informação (CI) poderá melhor dialogar com outras ciências e se posicionar frente os desafios impostos pela Era Digital se adquirir maior consciência sobre si mesmo e sobre a área profissional elegida, durante sua formação.

O mundo mudou e a transformação digital já é um fato. Estamos naturalmente nos transformando, deixando de lado um cenário mais previsível, no qual as organizações eram simples, controláveis e o ambiente estável. O indivíduo se submetia as regras da organização, se inseria num grupo e seguia adiante, com a certeza de que ali se daria seu desenvolvimento intelectual e moral. Contudo, com a globalização, mudanças radicais na economia e na sociedade e inovações disruptivas nas Tecnologias de Informação e Comunicação, as fronteiras de todas as profissões estão sendo derrubadas, estabelecendo-se um novo paradigma nas profissões e no planejamento de carreiras.

## **6 | O FUTURO DO TRABALHO**

Segundo dados apresentados pelo *World Economic Forum / Committee to Improving the State of de World*, nos relatórios de 2016 e de 2018, as habilidades digitais de uma geração de profissionais que atuará em uma economia digital devem ser levadas em conta como um diferencial capaz de manter as engrenagens da economia. Nesses relatórios foram apresentados informações e dados do mundo todo, com o objetivo de informar outras

empresas, governos e trabalhadores na tomada de decisões quanto a gestão do futuro do trabalho, fornecendo uma projeção até 2022 (WORLD ECONOMIC FORUM, 2018).

Do ponto de vista que nos compete neste artigo, ele nos apresenta várias reflexões sobre o quanto em uma sociedade do entorno digital, o profissional da informação que deseja atuar neste ambiente poderá ter uma profissão revitalizada e assegurar que sua formação lhe dê uma maior garantia de empregabilidade e realização pessoal pelos investimentos feitos, já que como bem nos relata Schein (1978), trabalhar pode significar algo mais que ter um emprego. E neste contexto, a segurança relacionada ao trabalho pode estar mais voltada à qualificação profissional e às atitudes pessoais do que às possibilidades atreladas ao empregador.

HOJE, 2018	TENDÊNCIAS, 2022	DECLÍNIO, 2022
Pensamento analítico e inovação	Pensamento analítico e inovação	Destreza manual, resistência e precisão
Solução complexa de problemas	Aprendizagem ativa e estratégias de aprendizagem	Memória, habilidades verbais, auditivas e espaciais
Pensamento e análise críticos	Criatividade, originalidade e iniciativa	Gerenciamento de recursos financeiros e materiais
Aprendizagem ativa e estratégias de Aprendizagem	Projeto e programação de tecnologia	Instalação e manutenção de tecnologia
Criatividade, originalidade e iniciativa	Pensamento e análise críticos	Leitura, escrita, matemática e audição ativa
Atenção aos detalhes, confiabilidade	Solução complexa de problemas	Gestão de pessoal
Inteligência emocional	Liderança e influência social	Controle de qualidade e conscientização de segurança
Raciocínio, resolução de problemas e Ideação	Inteligencia emocional	Coordenação e gestão do tempo
Liderança e influência social	Raciocínio, resolução de problemas e ideação	Habilidades visuais, auditivas e de fala
Coordenação e gestão do tempo	Análise e avaliação de sistemas	Uso, monitoramento e controle de tecnologia

Tabela 2: Síntese da Demanda por Habilidades (2018 versus 2022)

Fonte: autoria própria adaptada de World Economic Forum Centre for the New Economy and Society (2018)

A pesquisa do *World Economic Forum* se concentrou em reunir os pontos de vista de executivos de negócios – principalmente *Chief Human Resources* (CHROs), enfrentando mudanças em andamento na força de trabalho nas empresas da atualidade. As perguntas realizadas se dividiram em três partes: perguntas destinadas a mapear as transformações atualmente em andamento; questões focadas na documentação de tarefas de trabalho em mudança e, portanto, habilidades e requisitos para as funções desempenhadas por indivíduos no local de trabalho até 2022; e questões destinadas a entender as prioridades

e objetivos que as empresas têm que estabelecer em termos de treinamento da força de trabalho e qualificação. A tabela 2 apresenta uma síntese das habilidades que serão mais demandadas nos próximos anos, segundo os autores.

## 71 O PROFISSIONAL DA INFORMAÇÃO

Afinal que competências são esperadas de um profissional da informação que atua na curadoria digital, no âmbito da gestão da informação? Inicialmente é importante resgatarmos o conceito profissional da informação que como bem nos apontou Boeres (2018), já vem sendo estudado desde o ano de 1876. Um dos marcos importantes foi o ano de 1993, quando a Federação Internacional de Informação de Documentação (FID) avivou a discussão ao publicar uma pesquisa identificando os mercados emergentes destes profissionais de informação. O campo das tecnologias da informação, aumentou a discussão em torno de profissões como bibliotecário, arquivista, ampliando o escopo denominado profissional da informação, abarcando a pós-graduação em Ciência da Informação (CI).

É importante uma compreensão das reconfigurações profissionais nas profissões de informação-documentação.

Surge com Crickman (1979) e foi divulgado por Richard Mason (1990) e Ponjuan Dante (1993). A expressão foi posteriormente reforçada pela Federação Internacional de Informação e Documentação (FID) que criou um grupo de estudo dedicado ao moderno profissional de informação (*Special Interest Group / Modern Information Professional* - MIP). O campo informacional não é ocupado completamente por nenhum grupo profissional, sendo uma área permeável a caminhos diversificados e a uma justaposição dos seus espaços, dos seus atores e das suas funções (Cunha, 2009). O Euro-referencial, produzido por um conjunto de associações profissionais europeias, adotou também esta designação (1.<sup>a</sup> edição, 1998), traduzindo a evolução de vários campos profissionais através das suas relações de proximidade e competências. (TELO; PINTO, 2019, p.17).

Segundo Telo e Pinto (2019), o profissional de informação vai sendo definido, a partir da área, dos caminhos que ele segue, dos espaços que ocupa e dos atores com quem ele convive e suas funções. Como vai perpassar por vários campos profissionais, vai sendo desenhado a partir das suas relações de proximidade e competências.

A partir dos relatos trazidos nas seções 3 e 4, podemos inferir que o Profissional da Informação é aquele que faz a mediação entre o criador, o provedor, o usuário e a pessoa da tecnologia da informação, considerando o ciclo documental, desde a geração até o uso da informação pela sociedade, facilitando o acesso a ela.

Para Westerman, Bonnet e McAfee (2014) organização para se tornar digital precisa conseguir construir um ambiente em operações em tempo real e ter 3 princípios fundamentais: padronizar sistemas, processos e informações; automatizar para eliminar interação não validada; e acelerar tomada de decisão através de informações em tempo

real. Ao longo dos três princípios, a interação com o objeto informação é constante e o grande desafio para se iniciar uma jornada digital não está na adoção das tecnologias e, sim nas pessoas. E começar um processo de transformação é mais fácil do que dar continuidade e produzir o engajamento dos colaboradores, para que todos olhem para o mesmo propósito da organização.

## 8 | MAS AFINAL, O QUE É COMPETÊNCIA?

A palavra “competência” está associada à qualidade de quem é capaz de apreciar e resolver determinado assunto ou realizar determinada tarefa. A competência profissional remete à ideia de capacidade, soma de conhecimentos ou habilidades. Boeres e Cunha (2016), mencionam que é importante trabalhar a competência sob vários aspectos à luz de alguns autores, conforme apresentamos na Tabela 3 a seguir:

AUTORES	VISÕES
Gilbert (1978 <i>apud</i> BRANDÃO, 2009)	Competência <b>expressa o desempenho ou comportamento da pessoa no trabalho</b> . Sua proposta sofreu influência do <b>movimento S-R</b> , segundo o qual a compreensão do comportamento deve estar vinculada à investigação das relações entre <b>estímulos (S) e respostas (R)</b> .
Durand (2000)	Identificou como componentes, ou <b>elementos, da competência: conhecimentos, habilidades e atitudes do indivíduo</b> . Sua proposta parece ter sido influenciada pelo <b>movimento S-O-R</b> , segundo o qual essas afinidades precisam considerar o que ocorre no <b>indivíduo (O)</b> , que mediará as relações entre <b>estímulos (S) e respostas (R)</b> .
Brandão e Borges-Andrade (2007), Brandão (2007 : 9, p. 11)	<b>Interação da pessoa com o ambiente resulta em processos cognitivos ou na aquisição de conhecimentos, habilidades e atitudes.</b> Considera <b>três Dimensões da competência</b> : conhecimento, habilidade, atitude.
Para Valentim (2002, p. 122)	competências profissionais são o "(...) <b>conjunto das habilidades, as destrezas, as atitudes e os conhecimentos teórico-práticos necessários para cumprir uma função especializada, de um modo socialmente reconhecido e aceitável</b> ".
Freitas e Brandão (2006 <i>apud</i> BRANDÃO, 2009)	competência é a <b>união dos conceitos de conhecimentos, habilidades e atitudes, divulgados pelo desempenho profissional em um contexto organizacional</b> , agregando valor a pessoas e organizações.
Dewey (1979)	Segundo ele, é a <b>síntese das experiências de vida</b> que cada profissional carrega, <b>produzindo um saber fazer consciente</b> .
Zarifian (1999)	<b>competências profissionais</b> mostram-se quando as <b>pessoas agem nas situações</b> com as quais se deparam <b>nas organizações</b> e <b>unem atributos individuais e a estratégia</b> traçada pela organização

Tabela 3: Competências - seus vários conceitos

Fonte: autoria própria, adaptada de Boeres e Cunha (2016)

## 9 | COMPETÊNCIAS NA CURADORIA E PRESERVAÇÃO DIGITAIS

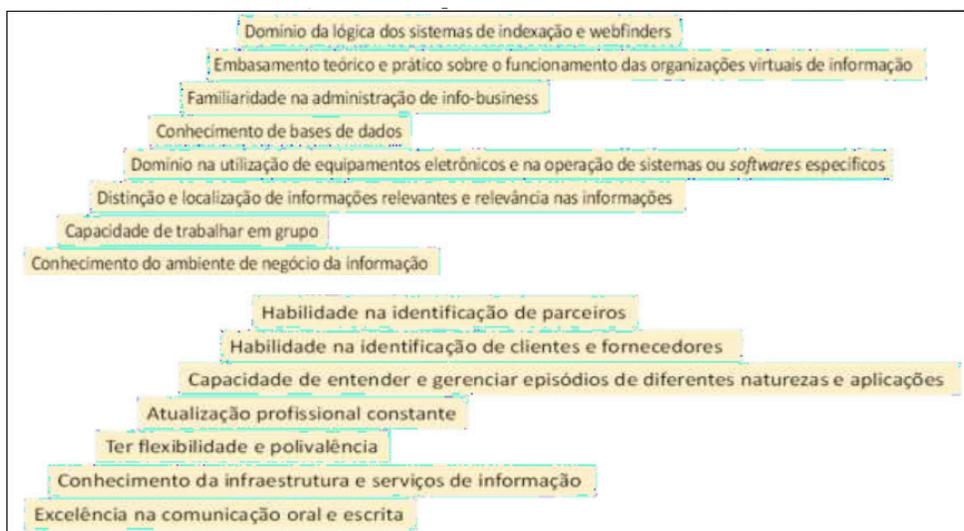
Para Davenport e Prusack (1998) citado por Faria e Castro Filho (2014) bons profissionais de informação digital necessitam de habilidades *hard* (conhecimento estruturado, qualificações técnicas e experiência profissional); e atributos *soft* (senso dos aspectos culturais, políticos e pessoais do conhecimento). Então, uma das reflexões que se

pode ter neste momento é que o profissional da informação digital é um profissional híbrido, que deverá ter em conta competências técnicas e pessoais. Para Westerman, Bonnet e McAfee (2014), a transformação digital das organizações passa a necessitar de profissional com habilidades digitais e mentalidade para trabalhar na economia digital.

Ambos os autores estão falando de profissionais com habilidades híbridas – pessoas técnicas que precisam ter mais experiência em negócios e ao mesmo tempo ser mais experientes em tecnologia. Dentro deste contexto, será que estamos formando bibliotecários como o profissional da informação digital, que deverão em sua formação unir lacunas de comunicação entre competências digitais e empresariais, mesmo quando sua atuação for na cultura?

Segundo Vendrell e Miranda (1999) as competências necessárias para o profissional de informação são de seis tipos: competência intelectual, prática, interativa, social, ética e estética.

Para Faria (2015) as competências mais importantes demandadas pelo mercado de trabalho, no âmbito da Ciência da Informação são:



Quadro 1: Competências, no âmbito da CI

Fonte: autoria própria, adaptada de Faria (2015, p. 49)

Percebemos que Faria (2015) e Westerman, Bonnet e McAfee (2014), embora sejam de ciências distintas, se amparam em um profissional formado por competências que vão desde a compreensão do modelo de negócio até os saberes tradicionais da Biblioteconomia.

Um achado interessante que data de um período antes da explosão da transformação

digital, mas que dialoga com os demais autores aqui já citados, foi o que nos apresenta Valentin (2002): no contexto do trabalho de CI, existem dois tipos de competências, as técnico- científicas e as gerenciais. As técnico-científicas pressupõem:

Executar o processamento de documentos em distintos suportes, em unidades, sistemas e serviços de informação;

Armazenar e recuperar informação guardada em qualquer meio para os usuários de unidades, serviços e sistemas de informação;

Utilizar e disseminar fontes, produtos e serviços de informação em diferentes suportes;

Executar procedimentos automatizados;

Planejar e construir redes de informação.

Quadro 2: Competências, segundo Valentin

Fonte: autoria própria, adaptada de Valentin (2002)

Se o mercado carece de um profissional portador de competências, sendo o bibliotecário possuidor de muitas delas, talvez o desafio esteja em o bibliotecário tomar consciência de suas aptidões e passar a dar vazão para uma atuação maior como gestor do conhecimento, segundo nos apresenta Souto (2005).

Segundo Valentin (2004), as competências e habilidades do Bibliotecário vem sendo aceleradamente ditadas de fora para dentro. Este profissional passa a ter múltiplos espaços de atuação, inclusive já sendo reconhecido a partir dos anos 2010 como Bibliotecário de Dados, pela possibilidade de desenvolver ao longo de sua formação habilidades necessárias para trabalhar com qualquer tipo de dados, sejam eles observacionais, computacionais ou experimentais.

Para Paletta e Milanesi (2016), à medida que o mundo se torna mais sofisticado e diversificado, a competência exigida do bibliotecário profissional, que em uma primeira etapa é altamente técnica, cresce e inclui outras responsabilidades importantes no novo ambiente organizacional.

Além do domínio de múltiplos saberes, para se construir um perfil de bibliotecário protagonista, Varela e Barbosa (2012), explicam que é preciso empreender ações durante a formação promovendo o desenvolvimento cognitivo e o domínio de conteúdos e metodologias profissionalizantes, especializados do campo da Biblioteconomia e da Ciência da Informação e de áreas do conhecimento correlatas, de modo a solidificar essa formação, expandindo a visão crítica sobre a pluridisciplinaridade e complexidade, conduzindo o bibliotecário a focar no princípio de formação integral e integradora do indivíduo e dele com o ambiente, tendo como pilar de sustentação as demandas sociais e a função social da profissão.

Ao compreender a necessidade de integrar múltiplos saberes, desenvolvendo um conjunto de conhecimentos e habilidades, o bibliotecário que atuar no campo da preservação digital pode se sentir cada vez mais preparado para exercer seu papel e se legitimar frente a Sociedade que poderá passar a respeitá-lo como tal.

## REFLEXÕES

Sem dúvida os avanços tecnológicos das últimas décadas aceleram a produção e transformação de informação e documentos e mudam rapidamente a fronteira entre as tarefas de trabalho realizadas por humanos. Para aproveitar o potencial transformador da Quarta Revolução Industrial, líderes empresariais de todos os setores e regiões cada vez mais são chamados a formular uma abrangente estratégia de força de trabalho pronta para enfrentar os desafios desta nova era de mudança acelerada e inovação.

Com tantas transformações em ritmo acelerado, a janela de oportunidade para ações proativas e para o gerenciamento simultâneo dessa mudança está se fechando rapidamente e as empresas, instituições, governos e trabalhadores devem ter a iniciativa de planejar e implementar uma nova visão para o mercado de trabalho global.

A preservação da informação e dos documentos digitais, como visto anteriormente, traz vários desafios de longo prazo, ligados às questões de *hardware*, *software*, formatos, metadados, mídias, políticas e procedimentos, padrões, formatos entre outros.

Assim, o conhecimento e as aptidões necessárias para os profissionais da informação envolvidos na gestão da informação, preservação e curadoria digital incluem conhecimentos integrados em ciência da computação, ciência da informação, biblioteconomia, arquivística e outras disciplinas e domínios relacionados com a criação, utilização, armazenamento, disponibilização e preservação de dados.

Analisando dados trazidos pelos relatórios 2016 e 2018 do *Word Economic Forum*, nos perguntamos quais são as implicações no desenvolvimento da trajetória do profissional da informação digital, enquanto campo da Ciência da Informação (CI)? Este profissional está sendo preparado para os desafios que a preservação digital enfrenta, a partir da escolha pela formação em Biblioteconomia? O estudante de biblioteconomia tem entendimento sobre suas competências profissionais? Os profissionais que lidam com o serviço de recrutamento e seleção compreendem as competências técnicas e funcionais que o bibliotecário desenvolve, ao longo de sua formação e como podem ser utilizadas pelas organizações?

Por exemplo, no que se refere à empregabilidade dos profissionais da curadoria digital, foram identificadas as seguintes competências funcionais exigidas por parte de entidades empregadoras: criação e edição de diversos tipos de objetos digitais, digitalização, criação de metadados, gestão de repositórios e implementação de estratégias e preservação. Além disso, também foram exigidos conhecimentos ao nível de: formatos de arquivos, workflow de curadoria, princípios de gestão de dados, arquitetura de repositórios,

tecnologia de pesquisa e recuperação, estratégias de preservação, aplicação dos direitos de autor, políticas de investigação e procedimentos relacionados com gestão de dados.

O *Committee on Future Career Opportunities and Educational Requirements for Digital Curation* procurou identificar e analisar as atividades efetuadas no âmbito da curadoria digital com o propósito de determinar, com rigor, os conhecimentos e competências exigidas no mercado de trabalho e concluiu que, apesar da crescente oferta de formação dessa área, esta realiza-se quase que exclusivamente ao nível de estudos avançados (pós-graduações e especializações).

Falta também a definição sobre quem estará adiante das unidades de informação digitais, gerindo-as, se bibliotecários, se o pessoal de TI, se arquivistas, pois são muitos são habilitados. Quem sabe o futuro virá, como sugere o *National Science and Technology Council* (2009), com a colaboração entre as organizações, entidades e pessoas para levar adiante a gestão e responsabilidade pela preservação e curadoria de dados digitais.

## CONSIDERAÇÕES FINAIS

Com a transformação digital, gerenciar a informação torna-se uma atividade cada vez mais complexa. As organizações precisam estar atentas para poderem readequar suas estruturas, processos e competências profissionais, pois o aumento do valor da informação juntamente com a crescente necessidade de interoperabilidade de sistemas, requer um olhar da GI para além dos limites do uso da informação, onde a preservação da informação digital torna-se uma questão estratégica.

Garantir que a informação que é produzida hoje será acessível daqui a 50, 100 ou 200 anos é um dos maiores desafios dos tempos atuais. Com o tempo, se for adequadamente preservada, será deslocada para os arquivos e museus transferindo para estes a necessidade de lidar com a preservação digital. Acostumados a essencialmente lidar com objetos e documentos físicos, estes equipamentos têm pela frente um grande desafio que deve começar a ser resolvido agora, para garantir o futuro dos itens que ainda receberá. Este é um tema pouco abordado na literatura, e que deve ser objeto de estudo futuro.

Também se identifica a necessidade de identificar os fatores que determinam a ocupação das vagas que poderiam ser preenchidas por bibliotecários e as são por profissionais ditos “da informação”, mais especificamente por possuírem habilidades digitais ou domínio de modelos de negócios em uma economia digital. Assim, pretende-se expandir os resultados coletados nesta pesquisa bibliográfica sobre o profissional da informação para uma pesquisa futura, pois acredita-se que preparar os novos profissionais da informação, com formação adequada para os enfrentamentos de uma economia digital é um intento a ser considerado pelas universidades que pretendem se engajar na transformação digital.

## REFERÊNCIAS

ARELLANO, Miguel Ángel Márdero. Preservação de documentos digitais. **Ciência da Informação**, v. 33, n. 2, p. 15–27, 2004. Disponível em: [https://www.scielo.br/scielo.php?pid=S0100-19652004000200002&script=sci\\_arttext&lng=pt](https://www.scielo.br/scielo.php?pid=S0100-19652004000200002&script=sci_arttext&lng=pt).

ARELLANO, Miguel Ángel Márdero. Cariniana: uma rede nacional de preservação digital.

**Ciência da Informação**, v. 41, n. 1, p. 83–91, 2012. Disponível em: <http://revista.ibict.br/ciinf/article/view/1354>.

BOERES, Sonia. Literacy and the organization of digital information coupled with lifelong learning. **Revista Digital de Biblioteconomia e Ciência da Informação**, v. 16, n. 2, p. 483–500, 2018. DOI: <https://doi.org/10.20396/rdbci.v16i2.8651507>.

BOERES, Sonia; CUNHA, M. B. Competência para a preservação e curadoria digitais. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação** 14, v. 14, n. 3, p. 426–449, 2016. DOI: <https://doi.org/10.20396/rdbci.v14i3.8646303>.

CASE, Steve. **Terceira onda da internet: a reinvenção dos negócios na era digital**. São Paulo: HSM, 2017.

COLUMBIÉ, Radames Limares. Presentación. *In: Gestión de información. Dimensiones e implementación para el éxito organizacional*. Gijón: Ediciones Trea SL, 2007.

CONSELHO NACIONAL DE ARQUIVOS, CONARQ. **Carta para a Preservação do Patrimônio Arquivístico Digital**. Brasília, 2004

CROSAS, Mercè; KING, Gary; HONAKER, James; SWEENEY, Latanya. Automating Open Science for Big Data. **Annals of the American Academy of Political and Social Science**, v. 659, n. 1, p. 260–273, 2015. DOI: <https://doi.org/10.1177/0002716215570847>. Disponível em: <http://nrs.harvard.edu/urn-3:HUL.InstRepos:25290370>.

DAVENPORT, Thomas H. **Big data at work: dispelling the myths, uncovering the opportunities**. Boston, Massachusetts: Harvard Business School Publishing Corporation, 2014.

DETLOR, Brian. Information management. **International Journal of Information Management**, v. 30, n. 2, p. 103–108, 2010. DOI: <https://doi.org/10.1016/j.ijinfomgt.2009.12.001>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0268401209001510?via=ihub>.

FARIA, A. C. ; CASTRO FILHO, C. Profissional da informação: estudos dos egressos no estado de São Paulo, mundo do trabalho, habilidades e competências. **Ponto de Acesso**, v. 8, n. 3, p. 44–63, 2014. Disponível em: <https://periodicos.ufba.br/index.php/revistaici/article/view/8465>.

FARIA, Ana Carolina Cintra. **A inserção do bibliotecário no mercado de trabalho: fatores de influência e competências**. 2015. Universidade de Brasília, Brasília, 2015.

FLORIDI, Luciano. **Information: A very Short Introduction**. New York: Oxford University Press INC., 2010.

FORMENTON, D.; GRACIOSO, L. S.; CASTRO, F. F. Revisitando a preservação na perspectiva da ciência da informação: aproximações conceituais. **Revista Digital de Biblioteconomia e Ciência da Informação**, v. 13, n. 1, p. 170–191, 2015. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/1587>.

KAISLER, Stephen; ARMOUR, Frank; MONEY, William; ESPINOSA, J. Alberto. Big Data Issues and Challenges. **Encyclopedia of Information Science and Technology, Third Edition**, n. 2013, p. 363–370, 2014. DOI: 10.4018/978-1-4666-5888-2.ch035. Disponível em: <https://ieeexplore-ieee-org.ez67.periodicos.capes.gov.br/document/6479953/>.

MIDDLETON, Michael Robert. **Information Management: a consolidation of operations, analysis and strategy**. Wagga Wagga: Centre for Information Studies Charles Sturt University CIS, 2002.

MONTEIRO, Samuel Alves; DUARTE, Emeide Nóbrega. Bases teóricas da gestão da informação. **InCID: Revista de Ciência da Informação e Documentação**, v. 9, n. 2, p. 89–106, 2018. DOI: 10.11606/issn.2178-2075.v9i2p89-106. Disponível em: <http://www.revistas.usp.br/incid/article/view/133677>.

NATIONAL SCIENCE AND TECHNOLOGY COUNCIL. **Interagency Working Group on Digital Data. Harnessing the power of digital data for science and society**. 2009 Disponível em: [http://www.nitrd.gov/about/harnessing\\_power\\_web.pdf](http://www.nitrd.gov/about/harnessing_power_web.pdf).

PALETTA, Francisco Carlos; MILANESI, L. Current findings from research on professional librarian skills development and needs in brazil. **Informação & Sociedade:Estudos**, v. 26, n. 1, p. 199–209, 2016. Disponível em: <http://www3.eca.usp.br/sites/default/files/form/biblioteca/acervo/producao-academica/002789199.pdf>.

PINTO, Maria Manuela. Um modelo sistémico de gestão da informação: da flexibilidade organizacional à interoperabilidade do sistema. **Revista Ibero-Americana de Ciência da Informação**, v. 12, n. 1, p. 22–52, 2018. Disponível em: <http://periodicos.unb.br/index.php/RICI/article/view/11015/18565>.

PONJUÁN DANTE, Gloria. **Gestión de información en las organizaciones**. Santiago: CECAPI, 1998.

PONJUÁN DANTE, Gloria. **Gestión de información - Dimensiones e implementación para el éxito organizacional**. 2ª ed. Gijón: Ediciones Trea SL, 2007.

PONJUÁN DANTE, Gloria. La gestión de información y sus modelos representativos. Valoraciones. **Ciencias de la Información**, v. 42, n. 2, p. 11–17, 2011. Disponível em: [https://www.researchgate.net/profile/Gloria\\_Ponjuan/publications](https://www.researchgate.net/profile/Gloria_Ponjuan/publications).

SCHEIN, Edgar H. **Career dynamics**. Reading: Addison-Wesley, 1978.

SFETCU, N. **Big Data Ethics in Research**. MultiMedia Publishing, 2019. DOI: 10.13140/RG.2.2.11054.46401.

SITA, Mauricio (Coord. .. **Capital intelectual: a fórmula do sucesso**. São Paulo: Ser Mais, 2013.

SOUTO, Leonardo Fernandes. **O profissional da informação em tempos de mudanças**. Campinas: Alínea, 2005.

TARAPANOFF, Kira. **Inteligência organizacional e competitiva**. Brasília: UnB, 2001. TELO, P. A. O. de C.; PINTO, M. L. B. G. Debater a visão europeia sobre competências de informação-documentação: perspectivas, estratégias e posicionamento. **Ciência da Informação**, v. 48, n. 2, p. 25–49, 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/4695>.

UNITED NATIONS EDUCATIONAL SCIENTIFIC AND CULTURAL

ORGANIZATION, UNESCO. **Charter on the preservation of the digital heritage**. Paris. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000229034?posInSet=3&queryId=87a30236-5b54-4567-90ad-1060384b7752>.

VALENTIN, Marta Lígia. **Formação do profissional da informação**. São Paulo: Polis, 2002.

VALENTIN, Marta Lígia. **Atuação profissional na área de informação**. São Paulo: Polis, 2004.

VARELA, A.; BARBOSA, M. L. A. Trajetórias cognitivas subjacentes ao processo de busca e uso da informação: fundamentos e transversalidades. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 17, n. 1, p. 142–168, 2012. DOI: 10.5007/1518-2924.2012v17nesp1p142. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17nesp1p142/22730>.

VENDRELL, B. B. De; MIRANDA, M. J. Competencias del profesional de la información: un andén en el Mercosur. *In*: ENCUENTRO DE DIRECTORES Y DOCENTES DE ESCUELAS DE BIBLIOTECOLOGÍA DEL MERCOSUR 1999, Santiago. **Anais [...]**. Santiago: UTEM, 1999. p. 99–104.

WESTERMAN, G.; BONNET, D.; MCAFEE, A. **Leading digital: turning technology into business transformation**. Boston: Harvard Business School Publishing Corporation, 2014.

WORLD ECONOMIC FORUM, . **The future of Jobs report 2018**.

# PRESERVAÇÃO DIGITAL: CONSIDERAÇÕES SOBRE ARQUIVAMENTO DE CONTEÚDO DA WEB

**Antonio Paulo Carretta,**

Doutorando em Ciência da Informação pelo PPGCI ECA-USP

**RESUMO:** Artigo aborda questões sobre preservação e justificativas para arquivamento de conteúdo Web. Apresenta breve panorama de evolução da Web; formato de arquivamento (WARC), padrões e recomendações de preservação digital. Por fim, observa dimensões de preservação, seleção e políticas de Web arquivamento.

**PALAVRAS-CHAVE:** Preservação Digital; Arquivamento da Web; Preservação da Web.

## DIGITAL PRESERVATION: CONSIDERATIONS ON WEB ARCHIVING

**ABSTRACT:** Article presents digital preservation issues and justifications for Web archiving. It contains a brief overview of the evolution of the Web; archiving format (WARC), digital preservation standards and recommendations. Finally, it observes dimensions of preservation, selection, and Web archiving policies.

**KEYWORDS:** Digital Preservation; Web Archiving; Web Preservation.

## INTRODUÇÃO

Presente em nosso cotidiano digital, a Web é uma importante camada tecnológica com registros que constroem a memória de nossa sociedade. Para entender melhor esse quadro, vale a pena fazer a seguinte associação: assim como publicações impressas e outros documentos físicos compõem o patrimônio da humanidade, permitindo resgatar a memória de diversas épocas por meio da representação do pensamento e da cultura em sociedade, de maneira similar o conteúdo digital disponível na Web registra nosso passado mais recente e se constitui fonte de referência para estudos futuros, bem como herança cultural de uma nação.

Neste sentido, em 2003, a UNESCO reconheceu que estamos diante de um novo legado cultural: o patrimônio digital, que consiste em recursos únicos, derivados do conhecimento ou expressão de seres humanos, gerados exclusivamente em formato digital, e que incluem informações de diversas naturezas: cultural, educacional, científica e outras.

Porém, a condição de existência e manutenção do conteúdo Web depende de ações de preservação que minimizem as perdas contínuas de informação digital, decorrentes de erros de endereçamento de páginas Web, mudanças de plataformas, descontinuidade

de projetos ou desaparecimento de blogs e sites. Enfrentar com estratégias e políticas de preservação essa fragilidade do ambiente é o grande desafio dos países que pretendem manter registros de memória para futuras gerações.

Para refletir sobre esse tema, este artigo traça um breve panorama de evolução da Web; considera aspectos atrelados às tecnologias e estruturas da Internet, o formato adequado para arquivamento, recomendação de preservação e, por fim, levanta pontos de discussão sobre políticas de arquivamento, curadoria e reuso de conteúdo da Web.

## **CONTEXTO HISTÓRICO E PANORAMA DA WEB**

Passados mais de 30 anos da criação da Web, proposta por Tim Berners-Lee, observamos um ambiente de comunicação e disseminação de informação digital que se tornou parte do cotidiano das pessoas. A evolução ocorreu com velocidade, volume e variedade de conteúdo. De apenas uma página experimental, em 1990, atualmente existem mais de 1.5 bilhões de sites na World Wide Web. Desse total, segundo dados do órgão de controle de domínios nacionais, o Registro.br, a Web brasileira já representa mais de 4 milhões de domínios com terminação “.br”.

Entretanto, do total de sites no mundo, menos de 200 milhões estão ativos<sup>1</sup> e a explicação está no fato que, ao longo do tempo, os conteúdos publicados na Web podem perder a dinâmica de atualização ou atividade. Mike Ashenfelder, membro da Biblioteca do Congresso dos EUA, destaca que a vida útil de uma página na Web é de cerca de 100 dias, pouco mais de 3 meses (ROCHA, 2015). Essa natureza efêmera indica que depois dos 3 meses, muitos endereços deixam de ser atualizados ou perdem a frequência de postagens e acabam caindo no esquecimento.

Segundo estimativa do Internet Live Stats, que reforça ainda mais a fragilidade de manutenção do ambiente Web, 75% dos sites já criados estão vivos, mas não são ativos. Outros indicadores apontam que 80% das páginas web não estão disponíveis na sua forma original após 1 ano; 13% de artigos acadêmicos online desaparecem após 27 meses e 11% dos recursos de mídias sociais são perdidos depois de 1 ano (COSTA, GOMES, SILVA, 2016).

A percepção dessa condição é mais evidente ao buscarmos por páginas da Web e nos depararmos com o “Erro 404”, mensagem automática enviada pelo servidor de hospedagem do site avisando que determinada página não foi encontrada. Na opinião de especialistas, respectivamente em estrutura da internet e curadoria digital, estamos diante de um grande problema que exige analisarmos o impacto da ausência de ações de preservação da Web. Segundo Vint Cerf, vivemos o momento da chamada “era das trevas digital”, marcada pela incerteza de recuperação de conteúdos produzidos e que levanta uma questão estratégica e com projeção para o futuro: o que poderia acontecer

<sup>1</sup> Dados estatísticos indicados nos sites Internet Live Stats (<https://www.internetlivestats.com>) e Registro.br (<https://registro.br/dominio/estatisticas/>), consultados em julho de 2020.

com o conteúdo de publicações acadêmicas e diversas referências online, quando daqui a dez, 20 ou 50 anos os domínios mudarem, forem abandonados ou simplesmente deixarem de ser pagos? Nessa linha de raciocínio, Aquiles Brayner alerta que este século vai se configurando como a ‘Era do Erro 404’ e “o futuro tende a ser marcado pela escassez de fontes históricas. Não se pode pensar a história do século 21 sem levar em conta o que se é publicado na internet, incluindo as redes sociais” (ABDALLA, 2018).

Em oposição a esse contexto alarmante, ao longo do tempo de existência da Web, ações e serviços foram surgindo para preservação de parte da memória digital de nossa sociedade. Por exemplo, iniciado em 1996 e atuando como arquivo para armazenamento de páginas da Web, o Internet Archive é referência nos estudos sobre preservação digital e colaborou no desenvolvimento do atual formato WARC (Web ARChive), utilizado pelos participantes do International Internet Preservation Consortium (IIPC) como padrão para armazenamento de documentos gerados na Web (INTERNET ARCHIVE, 2020; INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2009).

Por sua vez, o IIPC surgiu em 2003 para agregar esforços no desenvolvimento de recomendações e melhores práticas para preservação digital. Formado por instituições públicas, privadas e sem fins lucrativos, em sua maioria bibliotecas e arquivos nacionais, localizados em alguns países da África, Ásia, América do Norte, América do Sul (apenas o Chile), Europa e Oceania. Atualmente conta 37 países e 57 membros, dos quais 16 países declaram em sua legislação de depósito legal propósitos para preservação do conteúdo Web (IIPC, c2020). Aliás, o diferente contexto jurídico de cada país participante esbarra em questões legais e, por isso, apresentam diferentes diretrizes de preservação e autorizações de acessos aos documentos Web armazenados.

Entre os membros participantes do IIPC, podemos mencionar o caso do Arquivo.pt, serviço de arquivamento e preservação da Web portuguesa, que iniciou oficialmente em 2008. Este serviço contém páginas no idioma português desde 1996, mantém estrutura amigável de pesquisa, não apenas por url, mas também por termos, e oferece espaço para reflexão do tema, bem como estimula pesquisas inovadoras com base no reuso do conteúdo histórico arquivado (ARQUIVO.PT, 2020a).

Paralelamente, outras iniciativas e projetos especiais tem se preocupado com a preservação do conteúdo originado no ambiente Web. Com dimensões e propósitos diferentes, diversas instituições atuam em ações de preservação de seu próprio conteúdo, registros de redes sociais e informações temáticas, de contexto político ou cultural, como é o caso do consórcio The New York Art Resources Consortium (NYARC, 2020), que integra três importantes museus de arte na cidade de Nova York: Museu do Brooklyn, Coleção Frick e Museu de Arte Moderna. O projeto da NYARC arquivava coleções selecionadas de sites de exposições, em áreas que correspondem ao escopo e pontos fortes dessas instituições, bem como sites de projetos do consórcio e sites institucionais dos três museus.

Embora o Brasil tenha representatividade na Web e intuições públicas preocupadas

com o tema, como Arquivo e Biblioteca Nacional, alinhadas com programas de preservação digital, o país não possui uma legislação de apoio, nem serviços de abrangência nacional dedicados à preservação de informações do domínio.BR. Pontualmente, existem casos e projetos especiais relacionados com a Web brasileira (ROCKEMBACH, 2018), por exemplo, o Latin American Web Archiving Project, hospedado em endereço eletrônico da Universidade do Texas em Austin<sup>14</sup>, cuja coleta para arquivamento iniciou em 2005 e tem foco em documentos governamentais e de expressão política.

Outra iniciativa pontual foi do Grupo de Trabalho de Desenvolvimento de Conteúdo vinculado ao Consórcio Internacional de Preservação da Internet, que coletou sites, artigos, notícias, blogs e mídias sociais (Twitter, Facebook) sobre as Olimpíadas Rio 2016. O grupo realizou formulário público e incentivou pessoas para colaborarem com a seleção de temas relacionados às Olimpíadas 2016. A hashtag '#RIO2016WA' no Twitter também foi uma forma de marcar e acompanhar informações relativas ao arquivamento das Olimpíadas 2016. Resumidamente, este recorte histórico e o panorama apresentado trazem um contexto preocupante, principalmente para países que retardam o enfretamento desse processo de perda de memória digital. Embora o Brasil seja exemplo de discussões sobre governança da internet, educação e direitos nas redes digitais, faltam ainda estratégias e ações, concretas e colaborativas, que impulsionem um programa nacional para arquivamento de conteúdo da Web brasileira. Além disso, carece do fortalecimento de base teórico-empírica, apoiada por pesquisas nacionais e amplas reflexões nos campos da Ciência da Informação e Humanidades Digitais.

## **ARGUMENTOS E RECOMENDAÇÕES SOBRE ARQUIVAMENTO E PRESERVAÇÃO DA WEB**

O processo de arquivamento e preservação da Web é amplo e complexo. Para refletir sobre sua função e a natureza desse recurso, podemos destacar alguns conceitos ou argumentos de apoio: memória digital, o formato digital e as recomendações de preservação de conteúdos em ambientes digitais.

Segundo Vera Dodebei (2006), na Web a característica de acumulação do conhecimento ocorre no domínio coletivo, provocando uma permanente construção e reconstrução da informação, gerando não apenas bancos de dados ou bases de dados, mas centros de conhecimento compartilhado. Neste contexto, a Web passa ser vista como espaço da memória social, permitindo pesquisas sobre comportamento e meios de produção do conhecimento, seja ele de natureza histórica, artística ou técnica. Este domínio coletivo “faz da Web um grande centro virtual da memória do mundo.”

Analisando essa construção de uma memória digital, Dodebei (2006) elabora uma questão essencial: “Como propiciar a preservação das memórias que circulam na web?” Nesta indagação a autora deixa claro que “a existência do patrimônio digital e sua

permanência na memória virtual do mundo estarão intimamente ligadas às condições de preservação que forem proporcionadas por aqueles que se responsabilizarem pela inclusão e manutenção de um objeto na rede mundial.” Muito embora exista consciência da quase impossibilidade de preservação total da informação em meio digital, grandes são os esforços para manter um processo de transmissão para outras gerações, impedindo seu desaparecimento, absorção ou dispersão entre domínios hegemônicos, de ordem idiomática ou geopolítica. Para evitar isso, é necessário não apenas favorecer a migração de informação para novos suportes, mas dotar os objetos digitais de significado, preservando seu conteúdo intelectual e garantindo sua autenticidade e acessibilidade.

Preservar a integridade de um documento online é outro ponto essencial de análise. A Web opera com camadas. Percebemos que a busca por uma melhor solução para lidar com o problema dos dados estruturados, considerando sua migração para um formato de arquivamento adequado, está baseada na moderna construção de websites que, basicamente, funcionam com uma arquitetura em três camadas: camada de metadados oferecem estrutura e capacidade de recuperação do conteúdo; camada de apresentação dos dados, definida pela codificação em HTML do documento acessado; e camada lógica, definida pelo software que permite comunicação, leitura e exibição dos elementos das camadas de dados e apresentação (MASANÉS, 2006).

Essa estrutura em camadas é complexa e opera de forma dinâmica. Uma página da Web pode conter vários elementos (texto, vídeo, áudio, imagens), que refletem um esforço editorial e estético, por padrão ou livre desenvolvimento, e pode conter diferentes combinações de visualização e interatividade. Todas estas características precisam ser mantidas ao migrarmos esse conteúdo para um formato de preservação. No caso da Web, para fins de armazenamento, o formato atualmente utilizado é WARC (Web ARChive), cujo requisito geral permite carregar de forma segura todos os objetos digitais constituintes da estrutura do site, facilitando o processo de coleta, acesso e organização do conteúdo armazenado.

O formato WARC oferece um padrão estruturador para gestão e armazenamento de milhões de recursos digitais coletados da web, por meio da construção de aplicativos para coleta, gerenciamento, acesso, mineração e troca de conteúdo. Embora seja considerado um formato exclusivo para arquivos da web, pode ainda ser adotado para armazenar objetos natos digitais ou digitalizados. Adicionalmente, a maneira como os arquivos WARC são criados e os recursos armazenados depende do uso de software e aplicativos específicos.

Segundo a norma ISO 28500:2009 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2009) o formato de arquivo WARC tem como propósito:

- armazenar o conteúdo da carga útil (corpo de dados) e as informações de controle dos principais protocolos da camada de aplicativos da Internet, como o HTTP (Hypertext Transfer Protocol), o DNS (Sistema de Nomes de Domínio) e o FTP (File Transfer Protocol);

- armazenar metadados complementares vinculados a outros dados armazenados (por exemplo, classificador de assunto, idioma e codificação);
- apoiar a compactação de dados e manter a integridade do registro de dados;
- armazenar todas as informações de controle do protocolo de coleta (por exemplo, cabeçalhos de solicitação), não apenas informações de resposta;
- armazenar os resultados de transformações de dados vinculados a outros dados armazenados;
- armazenar indicações de duplicidade vinculadas a outros dados armazenados (para reduzir o armazenamento na presença de recursos idênticos ou substancialmente semelhantes);
- ser estendido sem interromper funcionalidades existentes;
- suportar o manuseio de registros excessivamente longos por truncamento ou segmentação, quando desejado.

Por conta dessas características e a natureza que adquire esse formato, outro conceito pode ser empregado para definir os arquivos que surgem da transição estrutural: o renascido digital. Segundo Brügger (2017), materiais digitais podem ter diferentes processos de formação: digitalizado, nato digital e renascido digital.

O material digital renascido é um material digitalizado ou nato digital que foi coletado e preservado, mas em algum grau foi alterado neste processo, por exemplo, como acontece após a transição de um formato para um arquivo da Web. Considerando essa tríade digital, a Web pode desempenhar três papéis diferentes: 1º como plataforma para distribuição de materiais analógicos digitalizados; 2º como fonte que já nasce de processo e formato digital, e 3º como fonte digital renascida em um arquivo da Web, estruturado com propósito de armazenamento e preservação.

Nas últimas décadas, propostas para organização e armazenamento da memória digital tentam encontrar maneiras adequadas para coletar, em grande escala, a imensa quantidade de conteúdo gerado na Web. Para auxiliar nessa tarefa, são utilizadas ferramentas como “crawlers”, rastreadores que navegam por websites de maneira automatizada e conforme um conjunto de políticas de coleta. O fluxo de operação desses rastreadores segue a seguinte lógica: a partir de uma lista de URLs, este programa salva cada página contida em determinada URL, encontra os hiperlinks da página (links para outros endereços, imagens, vídeos, scripts ou alguma instrução do código fonte) e os adiciona em uma lista controlada de URLs que visita recursivamente para novas ações de coleta.

Esse processo para armazenar e gerenciar tal quantidade de páginas da Web representa um grande desafio. Por conta disso, diversas instituições de memória definem algumas recomendações e práticas que podem auxiliar a operação dos rastreadores e, de certa forma, tornar um website mais ou menos preservável. Entre as recomendações

sugeridas pelo Arquivo.pt (2020b) podemos apontar:

1. Utilize diretrizes de acessibilidade, além disso, evite lacunas na identificação do conteúdo, por exemplo: observe se incluiu corretamente a data de publicação dos documentos;
2. Defina códigos e use provedores que ajudem os rastreadores de arquivamento, para isso, use corretamente o protocolo de exclusão de robots > robots.txt é um arquivo, salvo na pasta raiz do site, que serve para dar ordens específicas aos robôs de busca e rastreadores indicando páginas do site que podem (ou não) ser acessadas para arquivamento e apresentam conteúdo textual relevante;
3. Considere usar uma licença Creative Commons, indicações que expressam claramente quando determinado conteúdo é livre para uso ou possui restrições de coleta;
4. Mantenha URLs estáveis e redirecione quando necessário. Qualquer mudança na URL pode tornar um site inacessível e fora do alcance dos rastreadores; use um endereço para cada conteúdo;
5. Utilize formatos adequados e recomendados para preservação, por exemplo: HTML ou XML (texto), PNG (imagem), AVI (vídeo);
6. Publique metadados para enriquecer os conteúdos, por exemplo, metadados descritivos do padrão Dublin Core, estas informações enriquecem ou complementam conteúdos capturados. Este conjunto de aspectos nos ajuda entender a complexidade de preservação da Web e quais protocolos apoiam o processo de arquivamento. No entanto, para pensarmos em ações mais amplas, a definição de políticas de coleta, estratégias de curadoria e acesso ao conteúdo arquivado são fundamentais para proteção do patrimônio digital em evolução.

## **POLÍTICAS DE ARQUIVAMENTO WEB, CURADORIA E REUSO DE INFORMAÇÃO**

Abrindo a discussão sobre políticas e estratégias de preservação da herança digital, Masanès (2006) considera 3 pontos críticos muito comuns em argumentações iniciais contra o arquivamento de conteúdo da Web: a qualidade do conteúdo publicado; a condição tecnológica de preservação do ambiente digital e a imensurável tarefa de armazenar e permitir acesso ao conteúdo preservado. Como contraponto a complexidade dessas questões, destacamos categorias possíveis para seleção de conteúdo, dimensões estruturais do processo de preservação digital e os desafios para promover políticas de preservação da Web brasileira.

De forma objetiva, podemos definir o arquivamento da web como um processo que compreende coletar, armazenar e disponibilizar a informação retrospectiva da World Wide Web para futuros pesquisadores. Para isso, a seleção é uma questão fundamental da web arquivamento. Dito de outra forma, o arquivamento na Web vai além das questões tecnológicas e da mera coleção de material da Internet. Questões relacionadas ao uso precisam ser vistas pela perspectiva organizacional, funcional e tecnológica. Pensar no uso

de uma iniciativa de arquivamento na Web abrange da seleção e coleta, ao gerenciamento e preservação da informação (ROCKEMBACH, 2018; MASANÈS, 2006).

Nesse contexto, o processo de seleção para arquivamento pode ser classificado em quatro categorias (BRAYNER, 2016; ROCKEMBACH, PAVÃO, 2018):

- a) seleção extensiva (que abrange domínios inteiros como .fr ou .br);
- b) seleção intensiva (que coleta em profundidade, arquivando os diversos níveis e hierarquias de um site selecionado por estratégias de curadoria)
- c) seleção temática: definida por critérios para escolha de um tema ou evento relevante, por exemplo, a coleta de conteúdo sobre o impacto social da COVID- 19;
- d) seleção autorizada: sites que possuem diretrizes de coleta autorizadas por seus editores e autores.

Devido à natureza descentralizada e a rápida obsolescência tecnológica, muitos desafios são impostos para preservação da Web. O ponto de partida para enfrentá-los é o planejamento e a definição de políticas de arquivamento (se centrado em um tema, site ou domínio), a metodologia e as plataformas para coletar, armazenar e disponibilizar o conteúdo Web.

Estas palavras-chave (coletar, armazenar e disponibilizar) correspondem às etapas do processo de arquivamento Web que, em certa medida, conforme a Figura 1, nos permitem visualizar três dimensões de preservação entrelaçadas:

- Dimensão política > baseada em questões de governança da internet, legislação, normas e padrões para coleta de conteúdo, procedimentos de preservação digital autorizada e condições de acesso;
- Dimensão tecnológica > baseada em tecnologias de preservação, aplicação de metadados, capacidade de interoperabilidade e nas ferramentas disponíveis para rastreamento, transição de formato e recuperação do conteúdo armazenado;
- Dimensão curatorial > baseada na escolha de estratégias para produção de conteúdo da Web, medidas de preservação seletiva ou integral, formas de exploração e reuso do conteúdo armazenado.



Figura 1: Dimensões de preservação do conteúdo Web.

Fonte: Elaborado pelo autor.

Estas dimensões configuram um quadro complexo para elaboração de estratégias e políticas de preservação e, quando projetados no contexto nacional, esbarramos numa lista de desafios (DANTAS, 2015):

- Apoio governamental e envolvimento de empresas públicas e privadas
- Definição de técnicas e tecnologias acessíveis e seguras
- Brechas jurídicas (por exemplo, regulamentação da Lei de Depósito Legal 10994/04 e 12192/10, que define: “publicações produzidas por qualquer meio ou processo, para distribuição gratuita ou venda”)
- Formação de competências profissionais e ampla inserção em programas acadêmicos
- Sustentabilidade econômica e tecnológica de serviços de arquivamento nacional
- Avaliação e monitoramento sistemático de coletas nacionais
- Acesso e difusão do conteúdo Web armazenado

Cabe ainda inserir nesse contexto de desafios a preocupação apresentada no relatório “Best Practice Forum on Local Content”, produzido em 2020 pelo Fórum de Governança da Internet. Nesse documento são destacadas questões sobre proteção, preservação e promoção das línguas locais e indígenas e do patrimônio cultural na era digital. Considerando a existência de conteúdos digitais vulneráveis, que registram idiomas ou práticas culturais de diversas comunidades, como de indígenas e artesões populares brasileiros, podemos integrar ao escopo de ações a segurança e preservação de heranças

culturais fragilizadas (IGF, 2020).

Buscar respostas para esses desafios indicam o primeiro passo para o desenvolvimento de um serviço de preservação e armazenamento da Web brasileira e, conseqüentemente, formação do patrimônio digital nacional.

## CONSIDERAÇÕES FINAIS

É inevitável mencionar que as interações sociais em ambiente digital intensificaram a produção e disseminação de informação. O contexto da sociedade digital indica, agora de forma mais evidente, que lidamos com um intenso processo de transformação do comportamento digital adquirido na última década. Além disso, a exemplo de outros momentos históricos, existem esforços mundiais para arquivar as diversas narrativas digitais da memória social: ações governamentais, conteúdo científico, textos jornalísticos, recursos educacionais, relações profissionais e trabalhistas, manifestações sociais em rede e seu impacto no cotidiano das pessoas.

Tomando esse momento como referência, ressaltamos a importância do arquivamento de conteúdo da Web não apenas para construção de registros digitais de memória, mas também para evitar um processo de esquecimento digital. Por essa perspectiva, parece urgente desenvolver projetos estruturais de apoio ao Web arquivamento, assim como ampliar abordagens de pesquisas sobre esse tema e analisar ações concretas sobre o patrimônio digital nacional, a amplitude de governança da internet, direitos e garantias de preservação, arquivamento, acesso à informação e ao conhecimento produzido na Web brasileira.

## REFERÊNCIAS

ABDALLA, Anita. Era do Erro 404: vamos perder a história registrada na internet? **UOL Tecnologia**, 15/08/2018. Disponível em: <https://www.uol.com.br/tilt/noticias/redacao/2018/08/15/a-era-do-erro-404-ja-esta-entre-nos.htm>. Acesso em: 20 jun. 2020.

ARQUIVO.PT. Arquivo.pt: informações acerca do serviço de pesquisa sobre o passado. 18/03/2020a. Disponível em: <https://sobre.arquivo.pt/pt/> Acesso em: 12 jun 2020.

\_\_\_\_\_. Criação e gestão de sites preserváveis: Recomendações do Arquivo.pt. [Slide]. 19/06/2020b. Disponível em: [https://sobre.arquivo.pt/wp-content/uploads/Sessao11\\_versao1.pdf](https://sobre.arquivo.pt/wp-content/uploads/Sessao11_versao1.pdf) Acesso em: 26 jun.2020.

BRAYNER, Aquiles Alendar. UK Web Archive Programme. A Brief History of Opportunities and Challenges. **Revista Digital de Biblioteconomia e Ciência da Informação (RDBCI)**. Campinas, v. 14, nº 2, mai/ago. 2016, p. 318–333. Disponível em: <<https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8645982>>. Acesso em: 10 jun. 2020.

BRÜGGER, Niels. When the present web is later the past: web historiography, digital history, and internet studies. **Historical Social Research**, 37 (2012) 4. p. 102–117.

COSTA, Miguel; GOMES, Daniel; SILVA, Mário J. **The evolution of web archiving** **International Journal on Digital Libraries**, p. 1-15, 2016.

DANTAS, Camila Guimarães. **Criptografias da memória: um estudo teórico-prático sobre o arquivamento da web no Brasil**. 2015.

DODEBEI, Vera. Patrimônio e Memória Digital. **Morpheus** - Revista Eletrônica em Ciências Humanas, v.4, n.8, 2006.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM (IIPC), c2020. Disponível em: <https://netpreserve.org/about-us/> Acesso em: 15 jun. 2020.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 28500:2009. Information and documentation - WARC file format**. Geneva: ISO, 2009. Disponível em: <https://www.iso.org/obp/ui/#iso:std:iso:28500:ed-1:v1:en> Acesso em: 18 jun.2020.

INTERNET ARCHIVE. About the Internet Archive. Disponível em: <https://archive.org/about/> Acesso em: 9 jun. 2020.

IGF – INTERNET GOVERNANCE FORUM. **Best Practice Forum on Local Content: Local and indigenous content in the digital space: Protection, preservation and sustainability of creative work and traditional knowledge**. United Nations, December 2020. Disponível em: [https://www.intgovforum.org/multilingual/index.php?q=filedepot\\_download/5005/2372](https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/5005/2372) Acesso em: 20 jan.2021.

MASANÈS, Julien. **Web Archiving**. Berlin, Heidelberg, Springer, 2006.

NEW YORK ART RESOURCES CONSORTIUM (NYARC). About [Página]. Disponível em<<http://nyarc.org/content/about>> Acesso em 22 jun. 2020.

ROCHA, Caroline. Quantos sites existem na internet? **OLHAR DIGITAL**, 30/09/2015. Disponível em: <https://olhardigital.com.br/noticia/quantos-sites-existem-na-internet/51803> Acessado em: 15 jul. 2020.

ROCKEMBACH, Moisés. Arquivamento da web: estudos de caso internacionais e o caso brasileiro. **Revista Digital de Biblioteconomia e Ciência da Informação (RDBCI)**. Campinas, v. 16, nº 1, já. /abr. 2018, p. 7-24. Disponível em:<<https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8648747/pdf>>. Acesso em: 10 jun. 2020.

ROCKEMBACH, Moisés; PAVÃO, Caterina Marta Groposo. Políticas e tecnologias de

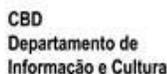
preservação digital no arquivamento da web. **RICI**: R. Ibero-Americana de Ciência da Informação. Inf., ISSN 1983-5213, Brasília, v. 11, n. 1, p. 168-182, jan. /abril.2018. Disponível em: < <https://lume.ufrgs.br/handle/10183/175153>>. Acesso em: 11 jun. 2020.

UNESCO. **Carta sobre la preservación del patrimonio digital** [Anexo N° 7], 15 out. 2003. Disponível em: <http://www.culturaydeporte.gob.es/planes-nacionales/dam/jcr:e840703d-133c-42e2-b3a2-90ab13e3bcdf/carta-sobre-la-preservaci-n-del-patrimonio-digital.pdf>. Acesso em: 20 jun. 2020.

**FRANCISCO CARLOS PALETTA** – Professor Livre Docente, Universidade de São Paulo, Escola de Comunicações e Artes, Departamento de Informação e Cultura. Pós-doutorado pelo Instituto de Pesquisas Energéticas e Nucleares. Doutor em Ciência pela Universidade de São Paulo. Mestre em Engenharia de Produção, Mestre em Gestion de l' Information et de la Connaissance Université Paul-Valéry Montpellier III - França. MBA em Altos Estudos de Estratégia e Geopolítica, MBA em Marketing, MBA em Data Science e Analytics, Bacharel em Engenharia Elétrica. Pós-doutorado em Ciência da Informação pela Universidade do Porto, Portugal. Doutor em Ciência da Informação pela Universidad Carlos III de Madrid. Líder do Grupo de Pesquisa Observatório do Mercado de Trabalho Profissional da Informação na Era Digital CNPq ECA USP. Docente no Programa de Pós-Graduação em Ciências da Informação PPGCI e Mestrado Profissional em Gestão da Informação.

**AUDILIO GONZALES AGUILAR** - Université Paul Valéry Montpellier III. Professor titular a Universidade Paul Valéry Montpellier III, Departamento de Documentação. Doutor em Direito e Informática pela Universidade de Montpellier I. Especialista acreditado pelo ministério de França em pesquisa e inovação para empresas. Membro do Laboratório Paragraphe EA 349 Universidad de París 8 . <http://paragraphe.site/> e <https://www.univ-paris8.fr/UR-Laboratoire-Paragraphe>

## REALIZAÇÃO



## FOMENTO



## COLABORAÇÃO CIENTÍFICA



## COLABORAÇÃO UNIVERSITÁRIA: ENSINO, PESQUISA E EXTENSÃO



## BIBLIOTECAS EM COLABORAÇÃO



## PARCEIROS INSTITUCIONAIS





**i2Δi**  
A Connected AI World



Pós-Graduação em Economia  
Comportamental e Tomada de Decisão



SÉRIE TECNOLOGIA E ORGANIZAÇÃO DA INFORMAÇÃO

# CURADORIA DIGITAL

Contribuições para a Ciência da Informação

 [www.atenaeditora.com.br](http://www.atenaeditora.com.br)

 [contato@atenaeditora.com.br](mailto:contato@atenaeditora.com.br)

 [@atenaeditora](https://www.instagram.com/atenaeditora)

 [www.facebook.com/atenaeditora.com.br](https://www.facebook.com/atenaeditora.com.br)

  
Ano 2023

SÉRIE TECNOLOGIA E ORGANIZAÇÃO DA INFORMAÇÃO

# CURADORIA DIGITAL

Contribuições para a Ciência da Informação

 [www.atenaeditora.com.br](http://www.atenaeditora.com.br)

 [contato@atenaeditora.com.br](mailto:contato@atenaeditora.com.br)

 @atenaeditora

 [www.facebook.com/atenaeditora.com.br](http://www.facebook.com/atenaeditora.com.br)

  
Ano 2023