

# Tendências para a gestão e preservação da informação digital

Organizadores

Miguel Ángel Márdero Arellano  
Luiza Martins de Santana Araújo







**Tendências para a gestão e preservação  
da informação digital**

*Tendencias para la gestión y preservación de la información digital*

Organizadores

Miguel Ángel Márdero Arellano

Luiza Martins de Santana Araújo

Brasília, DF  
2017

Diretoria

**Cecília Leite Oliveira**

Coordenação Geral de Pesquisa e Desenvolvimento de  
Novos Produtos (CGNP)

**Arthur Fernando Costa**

Coordenação Geral de Pesquisa e Manutenção de Produ-  
tos Consolidados (CGPC)

**Lillian Maria Araújo de Rezende Alvares**

Coordenação Geral de Tecnologias de Informação e Infor-  
mática (CGTI)

**Marcos Pereira Novais**

Coordenação de Ensino e Pesquisa, Ciência e Tecnologia  
Da Informação (COEPPE)

**Lena Vania Ribeiro Pinheiro**

Coordenação de Planejamento, Acompanhamento e Ava-  
liação (COPAV)

**José Luis dos Santos Nascimento**

Coordenação de Administração (COADM)

**Reginaldo de Araújo Silva**

Seção de Editoração

**Ramón Martins Sodoma da Fonseca**





**Tendências para a gestão e preservação  
da informação digital**  
*Tendencias para la gestión y preservación de la información digital*

Organizadores

Miguel Ángel Márdero Arellano

Luiza Martins de Santana Araújo

Brasília, DF

2017





## 2017 Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

Os autores são responsáveis pela apresentação dos fatos contidos e opiniões expressas nesta obra.

### Equipe técnica

#### Organizador

Miguel Ángel Márdero Arellano

#### Editor executivo

Ramón Martins Sodoma da Fonseca

#### Editoras assistentes

Gislaine Russo de Moraes Brito

#### Normalização de referências

Priscilla Mara Bermudes

#### Revisão gramatical e visual

Margaret de Palermo Silva

#### Projeto Gráfico

Seção de Editoração - Sedit/Ibict

#### Capa

Rodrigo Azevedo

#### Tradução

Seção de Editoração - Sedit/Ibict

---

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

Tendências para a gestão e preservação da informação digital [recurso eletrônico]  
/ organizadores: Miguel Ángel Márdero Arellano, Luiza Martins de Santana Araújo.  
- Brasília : 2017.

228 p.

ISBN número 978-85-7013-136-2

1. Preservação digital. 2. Gestão da informação. I. Tendências para a gestão e preservação da informação digital.

CDU 025.85(0.034.1)

---

### Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)

Setor de Autarquias Sul (SAUS)

Quadra 05, Lote 06, Bloco H - 5º Andar

Cep: 70070-912 - Brasília, DF

Telefones: 55 (61) 3217-6360 / 55 (61) 3217-6350

[www.ibict.br](http://www.ibict.br)

Rua Lauro Muller, 455 - 4º Andar - Botafogo

Cep: 22290-160 - Rio de Janeiro, RJ

Telefones: 55 (21) 2275-0321

Fax: 55 (21) 2275-3590

[http://www.ibict.br/capacitacao-e-ensino/pos-](http://www.ibict.br/capacitacao-e-ensino/pos-graduacao-em-ciencia-da-informacao)

[graduacao-em-ciencia-da-informacao](http://www.ppgci.ufrj.br)

<http://www.ppgci.ufrj.br>

# Comitê Editorial

## **Dunia Llanes Padrón**

Pós-Doutorado pela Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp) - SP, Brasil. Doutora em Biblioteconomia y Documentacion pela Universidad de Salamanca (USAL) - Salamanca, Espanha. Professora da Universidad de La Habana (UH) - Havana, Cuba.  
<http://lattes.cnpq.br/9392669707310310>  
<https://www.directorioexit.info/ficha4449>  
*E-mail:* [duniallp@yahoo.es](mailto:duniallp@yahoo.es)

## **Daniel Flores**

Pós-Doutorado pela Fundación Carolina/ Universidad de Salamanca (USal), Espanha. Doutor em Ciência da Informação pela Universidade Federal do Rio de Janeiro (UFRJ) - Rio de Janeiro, RJ - Brasil. Professor da Universidade Federal de Santa Maria (UFSM) - Santa Maria, RS - Brasil.  
<http://lattes.cnpq.br/9640543272532398>  
*E-mail:* [dfloresbr@gmail.com](mailto:dfloresbr@gmail.com)

## **Gildenir Carolino Santos**

Pós-Doutorado pela Universidade Estadual de Campinas (Unicamp), Brasil. Doutor em Educação pela Universidade Estadual de Campinas (Unicamp), Brasil. Bibliotecário da Universidade Estadual de Campinas (Unicamp) - Campinas, SP - Brasil.  
<http://lattes.cnpq.br/1221773207784315>  
*E-mail:* [gilbfe@unicamp.br](mailto:gilbfe@unicamp.br)

## **Miquel Tèrmens**

Doutor em Documentació pela Universitat de Barcelona (UB) - Barcelona, Espanha. Professor da Universitat de Barcelona (UB) - Barcelona, Espanha.  
<http://lattes.cnpq.br/0754437875262792>  
<http://bd.ub.edu/pub/terms/>  
*E-mail:* [termens@ub.es](mailto:termens@ub.es)

## **Rodrigo Rabello da Silva**

Pós-Doutorado pela Universidade de Brasília (UnB), Brasil. Pós-Doutorado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Brasil. Doutor em Ciência da Informação pela Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp) - SP, Brasil. Professor da Universidade de Brasília (UnB) - Brasília, DF - Brasil.  
<http://lattes.cnpq.br/3092147925440268>  
*E-mail:* [rdabello@gmail.com](mailto:rdabello@gmail.com)

---

<b>Prefácio</b>	<b>8</b>
<b>Acesso, memoria e preservação da informação digital</b> <i>Access, memory and preservation of digital information</i> <i>Acceso, memoria y preservación de la información digital</i> Emir Suaiden	
<hr/>	
<b>CAP. 1   La brecha digital urbano-rural de los jóvenes en Colombia: limitación a su ejercicio político</b>	<b>11</b>
<i>A brecha digital urbano-rural dos jovens na Colômbia: limitação ao exercício político</i> <i>The urban-rural digital divide of young people in Colombia: limitation to political exercise</i> Irene Sofía Romero Otero Adriana Otalora-Buitrago	
<hr/>	
<b>CAP. 2   Lecciones aprendidas en proyectos de gestión de documentos electrónicos</b>	<b>30</b>
<i>Lessons learned in electronic document management projects</i> <i>Lições aprendidas no projeto de gestão de documentos eletrônicos</i> Vicente González Johann Pirela Morillo Nelson Javier Pulido Daza	
<hr/>	
<b>CAP. 3   Valor probatorio de los documentos electrónicos: Visiones desde Colombia y Venezuela</b>	<b>53</b>
<i>Valor probatório dos documentos eletrônicos: pontos de vista de Colômbia e Venezuela</i> <i>Probative value of electronic documents: views from Colombia and Venezuela</i> Johann Pirela Morillo María Yaneth Álvarez Nelson Javier Pulido Daza	
<hr/>	
<b>CAP. 4   Ontología Digital Arquivística</b>	<b>77</b>
<i>Digital archival ontology</i> <i>Ontologia digital archivística</i> Charlley Luz	



---

**CAP. 5 | Dados de pesquisa: o que são, impactos do grande volume produzido, como organizá-los e quais preservar** 98

*Research data: what are the impacts of large volume produced, how to organize them and what to preserve*

*Datos de investigación: qué son, impactos del gran volumen producido, cómo organizarlos y cuáles preservar*

Maíra Murrieta Costa

Murilo Bastos da Cunha

Sonia de Assis Boeres

---

**CAP. 6 | Gestão dos dados de pesquisa: oportunidades e desafios** 143

*Research data management: opportunities and challenges*

*Gestión de los datos de investigación: oportunidades y retos*

Anaíza Caminha Gaspar

Lillian Alvares

Maria de Nazaré Freitas Pereira

---

**CAP. 7 | Preservação do patrimônio cultural da música brasileira** 165

*Preservation of cultural heritage of Brazilian music*

*Preservación del patrimonio cultural de la música brasileña*

Fernando William Cruz

Juliana Faria Silva

Luiza Beth Nunes Alonso

---

**CAP. 8 | La preservación digital y la red Cariniana** 200

*Preservação digital e a rede Cariniana*

*Digital preservation and the Cariniana Network*

Miguel Ángel Márdero Arellano

---

**Sobre os autores** 221

# ACESSO, MEMÓRIA E PRESERVAÇÃO DA INFORMAÇÃO DIGITAL

EMIR SUAIDEN

Professor Titular e Diretor da Biblioteca Central  
da Universidade de Brasília

emir@unb.br

Com as grandes mudanças que a sociedade globalizada vem consolidando nas últimas décadas, aliadas a fluidez da rede e o impacto social como elementos transformadores da realidade dos sistemas de informação é preciso estar atento às necessidades de atualização dos serviços e produtos oferecidos e do seu próprio conceito para que possam atender as demandas da sociedade contemporânea. Mesmo com o desenvolvimento dos sistemas de informação e seus profissionais vêm construindo e com as conquistas significativas alcançadas, ainda assim, é essencial promover mudanças para que as unidades de informação possam chegar a um modelo adequado à realidade que a atualidade exige.

A globalização que se realiza pela ampliação dos meios de comunicação e por uma nova base tecnológica, rompe diversos paradigmas e altera as relações entre os sujeitos, as comunidades e as instituições. Inaugura novos processos informacionais, diferentes formas de integração social e inovadoras formas produtivas e culturais podem ser inventadas.

Nesse contexto o tema da inclusão social na sociedade da informação tem obtido muita importância e vem sendo discutido tanto no âmbito acadêmico como no político e jurídico. Estar excluído é ocupar um lugar a margem de um processo é permanecer na periferia. Estar incluído é ter participação ativa na sociedade, é fazer valer seus direitos e deveres, é ser um cidadão.

Da mesma forma que o acesso à informação é indispensável na formação da cidadania é também vital na produção científica e na visibilidade internacional.

Quando se pensa em informação e conhecimento na atualidade, é necessário conceituar esses valores sob a ótica da educação, da economia e da cultura. É essencial identificar a dimensão humana, onde a educação é o cerne da questão; a dimensão tecnológica, onde o fator econômico tem fundamental importância, onde a cultura se apresenta como resultado das transformações ocorridas na sociedade.

O processo de transformação é global, atinge a todos os países direta ou indiretamente, mais ou menos intensamente. Nos países desenvolvidos a sociedade da informação e do conhecimento surge como consequência natural do processo de desenvolvimento. Nos países em desenvolvimento é necessária a adoção de estratégias políticas, educacionais, sociais e principalmente informacionais traduzidas em forma de programas, ações e formação adequada dos profissionais da informação que possibilitem a inclusão destas sociedades no mundo globalizado.

Uma característica fundamental da sociedade da informação é gerar conhecimento que seja sustentado pela produção da informação. O fenômeno da produção da informação é disseminado pelas estruturas informacionais (bibliotecas, bases e bancos de dados, sistemas de informação, arquivos e museus), que determinam a efetividade da produção do conhecimento. As tarefas realizadas por essas estruturas são: produção do conhecimento, produção da informação organizada e transferência da informação. Essas estruturas têm diferentes lógicas operacionais. As tarefas relacionadas com a informação são determinadas por várias questões, especialmente tecnológicas que buscam a eficiência técnica e econômica. Eficiência é vital para a produção do conhecimento, no entanto, não produz conhecimento de fato, simplesmente contribui para organizar a explosão da informação. Esse processo organizacional se torna legítimo quando o profissional da informação tem clareza nos setores de acesso, memória e preservação.

Nos últimos anos Colômbia e Brasil realizaram um grande investimento na formação de recursos humanos em nível de pós-graduação e na implantação de bibliotecas, sistemas de informação, etc.

Esse esforço acaba sendo o caminho da construção de uma sociedade mais justa. Onde o acesso, memória e a preservação digital são peças fundamentais para a construção de uma geração de usuários críticos e produtores de informação.

Nas décadas de sessenta e setenta os especialistas afirmavam que o processo de dependência informacional na América Latina era devido à falta de um público leitor, à falta de indústria de conteúdos e ao aumento da distância entre o cidadão e o livro.

Hoje, no entanto, as pesquisas comprovam que tão importante como ter acesso é fundamental a noção de compreensão no processo informacional. Compreensão para entender que a informação gera riqueza em todos os setores do conhecimento é pode ser responsável para diminuir os desequilíbrios sociais e promover a geração de emprego e renda.

A revolução tecnológica nos dá a oportunidade de recuperação da memória do patrimônio histórico, cultura e científico através da preservação digital. Sabemos que até hoje a memória do patrimônio de alguns países da América Latina está muito mais na Library of Congress ou na British Library do que nas respectivas Bibliotecas Nacionais. Temos urgentemente que recuperar a memória para deixarmos de ser dependentes informacionais, além de econômicos.

Nesta sequência a preservação digital passa a ser a grande prioridade para os investigadores latino-americanos terem consciência do passado e com isso possibilitar o acesso indiscriminado na sociedade da informação. Esse acesso, baseado na preservação digital, é fator essencial para a utilização da memória do patrimônio para a construção de cenários adequados para o desenvolvimento da ciência da informação em nossa região.

○○○

# La brecha digital urbano-rural de los jóvenes en Colombia: limitación a su ejercicio político

**IRENE SOFÍA ROMERO OTERO**  
**ADRIANA OTALORA-BUITRAGO**

### RESUMEN

En este capítulo se plantea una reflexión crítica sobre la incidencia de la brecha digital urbano rural en los jóvenes rurales de Colombia, las que derivan en implicaciones de tipo político y social. De manera sucinta se discute sobre la inequidad en Colombia y sus consecuencias en el ejercicio ciudadano y el empoderamiento político, para luego realizar un análisis en torno al papel que desempeñan las TIC, que lejos de promover la inclusión constituyen un mecanismo invisibilizado de exclusión digital. En tal escenario las bibliotecas y sus profesionales hacen parte de los actores líderes implicados en la búsqueda, implementación y promoción de estrategias de información incluyentes para la reducción de esta brecha y así contribuir a la construcción de una Colombia más equitativa y en paz.

**Palabras claves:** Jóvenes rurales, brecha digital, inclusión social y política, TIC, bibliotecas digitales.

## INTRODUCCIÓN

La problemática derivada de la existencia de una brecha digital tiende a abordarse a partir de criterios técnicos que propenden a la ampliación de cobertura en el acceso a las TIC. Sin embargo, causas y efectos de tal fenómeno ameritan un mayor nivel de profundidad, que implica una reflexión en la que se entrecruzan elementos relacionados con la equidad, la justicia y la formación de capacidades.

Es una discusión que implica la generación de condiciones para la construcción de paz en el marco de la promoción de condiciones equitativas y democráticas. En efecto, uno de los elementos para la construcción de paz es la inclusión de las expectativas de los distintos actores sociales en los consensos de aquello que se considera como deseable en una sociedad. En tal sentido, la existencia de profundas inequidades y exclusiones en términos económicos, políticos y sociales, ha signado las dinámicas de violencia propias de las sociedades contemporáneas.

Colombia ha sido históricamente uno de los países más inequitativos de América Latina, para el año 2014 ocupaba en términos de Desarrollo Humano el puesto número 97 entre 188 países a nivel mundial, y el puesto número 19 entre 32 países de América Latina (PNUD, 2015). Lo anterior, es producto de la profunda desigualdad reinante en este país, así lo evidencia un índice de Gini de 53.5 entre 2011 y 2015 (Banco Mundial, 2016), en contraste con un país como Argentina, que ocupó el puesto número 40 en el Índice de Desarrollo Humano, con un Gini de 42,3 para el mismo periodo. Valga aclarar que, en este último índice, entre más cerca se encuentra a 100, mayor es la desigualdad existente al interior del país y que en casos de países con un desarrollo humano muy alto, como el de Noruega que para el 2015 ocupó el primer lugar en el IDH, su Gini se encuentra cercano al 25,9, indicando una sociedad más igualitaria si bien nunca se alcanzará la igualdad completa, lo que correspondería a un índice de 1.0.

Tales resultados exponen la existencia tangible de brechas insalvables al interior de la sociedad, en tal sentido, algunos autores destacan que la participación política a través del ejercicio de la ciudadanía está actualmente condicionada por factores principalmente sociológicos y económicos, es decir aquellos ciudadanos especialmente los urbanos con un poder adquisitivo

mayor, se ubican en una posición ventajosa frente a aquellos ciudadanos en su gran mayoría rurales, que no tienen los medios económicos ni las posibilidades sociales para poder ejercer una verdadera ciudadanía activa.

En tal sentido Rawls (2006) plantea la existencia de una “posición original” que es oculta por el “velo de la ignorancia” a razón de lo cual los ciudadanos no perciben la situación de ventaja o desventaja en la que han nacido y viven. Así, las instituciones son las llamadas a promover el acceso a oportunidades para que los ciudadanos en una y otra posición logren una participación en la esfera de lo público, de manera que sus intereses se vean representados en un “consenso traslapado” o acuerdos sociales de convivencia. La paz por tanto, demanda la solución de tales “posiciones originales” o brechas y el ejercicio activo de la ciudadanía en aras de la convivencia y el progreso social.

A partir de los desarrollos teóricos de Rawls, Sen (2010) expone la necesidad de promover las capacidades humanas especialmente en aquellos individuos que se encuentran en una situación de desventaja, de manera que se garantice el ejercicio de sus libertades, en concordancia con la realidad. Así, una idea de la justicia válida debe indagar por lo fáctico y no únicamente por lo teórico propiciando una mayor cercanía con la realidad de los individuos que hacen parte de una sociedad, a partir de un reconocimiento más fidedigno de su cotidianeidad se posibilita en una mayor medida el ejercicio de sus libertades y la generación de oportunidades que promuevan el desarrollo humano.

En tal sentido, el abordaje a un problema como en este caso sería el de la brecha digital urbano-rural, requiere una aproximación a la realidad que permita develar las condiciones de ventaja y desventaja de los distintos grupos de ciudadanos como lo son los jóvenes, de manera que se identifiquen puntos de intervención para la disminución de tales brechas en la sociedad, toda vez que sus implicaciones alcanzan el ejercicio político y ciudadano constituyéndose como una limitante al ejercicio de los derechos de las poblaciones excluidas en términos de la promoción de sus capacidades, e incluso, en cuanto a la posibilidad de planificar su propio futuro.

## **ALGUNAS APROXIMACIONES FRENTE AL FENÓMENO DE LA EXCLUSIÓN**

Como respuesta al anterior escenario, han surgido prácticas ciudadanas alternativas fruto de los movimientos sociales y sus luchas políticas, evento debidamente abordado en el mundo académico por investigaciones de diverso corte al fenómeno de la exclusión y de las brechas derivadas de la misma, en consonancia con otros temas como las políticas de asistencialismo ante situaciones de una clara vulnerabilidad, como es el caso de los refugiados políticos, las que implican una relación de inferioridad del ciudadano frente a un Estado que no promueve sus capacidades; y a una relación cada vez más explícita entre ciudadanía y pobreza fruto de las brechas sociales existentes, donde las demandas sociales ante tal problemática y las respuestas políticas que se plantean provienen de posiciones divergentes, si bien existiría consenso en cuanto a las consecuencias e implicaciones de la pobreza en el ejercicio ciudadano y el desarrollo humano de quienes la padecen. A continuación, se exponen las tres tendencias enunciadas anteriormente:

En primer lugar, encontramos un conjunto de artículos que hacen referencia a las prácticas ciudadanas alternativas, respecto a la protesta social (Caldeira, 2015) realiza un análisis a partir de las protestas sociales llevadas a cabo en Brasil en 2013, estableciendo una comparación con los movimientos sociales de las décadas de los 70 y 80. Concluye que las transformaciones que ha sufrido el ejercicio ciudadano al contar con una mayor cohesión social, así como el efecto de la pauperización en las condiciones de vida de las periferias conlleva a nuevas prácticas políticas en las que emergen demandas como el acceso a infraestructura, a medios masivos de comunicación, a estructuras democráticas, a la disminución de la violencia y al incremento en el consumo.

En contraposición (Leibner, 2015) asevera la existencia de un debate ambiguo en el ejercicio ciudadano de un grupo judíos Mizrahi quienes en 2011 realizan protestas públicas en la búsqueda de acceso a vivienda, logrando actuar como una entidad política independiente de manera temporal, así como estableciendo alianzas temporales de todo tipo, como lo fuera con los palestinos. Es este un tipo de ejercicio ciudadano que demanda el acceso a bienes y servicios como solución a la pobreza, pero que no alcanza la suficiente cohesión identitaria y social como para lograr una incidencia política efectiva.



Es esta una visión que se hace a un lado de la concepción legal o formal de la ciudadanía, la cual predominó hasta la década de los 90, perdiendo cada vez más espacio en respuesta a las dificultades para materializar el “deber ser” en acciones concretas conducentes a la garantía de los derechos a los ciudadanos. En un contexto como el latinoamericano, este debate cobra aún más vigencia, toda vez que el interés por la realidad se hace manifiesto en la gran cantidad de estudios de caso en el ámbito de la ciudadanía y el ejercicio ciudadano (Oliveira, 2009; Wanderley, 2009; Riachi, 2008; Rocchi y Venticinque, 2010; Gattoni, 2010; Gutiérrez, 2011; Saucedo y Taracena, 2011; Martínez, et al., 2010; Fernández Leffort, 2011; Costa y Teodósio, 2011; Ocampo, 2011; Madeiros, 2011; Mora y Anaya, 2013).

En segundo lugar, puede identificarse una preocupación frente al asistencialismo, producto de las respuestas estatales ante la condición de vulnerabilidad de poblaciones específicas. Así y en contraposición a los estudios de caso mencionados anteriormente en los que predomina el interés por comprender las dinámicas del ejercicio ciudadano, en esta nueva tendencia interesan los efectos sobre grupos particulares de las acciones del Estado u otros grupos como ONG. En la medida en que tales acciones se orientan desde una visión asistencialista, estas tienen un efecto lesivo sobre tales grupos humanos, llegando incluso a restringir sus posibilidades de organización, movilización y ejercicio ciudadano.

Es así como (Perdigon, 2015) explora la transformación de las subjetividades de tres refugiados palestinos en Líbano, dadas las condiciones de pobreza en las que sobreviven, las cuales se mantienen como resultado de las fuerzas políticas internacionales y derivan en un desentrañamiento del sí mismo de los refugiados. En el mismo sentido (Ross, 2015) señala a través de un trabajo etnográfico en Suráfrica la manera en que la pobreza afecta los conceptos de respeto y crudeza para quienes deben vivirla, modificando sus concepciones de ciudadanía diferenciada, así como sus formas de interacción social.

Por su parte (Cox, 2011), analiza las dinámicas históricas que han llevado a la discriminación racial de ciertas poblaciones en los Estados Unidos como quienes provienen de China, Japón, Vietnam, Sudáfrica, Cuba o Trinidad y Tobago, encontrando como común denominador el hecho de que estos inmigrantes hayan visto en su momento condicionada su estadia en Estados

Unidos a condiciones laborales precarias; así, la discriminación de la cual son objeto se ve justificada por las necesidades y restricciones laborales propias de su proceso de inmigración.

Son estos estudios que se adentran en el campo de las subjetividades políticas, cuestionando la forma en que las acciones públicas pueden derivar en la generación de una serie de condiciones que limitan el autoconcepto del individuo; así a través de la subsistencia en un contexto explícito de pobreza y de escasez los beneficiarios de ciertos programas sociales generan una mayor dependencia frente al Estado y en tanto dependencia, una restricción en su concepto de autonomía y ejercicio ciudadano.

En tercer lugar, se analiza la relación existente entre ciudadanía y pobreza, en tal sentido (Das & Randeria, 2015) analizan la forma en que la pobreza urbana emerge en el discurso público como producto de las demandas que realizan las comunidades pobres al Estado. En algunas ocasiones los Estados tienden a centrar su discurso en prácticas eficientistas conducentes al debilitamiento de los sistemas de aseguramiento social y a la restricción en la garantía de derechos a poblaciones vulnerables (Eckenwiler, 2011).

Tal discurso público es adoptado por otros actores como las transnacionales (Schwittay, 2011; Martin & Hill, 2012) que cuestionan las dinámicas existentes en la relación de mercado entre tales organizaciones y las poblaciones pobres, toda vez que las primeras crean necesidades frente a los cuales este tipo de población puede única y exclusivamente actuar como protoconsumidores, profundizando incluso su concepción de pobreza. Sin embargo, las respuestas que presentan tales individuos ante el empobrecimiento no mantienen una tendencia definida.

En la medida en que el mercado se constituye como un mecanismo de dominación de las clases económicas más favorecidas y por tanto productoras, sobre las clases sociales menos favorecidas, estas últimas ven restringido su derecho a la elección, el cual es condicionado por el mercado.

## **APORTES A LA CONSTRUCCIÓN DE PAZ EN COLOMBIA REQUIERE DE JÓVENES FORMADOS POLÍTICAMENTE E INCLUIDOS DIGITALMENTE**

En el caso colombiano y en un contexto histórico de construcción de paz, resulta de especial interés conocer la forma en que los jóvenes rurales son afectados directa o indirectamente por el fenómeno de la exclusión y de las brechas urbano-rural existentes.

En esta línea hay varios estudios (Caputo, 2006; Osorio, 2005 p. 125; Haimovich y Winkler, 2005; Bérlogo y Carvajal, 2008; Hopenhayn, 2008; Osorio, Jaramillo & Orjuela, 2010) que evidencian la existencia de las brechas y que sitúa en desventaja al mundo rural frente al urbano, en materia política, económica, social y de manera especial a los jóvenes rurales.

Centrando la mirada en esta población, resulta interesante conocer las prácticas y experiencias políticas de los jóvenes rurales, pues estas constituyen un camino para la inclusión de sus comunidades en condiciones más equitativas, como un mecanismo alternativo frente a las respuestas asistencialistas del Estado, las que en muchas ocasiones disminuyen al individuo.

En el caso Latinoamericano son varios los estudios que plantean la necesidad de estudiar los jóvenes rurales, pues se trata de una población clave con un gran potencial para el desarrollo de sus comunidades (Espíndola, 2002, p. 2; Jurado y Tobasura 2012; Gaviria y Luna, 2013; Patiño, Alvarado y Ospina, 2014).

En Colombia la participación política de los jóvenes rurales es muy tímida a través de “diferentes acciones colectivas, hacen visible su papel en el ámbito político y sociocultural, dan muestras de ir más allá del presente inmediato y de sus propias necesidades, para participar y dinamizar su comunidad. Con el referente identitario de “grupo juvenil”, acompañado por la pertenencia territorial, se va definiendo un “nosotros” desde el cual se intentan varias iniciativas. Algunas más autónomas y otras más influenciadas por las “oportunidades políticas” (Tarrow, 1997 en Osorio, 2005 p. 130), se ven acompañadas con frecuencia por el clientelismo y son tocadas, en mayor o menor grado, por la maraña de intereses y presiones propias de la guerra” (Osorio, 2005 p. 130).

Igualmente, Agudelo et al, en los resultados del proyecto titulado: *Sentidos de la participación ciudadana presentes en las prácticas políticas de jóvenes en contextos urbanos y rurales* evidenciaron que la juventud rural del municipio de Jamundí “se encuentran alrededor de sus procesos organizativos en el territorio y convergen desde su liderazgo en cada corregimiento, siendo referentes de vida para otros jóvenes, al tiempo que sus hogares, espacios académicos y ámbitos formales de incidencia política están atravesados por sus particulares apuestas transformadoras” (Agudelo, Murillo, Echeverry & Patiño, 2013, p. 593).

Además, sus formas de participación ciudadana son multidimensionales proyectando las causas y apuestas tanto a nivel personal como colectivo, a través del desarrollo de liderazgos polifacéticos llevados a cabo en sus propias comunidades y localidades es decir en su cotidianidad. Es por esto que se consideran como agentes de cambio en su trayectoria de liderazgo regional y como sujetos políticos (Agudelo, Murillo, Echeverry & Patiño, 2013, p. 601).

Es importante mencionar que cuando los jóvenes rurales enfocan su mirada hacia su propio contexto local “sienten la necesidad de reconocer y dinamizar los escenarios de acción política al afrontar una comunicación local o una comunicación ciudadana; con todo ello, se fortalece otra comunicación donde los jóvenes inciden en su cotidianidad y en las políticas públicas que definen su presente y su futuro” (Garcés y Betancur, 2006, p. 88). Es por esto que a través del uso de una comunicación ciudadana efectiva estos jóvenes pueden empoderarse y ser ciudadanos más activos frente a la transformación social y política del país.

Colombia se encuentra en un momento histórico de tránsito hacia la paz, el cual demanda la comprensión de diferentes dinámicas y prácticas en lo público que involucren los distintos actores sociales que hacen parte de la nación. Siguiendo la argumentación de Sen (2010) la consecución de una sociedad justa requiere el reconocimiento de las realidades y la comprensión del mundo de los distintos grupos sociales que se enlistan a la construcción de un nuevo modelo de sociedad, lo que es un requisito para la construcción de paz en Colombia.

Así, la inclusión de aquellos sectores que han sido marginados de la vida política del país se convierte en una pieza clave en la promoción de una cultura de paz y de un ejercicio político activo donde se respeten las individualidades y particularidades culturales de cada una de las comunidades que conforman una identidad nacional.

En tal sentido, cobra especial importancia la comprensión de las realidades y las dinámicas políticas de los jóvenes, en particular de los jóvenes rurales, quienes han estado por generaciones excluidos de los escenarios de toma de decisiones y participación política en Colombia.

Se asume que la formación política de los jóvenes rurales o de cualquier otro grupo de jóvenes, pasa por la comprensión y validación de sus formas de interacción, comunicación y participación, así como por la discusión de sus referentes ético morales, los que subyacen a su comprensión de lo político y por ende justifican sus formas de acción política y ejercicio ciudadano.

En este escenario, signado por los condicionantes de la sociedad de la información en la que nos encontramos donde las TIC como herramienta de interconexión e intercambio de información desempeñan un papel protagónico, donde deberían contribuir a la reducción de las brechas en general, y a la brecha digital en particular promoviendo una sociedad más justa e igualitaria.

No obstante, las TIC en los entornos urbanos han representado una mayor oportunidad de ejercicio ciudadano alternativo, promoción de las capacidades y empoderamiento digital, político e incluso económico de los jóvenes; mientras que en el caso de los jóvenes rurales al tener un limitado acceso a ellas y una escasa alfabetización digital, se ven enfrentados a una exclusión estructural, convirtiendo a las TIC en un escenario de desventaja frente a sus pares urbanos.

## **LAS TIC COMO INSTRUMENTO Y MECANISMO DE EMPODERAMIENTO**

Es una realidad que en la sociedad de la información, las TIC están desempeñando un papel protagónico de comunicación digital en los ciudadanos, esta situación ha generado nuevos escenarios de articulación social en el que el concepto de ciudadanía se introduce en los entornos digitales, impactando tanto en los ámbitos, económicos, políticos, sociales, éticos, culturales, entre otros.

Es por ello que gradualmente en diferentes partes del mundo van surgiendo estudios puntuales que evidencian el papel protagónico que pueden desempeñar los ciudadanos a través del uso de las TIC para ejercer su ciudadanía digital en los entornos virtuales. Muestra de esto son los resultados de los diferentes estudios (Bakker & de Vreese, 2011; Sampedro, Sánchez & Poletti 2013; Castells, 2012; Sloam, 2014; Xenos et al., 2014; Pardo, Pardo, 2015) que evidencian especialmente en la población juvenil una relación positiva en los diferentes usos que hacen de internet en relación con su participación política.

Asimismo, estos estudios dan indicio de que sí se está presentando un fortalecimiento de la ciudadanía digital global especialmente en la juventud, es decir, los jóvenes del mundo están manifestando preocupaciones de las diferentes problemáticas mundiales, están convocando, están movilizando masas a través del uso de las TIC y de un activismo virtual por medio de los blogs, redes sociales, páginas web, entre otras herramientas de movilización ciudadana reflejando que esta población se encuentra en un proceso de intensa politización es decir, tienen deseo de participar en la política pero no a través de los medios convencionales sino aprovechando las ventajas que ofrece el internet (Berlinguer y Martínez, 2014, p. 41). No obstante, es importante mencionar que para poder ser un ciudadano digital este debe tener básicamente tres requisitos mínimos tal como lo plantea Robles: “acceso a internet, habilidades digitales y percepción de la utilidad de las tecnologías” (Robles, 2009, p. 55).

Por otro lado, hay diferentes investigaciones en varias partes del mundo que han detectado la existencia real de una marcada brecha digital especialmente entre los ciudadanos urbanos y rurales (Oyedemi, 2015, p. 20; Villa & Moncada, 2011, p. 28; Akca, Sayili & Esengun, 2007). Centrando la mirada en los sectores

rurales, actualmente hay diversos proyectos a nivel latinoamericano y nacional encaminados a mejorar la gobernanza democrática en estos lugares con la implementación y uso de las TIC (MINTIC, 2015, p. 6-7; Villarroel, 2013, p. 79; MINSEGPRES, 2012; Bertón, Totorica y González, 2013), pero desafortunadamente muchas de estas iniciativas gubernamentales se han quedado únicamente en la etapa de la implementación tecnológica cuyos resultados se ven reflejados en el ascenso de las cifras de ciudadanía digital (todo parece un número más en los indicadores) y poco se ha pensado en cómo el ciudadano rural puede hacer un uso integral de las tecnologías acompañado de un proceso de formación y alfabetización digital para que de esta manera ellos mismos hagan conciencia del poder que tienen en sus manos para hacer cumplir sus derechos, empoderarse y participar activamente como ciudadanos digitales.

Es por ello que existen diferentes propuestas en torno al interés de disminuir la brecha digital, por ejemplo en el caso de la IFLA/UNESCO a través de su manifiesto sobre las bibliotecas digitales plantea la urgente necesidad de subsanarla y propone a los gobiernos como estrategia llevarlo a cabo a partir de la creación, fortalecimiento y uso de las bibliotecas digitales, pues las bibliotecas “han sido desde hace mucho tiempo instrumentos esenciales para fomentar la paz y los valores humanos. Su actual funcionamiento digital abre un nuevo cauce al universo del conocimiento y la información, estableciendo contactos entre culturas separadas por fronteras geográficas y sociales” (IFLA/UNESCO, 2013).

Por tanto, no cabe duda de que las bibliotecas en general y las digitales en particular se convierten en escenarios propicios de inclusión digital y para lograrlo se deben promover las siguientes estrategias tal como lo plantea Digital Inclusion Survey: “a). Proporcionar el libre acceso a las tecnologías de acceso públicos (hardware, software y de alta velocidad de conexión a Internet) en sus comunidades. b). Proporcionar acceso a una amplia gama de contenido digitales para sus comunidades. c). Proporcionar servicios de alfabetización digital que ayuden a las personas a desenvolverse, comprender, evaluar y crear contenido digital utilizando una amplia gama de tecnologías de la información y de la comunicación. d). Proporcionar programas y servicios clave para toda la comunidad en áreas como salud, bienestar, educación, empleo, desarrollo personal, y compromiso cívico” (Digital Inclusion Survey, 2014).

Así pues, no cabe duda que las bibliotecas y sus profesionales como agentes de promoción de la información y de la democracia, también desempeñan un papel esencial en la superación de la brecha digital pues propician el desarrollo de capacidades y a su vez generan oportunidades y para el caso del contexto colombiano se hace necesario llevar a cabo un proceso de formación pedagógica, inclusión y alfabetización digital con los ciudadanos rurales de diferentes zonas del país, con el fin de brindarles las herramientas teórico-prácticas necesarias para que por medio de un uso consciente de las TIC puedan ejercer una real ciudadanía digital desde sus localidades y así contribuir a través de su participación política y social en la construcción de la paz de la nación, donde sus voces sean escuchadas y donde puedan ejercer realmente sus derechos y deberes como ciudadanos colombianos y también como ciudadanos del mundo. Así pues, se convertirán en actores protagónicos y partícipes en el proceso de construcción de paz y así ayudarán a abrir nuevos escenarios de diálogo e intercambio de ideas, que de una u otra manera contribuyan a disminuir las brechas existentes y por ende empezar a edificar una cultura de paz de la mano de las TIC y apoyados por las bibliotecas.



## CONCLUSIONES

Las limitaciones en el acceso a las TIC y la escasa alfabetización digital en el caso de los jóvenes rurales en el país profundizan aún más las brechas digitales, sociales, políticas, culturales y económicas, fruto de la exclusión histórica a la que se han visto enfrentadas las poblaciones rurales en Latinoamérica.

En el caso colombiano, al ser este uno de los países más inequitativos de América Latina, tiende a reproducir de manera intergeneracional los patrones de la relación urbano-rural signados por la injusticia y la limitación de oportunidades. De esta forma, el mundo rural colombiano está ofreciendo a sus jóvenes posibilidades limitadas de acceso a las TIC, condenándolos a una situación de estancamiento y desventaja frente a sus pares urbanos.

Las TIC deberían ser una herramienta de inclusión, orientadas al adecuado manejo de la información. Sin embargo, desafortunadamente las TIC se han convertido en un mecanismo de exclusión digital la cual se profundiza aún más la brecha digital urbano-rural existente; esto se evidencia notoriamente en poblaciones tradicionalmente excluidas como es el caso de los jóvenes rurales colombianos.

Por otra parte, en términos de intervención, las TIC con apoyo y gestión del Estado, las bibliotecas, los profesionales de la información, entre otros, deberían constituirse como una oportunidad real de empoderamiento político y ciudadano de los jóvenes rurales colombianos, contribuyendo a cerrar la brecha digital existente y facilitando el ejercicio de formas alternativas de ciudadanía, democracia y construcción de paz.

Dada su función social, las bibliotecas en general y las digitales en particular, juegan un papel protagónico en la transformación de este escenario para reducir la brecha digital generando implicaciones sociales, culturales y políticas, aportando al fortalecimiento de la democracia y de construcción de paz. Así, el bibliotecólogo se constituirá en un líder y gestor de iniciativas de cambio social.



## REFERENCIAS

- AGUDELO, A. et al. Participación ciudadana y prácticas políticas de jóvenes en la cotidianidad. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, v. 11, n. 2, p. 587-602, 2013. Disponible en: <<http://revistaumanizales.cinde.org.co/index.php/Revista-Latinoamericana/article/view/937/444>>. Acceso en: 22 Mayo 2017.
- AKCA, H.; SAYILI, M.; ESENGUN, K. Challenge of rural people to reduce digital divide in the globalized world: theory and practice. *Government Information Quarterly*, v. 24, n. 2, p. 404-413, 2007. Disponible en: <<http://www.sciencedirect.com/science/article/pii/S0740624X06001663>>. Acceso en: 29 Mayo 2017. Disponible en: <<https://doi.org/10.1016/j.giq.2006.04.012>>. DOI: 10.1016/j.giq.2006.04.012
- BAKKER, T; VREESE, C. de. Good news for the future?: young people, internet use, and political participation. *Commun. Research*, v. 38, n. 4, p. 451-470. Disponible en: <<http://journals.sagepub.com/doi/pdf/10.1177/0093650210381738>>. Acceso en: 29 Mayo 2017.
- BÉRGOLO, M.; CARBAJAL, F. *Brecha urbana: rural de ingresos laborales en Uruguay para el año 2006: una descomposición por cuantiles*. Argentina: Universidad Nacional de la Plata, 2008. Documento de trabajo, n. 75. Disponible en: <[http://www.scielo.cl/scielo.php?script=sci\\_arttext&pid=S0718-88702010000200006](http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-88702010000200006)>. Acceso en: 29 Mayo 2017.
- BERLINGUER, M.; MARTINEZ, R. *Desconfiados: suspendidos entre búsqueda, resignación y revuelta: una situación inestable. Jóvenes, internet y política*. En Subirats, J.; Fuster, M.; Martínez, R.; Berlinguer, M. y Salcedo, J. (2014). Jóvenes, Internet y política. Madrid: Centro Reina Sofía sobre Adolescencia: Juventud y Fundación de Ayuda contra la Drogadicción. 2014. Disponible en: <[http://adolescenciayjuventud.org/que-hacemos/monografias-y-estudios/ampliar.php/Id\\_contenido/126670/](http://adolescenciayjuventud.org/que-hacemos/monografias-y-estudios/ampliar.php/Id_contenido/126670/)>. Acceso en: 29 Mayo 2017.
- BERTÓN, J.; TOTORICA, P.; GONZÁLEZ, S. *Estudio diagnóstico participación ciudadana electrónica*: AGESIC. 2013. Disponible en: <[http://www.agesic.gub.uy/innovaportal/file/3358/1/obsclu\\_infreferentes\\_v12.pdf](http://www.agesic.gub.uy/innovaportal/file/3358/1/obsclu_infreferentes_v12.pdf)>. Acceso en: 29 Mayo 2017.
- CALDEIRA, T. P. Social movements, cultural production, and protests: São Paulo's shifting political landscape. *Current Anthropology*, v. 56, n. 11, p. 126-136, 2015. Disponible en: <<https://doi.org/10.1086/681927>>.
- CAPUTO, Luis. *Estudios sobre juventud rural en América Latina: limitaciones y desafíos para una agenda de investigación sobre juventud rural*. 2006. Disponible en: <<http://juventudruralemprendedora.procasur.org/wp-content/uploads/2013/08/Ana%CC%81lisis-sobre-los-Estudios-sobre-Juventud-Rural-en-Ame%CC%81rica-Latina..pdf>>. Acceso en: 29 Mayo 2017.
- CASTELLS, Manuel. *Redes de indignación y esperanza*. Madrid: Alianza, 2012.

CAP 1 | La brecha digital urbano-rural de los jóvenes en Colombia: limitación a su ejercicio político

CHILE. Ministerio Secretaría General de la Presidencia. *Transparencia y participación a través de un gobierno abierto e interactivo: experiencia: estrategia de gobierno abierto de Chile*. Chile, 2012. Disponible en: <[http://centrodeinnovacion.gobiernoenlinea.gov.co/sites/default/files/fichaexp\\_2015\\_24\\_datosabiertoschile\\_v1.pdf](http://centrodeinnovacion.gobiernoenlinea.gov.co/sites/default/files/fichaexp_2015_24_datosabiertoschile_v1.pdf)>. Acceso en: 29 Mayo 2017.

COLÔMBIA. Ministerio de Tecnologías de la Información y las Comunicaciones. *Informe de gestión al congreso de la república*. Colômbia, 2015. Disponible en: <[http://www.mintic.gov.co/portal/604/articles-13320\\_doc\\_.pdf](http://www.mintic.gov.co/portal/604/articles-13320_doc_.pdf)>. Acceso en: 30 Mayo 2017.

COSTA, D. V. da; TEODÓSIO, A. dos S. de S. Desenvolvimento sustentável, consumo e cidadania: um estudo sobre a (des) articulação da comunicação de organizações da sociedade civil, do estado e das empresas. *Revista de Administração Mackenzie*, v. 12, n. 3, p. 114-145, 2011. <https://doi.org/10.1590/S1678-69712011000300006>

COX, O. C. Race prejudice, class conflict, and nationalism. *Race/Ethnicity*, v. 4, n. 12, p. 169-182, 2011.

DAS, V.; RANDERIA, S. Politics of the urban poor: a esthetics, ethics, volatility, precarity an introduction to supplement 11. *Current Anthropology*, v. 56, n. 11, p. 3-14, 2015. <https://doi.org/10.1086/682353>

DIGITAL inclusion survey: digital inclusion builds communities today (and tomorrow). 2014. Disponible en: <<http://digitalinclusion.umd.edu/>>. Acceso en: 29 Mayo 2017.

ECKENWILER, L. Women on the move: long-term care, migrant women, and global justice. *International Journal of Feminist Approaches to Bioethics*, v. 4, n. 2, p. 1-31, 2011. <https://doi.org/10.3138/ijfab.4.2.1>

ESPÍNDOLA, D. Nuevo enfoque de políticas públicas de juventud rural. In: CONGRESO DE LA ASOCIACIÓN LATINOAMERICANA DE SOCIOLOGÍA RURAL, 2002. *Anales electrónicos...* Disponible en: <<http://juventudruralemprendedora.procasur.org/wp-content/uploads/2013/08/NUEVO-ENFOQUE-en-POLI%CC%81TICAS-PU%CC%81BLICAS-de-JUVENTUD-RURAL.pdf>>. Acceso en: 29 Mayo 2017.

FERNANDEZ LEFFORT, I. Ciudadanía y desarrollo: conceptos, diferencias y tensiones: el caso de la campaña por la reducción de la mortalidad materna en Méjico. *Tiempos de Paz*, n. 102, p. 77-83, 2011.

FERREIRA, E. M. La ciudadanía y la identidad de los sujetos en la modernidad oculta. *Seqüência*, n. 33, p. 20-41, 2011.

GARCÉS, Ángela; BETANCUR, Tatiana. Comunicación alternativa: una mirada a través de las agrupaciones juveniles. *Anagramas*, v. 5, n. 9, p. 83-100, 2006. Disponible en: <<http://revistas.udem.edu.co/index.php/anagramas/article/view/781/723>>. Acceso en: 29 Mayo 2017.

GATTONI, M. S. Dispositivos participativos en el municipio bonaerense de Morón: reflexiones en torno al estado local, la ciudadanía y la legitimidad democrática. *Opera*, n. 10, p. 83-102, 2010. Disponible en: <<http://revistas.uexternado.edu.co/index.php/opera/article/view/3100>> . Acceso en: 29 Mayo 2017.

GAVIRIA-LONDOÑO, M. B.; LUNA-CARMONA, M. T. Pluralidad humana en el destierro. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, v. 11, n. 2, p. 475-491, 2013. Disponible en: <<http://www.scielo.org.co/pdf/rlcs/v11n2/v11n2a03.pdf>>. Acceso en: 29 Mayo 2017.

GUTIÉRREZ VIDRIO, S. Representaciones sociales y construcción de la ciudadanía en jóvenes universitarios. *Sinéctica*, n. 36, p. 1-18, 2011. Disponible en: <<http://www.redalyc.org/pdf/998/99815920002.pdf>>. Acceso en: 29 Mayo 2017.

HAIMOVICH, Francisco; WINKLER, Hernán. *Pobreza rural y urbana en Argentina: un análisis de descomposiciones*. 2005. Argentina: Universidad Nacional de la Plata, 2005. Documento de Trabajo, n. 25. Disponible en: <<http://sedici.unlp.edu.ar/handle/10915/3585>>. Acceso en: 29 Mayo 2017.

HOPENHAYN, Martin. Inclusión y exclusión social en la juventud latinoamericana. *Pensamiento Latinoamericano*, n. 3, p. 49-71, 2008. Disponible en: <<https://dialnet.unirioja.es/servlet/articulo?codigo=2781553>>. Acceso en: 29 Mayo 2017.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS (IFLA). *Manifiesto de las IFLA/ UNESCO sobre las bibliotecas digitales*. 2013. Disponible en: <<http://www.ifla.org/ES/publications/manifiesto-de-las-ifla-unesco-sobre-las-bibliotecas-digitales?og=2431>>. Acceso en: 29 Mayo 2017.

JURADO, C.; TOBASURA, I. Dilema de la juventud en territorios rurales de Colombia: ¿campo o ciudad?. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, v. 10, n. 1, p. 63-77, 2012. Disponible en: <<http://www.scielo.org.co/pdf/rlcs/v10n1/v10n1a03.pdf>>. Acceso en: 29 Mayo 2017.

LEIBNER, G. HaTikva encampment, 2011: the ambiguous agency of the marginalized. *Current Anthropology*, v. 56, n. 11, p. 159-168, 2015. <https://doi.org/10.1086/682241>

MARTIN, K. D.; HILL, R. P. Life satisfaction, self-determination, and consumption adequacy at the bottom of the pyramid. *Journal of Consumer Research*, v. 38, n. 6, p. 1155-1168, 2012. <https://doi.org/10.1086/661528>

MARTÍNEZ, M. L.; SILVA, C.; HERNÁNDEZ, A. C. ¿En qué ciudadanía creen los jóvenes?: creencias, aspiraciones de ciudadanía y motivaciones para la participación sociopolítica. *PSYKHE*, v. 19, n. 2, p. 25-37, 2010. <https://doi.org/10.4067/S0718-22282010000200004>

OCAMPO, Á. M. Ciudadanía juvenil, juventud y estado: discursos de gobierno sobre sus significados. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, v. 9, n. 1,

p. 287-303, 2011. Disponible en: <<http://revistaumanizales.cinde.org.co/index.php/Revista-Latinoamericana/article/view/357>>. Acceso en: 29 Mayo 2017.

OLIVERA RODRIGUEZ, I. Juventud rural y lucha por la ciudadanía: límites y posibilidades en los procesos de socialización. *Antropológica*, v. XXVII, n. 27, p. 7-27, 2009. Disponible en: <[http://www.scielo.org.pe/scielo.php?script=sci\\_arttext&pid=S0254-92122009000100002](http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S0254-92122009000100002)>. Acceso en: 29 Mayo 2017.

CAP 1 | La brecha digital urbano-rural de los jóvenes en Colombia: limitación a su ejercicio político

OSORIO, Flor. Jóvenes rurales y acción colectiva en Colombia. *Nómadas*, n. 23, p. 122-131, 2005. Disponible en: <<http://www.redalyc.org/pdf/1051/105116741014.pdf>>. Acceso en: 29 Mayo 2017.

\_\_\_\_\_ ; JARAMILLO, Olga; ORJUOLA, Amanda. Jóvenes rurales: identidades y territorialidades contradictorias: algunas reflexiones desde la realidad colombiana. *Boletín del Observatorio Javeriano de Juventud*, 2010. Disponible en: <[http://puj-portal.javeriana.edu.co/portal/page/portal/Centro\\_Atico/pruebas2/boletin\\_ojj/recursos\\_ojj2/OJJ\\_Tema%20central\\_b1.pdf](http://puj-portal.javeriana.edu.co/portal/page/portal/Centro_Atico/pruebas2/boletin_ojj/recursos_ojj2/OJJ_Tema%20central_b1.pdf)>. Acceso en: 29 Mayo 2017.

OYEDEMI, Toks. Participation, citizenship and internet use among South African youth. *Telematics and Informatics*, n. 32, p. 11-22, 2015. Disponible en: <<https://doi.org/10.1016/j.tele.2014.08.002>>. Acceso en: 29 Mayo 2017.

PARDO, Miguel Ángel. Nuevas formas de participación ciudadana en Chile: hacia una democracia del siglo XXI. *Revista de Estudios Políticos y Estratégicos*, v. 3, n. 1, p. 14-41, 2015. Disponible en: <<http://revistaepe.blogutem.cl/files/2015/09/revista-epe-vol3-n1-nuev-formas-de-participacion-ciudadana-en-chile-hacia-una-democracia-del-siglo-XXI-pag14-41.pdf>>. Acceso en: 29 Mayo 2017.

PATIÑO, J.; ALVARADO, S. V. ; OSPINA-ALVARADO, M. C. Ampliación de sentidos sobre las prácticas políticas de jóvenes con vinculación a siete movimientos sociales en Colombia. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, v. 12, n. 1, p. 257-275, 2014. Disponible en: <[http://revistalatinamericanaumanizales.cinde.org.co/wp-content/uploads/2014/03/Vol\\_12\\_n\\_1/articulo15.pdf](http://revistalatinamericanaumanizales.cinde.org.co/wp-content/uploads/2014/03/Vol_12_n_1/articulo15.pdf)>. Acceso en: 29 Mayo 2017.

PERDIGON, S. "For us it is otherwise" three sketches on making poverty sensible in the palestinian refugee camps of Lebanon. *Current Anthropology*, v. 56, n. 11, p. 88-96, 2015. Disponible en: <<http://www.journals.uchicago.edu/doi/abs/10.1086/682354>>

RAWLS, John. *Teoría de la justicia*. México: Fondo de Cultura Económica, 2006.

RIACHI, M. C. Lugares comunes: ciudadanía, comunidad y pluralismo jurídico. *Pensares*, v. 5, p. 529-542, 2008.

ROBLES, J. M. *Ciudadanía digital: una introducción a un nuevo concepto de ciudadano*. España: Editorial UOC, 2009.

ROCCHI, G.; VENTICINQUE, V. Calidad democrática, ciudadanía y participación en el ámbito local. *Espacio Abierto*, v. 19, n. 4, p. 601-620, 2010.

ROSS, F. C. Raw life and respectability poverty and everyday life in a postapartheid community. *Current Anthropology*, v. 56, n. 11, p. 97-107, 2015. <https://doi.org/10.1086/682078>

SAMPEDRO, Víctor F.; SÁNCHEZ, José Manuel; POLETTI, Monica. Ciudadanía y tecnopolítica electoral: ideales y límites burocráticos a la participación digital. *Co-herencia*, v. 10, n. 18, p. 105-136, 2013. Disponible en: <<http://www.scielo.org.co/pdf/cohe/v10n18/v10n18a04.pdf>>. Acceso en: 29 Mayo 2017.

SAUCEDO, I. A.; TARACENA, B. E. Habitar la calle: pasos hacia una ciudadanía a partir de este espacio. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, v. 9, n. 1, p. 269-285, 2011. Disponible en: <<http://www.redalyc.org/pdf/773/77320072014.pdf>>. Acceso en: 29 Mayo 2017.

SCHWITTAY, A. The marketization of poverty. *Current Anthropology*, v. 52, n. 3, p. 71-82, 2011. <https://doi.org/10.1086/656472>

SEN, A. *La idea de la justicia*. Colombia: Taurus, 2010.

SLOAM, James. "The outraged young": young europeans, civic engagement and the new media in a time of crisis. *Information, Communication & Society*, v. 17, n. 2, p. 217-231, 2014. <https://doi.org/10.1080/1369118X.2013.868019>

VILLA, Nora Helena; MONCADA, Yanet Maritza. Efectos de la implementación de un programa gubernamental orientado a la alfabetización digital en una comunidad rural. *Investigación & Desarrollo*, v. 19, n. 1, p. 26-41, 2011. Disponible en: <[http://www.scielo.org.co/scielo.php?pid=S0121-32612011000100005&script=sci\\_abstract&tlng=es](http://www.scielo.org.co/scielo.php?pid=S0121-32612011000100005&script=sci_abstract&tlng=es)>. Acceso en: 29 Mayo 2017.

VILLARROEL, Valentín. Gobierno electrónico y ciudadanía digital para zonas rurales de Perú. *Bit*, 194. 77-79, 2013. Disponible en: <<https://www.coit.es/archivo-bit/septiembre-2013/gobierno-electronico-y-ciudadania-digital-para-zonas-rurales-de-peru>>. Acceso en: 29 Mayo 2017.

WANDERLEY, F. Prácticas estatales y el ejercicio de la ciudadanía: encuentros de la población con la burocracia en Bolivia. *Iconos*, n. 34, p. 67-79, 2009. Disponible en: <<http://www.redalyc.org/html/509/50911338006/>> Acceso en: 29 Mayo 2017

XENOS, Michael; VROMEN, Ariadne; LOADER, Brian D. The great equalizer?: patterns of social media use and youth political engagement in three advanced democracies. *Information, Communication & Society*, v. 17, n. 2, p. 151-167, 2014. Disponible en: <<https://doi.org/10.1080/1369118X.2013.871318>>. Acceso en: 29 Mayo 2017.

## ***A brecha digital urbano-rural dos jovens na Colômbia: limitação ao exercício político***

### **RESUMO**

*Este capítulo apresenta uma reflexão crítica sobre a incidência da divisão digital rural urbana em jovens rurais na Colômbia, que leva a implicações políticas e sociais. Discute brevemente a desigualdade na Colômbia e suas conseqüências no exercício da cidadania e do empoderamento político, e depois faz uma análise sobre o papel das TICs, que, longe de promover a inclusão, constituem um mecanismo invisível de exclusão digital. Nesse cenário, as bibliotecas e seus profissionais estão entre os principais atores envolvidos na busca, implementação e promoção de estratégias de informação inclusivas para reduzir essa lacuna e assim contribuir para a construção de uma Colômbia mais equitativa e pacífica.*

**Palavras-chave:** Juventude rural. Divisão digital. Inclusão social e política. TIC. Bibliotecas digitais.

## ***The urban-rural digital divide of young people in Colombia: limitation to political exercise***

### **ABSTRACT**

*This chapter presents a critical reflection on the incidence of the urban rural digital divide in rural young people in Colombia, which lead to political and social implications. Briefly discusses inequity in Colombia and its consequences in the exercise of citizenship and political empowerment, and then make an analysis about the role of ICTs, which, far from promoting inclusion, constitute an invisible mechanism of digital exclusion. In such a scenario, libraries and their professionals are among the leading actors involved in the search, implementation and promotion of inclusive information strategies to reduce this gap and thus contribute to the construction of a more equitable and peaceful Colombia.*

**Keywords:** Rural youth. Digital divide. Social and political inclusion. ICT. Digital libraries.

# Lecciones aprendidas en proyectos de gestión de documentos electrónicos

Vicente González  
Johann Pirela Morillo  
Nelson Javier Pulido Daza

## RESUMEN

Con el desarrollo y la introducción de nuevas tecnologías en períodos cada vez más cortos, se ha evidenciado un aumento significativo en la producción de documentos electrónicos, lo que ha traído traumatismos y dificultades a las organizaciones al momento de gestionarlos, conservarlos y recuperarlos en entornos tecnológicos. Por ello, surge de la necesidad de aplicar nuevas metodologías y mejores prácticas para la gerencia de los proyectos de gestión documental electrónica en las diversas organizaciones, apoyando la transformación digital y la aplicación de principios archivísticos en el contexto empresarial. Para el desarrollo de la investigación, se asumió el enfoque cuantitativo, utilizando la técnica de la encuesta, sustentada en un cuestionario de recolección de datos, dirigido a 100 gerentes de proyectos, cursantes de Programas de Maestría en Gestión de Proyectos. Los resultados evidenciaron que las principales falencias de la gestión de proyectos sobre documentos electrónicos están en dos áreas estratégicas: los procesos de comunicación, que tienen que ver con la fluidez de la información y los canales para el intercambio de mensajes e ideas por parte de los actores participantes en la formulación y ejecución, y el levantamiento de requerimientos que las organizaciones presentan. Como conclusión se señala la necesidad de desarrollar acciones para lograr la efectividad en el diseño y ejecución de los proyectos de gestión de documentos electrónicos, considerando el conjunto de lecciones aprendidas.

**Palabras clave:** Proyectos. Gerencia. Formulación de proyectos. Gerencia de proyectos. Importancia de la gerencia de proyectos. Gestión de documentos electrónicos de archivo. Sistemas de gestión documental. Gobierno en línea. Evaluación de proyectos.



## INTRODUCCIÓN

En la primera mitad del siglo XX, los proyectos eran administrados con métodos y técnicas informales, basados en los gráficos Gantt – una representación gráfica del tiempo basada en barras, útil para controlar el trabajo y registrar el avance de tareas. En los años 50, se desarrollaron en Estados Unidos dos modelos matemáticos: PERT (Program Evaluation and Review Technique, técnica para evaluar y revisar programas, y CPM (Critical Path Method, método de ruta crítica, desarrollado por DuPont y Remington Rand, para manejar proyectos de mantenimiento de plantas). El PERT/CPM es, hasta la fecha, la base metodológica más utilizada por los gerentes de proyectos profesionales.

En 1969, se formó el PMI® (Project Management Institute, Instituto de Gerencia de Proyectos), bajo la premisa que cualquier proyecto, sin importar su naturaleza, podría utilizar las mismas bases metodológicas y herramientas. Es esta organización la más importante en América y establece estándares en esa materia.

La gran pregunta para los concedores y los que recién nacen en el tema es ¿Qué es un proyecto? Cuya respuesta articulada con el concepto de varios autores, entre ellos (PMBOK® Guide, 2013) “Un proyecto es un esfuerzo temporal que se lleva a cabo para crear un producto, servicio o resultado único. La naturaleza temporal de los proyectos implica que un proyecto tiene un principio y un final definidos”.

Estos eventos remiten a que los interesados en estructurar un proyecto de gestión electrónica de documentos, se pregunten por los elementos contentivos de un plan de proyecto; que deberá incluir: a). Fijación del objetivo del proyecto b). Definición de tareas y sus tiempos para su desarrollo c). Definición de hitos d). Definición de recursos e) Crear un cronograma base y f). Consolidar la ruta crítica del proyecto.

Fijación del objetivo del proyecto: En este punto se debe ser lo más específico como pueda, por cuanto se requiere del talento propio para la definición y la simplicidad al momento de redactar el objetivo del proyecto. La determinación y claridad en la definición permitirán así mismo lograr determinar el alcance o extensión, las personas o grupos afectados y el marco temporal.

**Definición de tareas y tiempo:** Luego de la fijación del o los objetivos, es de la naturaleza del proyecto adentrarse en delimitar los pasos necesarios y el tiempo de duración para conducir al éxito un proyecto; teniendo presente que las tareas se deben desarrollar en una secuencia determinada o paralela si es posible; pero en todos los casos debe recoger el espíritu del objetivo trazado por la naturaleza del proyecto, que constituirán el “PLAN DEL PROYECTO”.

**Definición de hitos:** Una de las dinámicas de la planeación es la definición o determinación de los HITOS que representan un evento o a actividad de significativa relevancia, puesto que un hito marca la finalización de un entregable o una fase por el momento en el que se causa.

**Definición de recursos:** Para la ejecución de un proyecto de gestión documental o documental electrónica, se requieren insumos que pueden incluir personas, equipos, software y cualquiera otro recurso necesario que permita el logro de los objetivos, teniendo en cuenta que su disponibilidad es básica en tiempo y espacio.

**Definición de ruta crítica:** Luego de realizadas las anteriores definiciones, llegamos al punto de reflejar las tareas críticas puesto que sin estas se retrasa el proyecto, por ello es imperativo determinar claramente cuáles son estas tareas y las actividades que pueden estar conexas.

**Cultura organizacional:** La cultura de las organizaciones a menudo tiene una influencia directa en el proyecto puesto que la estructura de las organizaciones limita la disponibilidad o los términos bajo los cuales los recursos se facilitan al proyecto.

**Preguntas de Chequeo:** Para determinar los posibles riesgos a los que se puede enfrentar el proyecto en su ejecución, sugerimos que se elabore un listado con los asuntos más relevantes y para ello enunciar preguntas como: ¿Quién o quiénes requieren ver los avances del proyecto? ¿Qué tipo de avances se deben informar? ¿El presupuesto debe tener algún tipo de seguimiento? ¿Cuáles son los entregables y los tiempos óptimos para su entrega? ¿Cuáles son las consecuencias de la no entrega o la entrega tardía? ¿Qué debería hacer si hay un entregable que definitivamente no podré lograr?

Este tipo de definiciones o conceptualizaciones muestra la importancia de la gerencia de proyectos de gestión de documentos electrónicos por cuanto se rescata que la Gerencia de proyectos es la aplicación de conocimiento, habilidades, herramientas y técnicas a las actividades propias de un proyecto para alcanzar los requerimientos del mismo. PMBOK® Guide.

## CONTEXTO DE LOS PROYECTOS DE GESTIÓN DOCUMENTAL ELECTRÓNICA EN LAS ORGANIZACIONES

La gestión documental electrónica cada vez se plantea como una necesidad de las organizaciones públicas y privadas, en un contexto signado por la maximización del acceso a la información y la transparencia en los procesos administrativos, que deben orientarse en todo momento a generar servicios y productos dirigidos a los ciudadanos, como base de la construcción de canales para la participación ciudadana. En este sentido, los proyectos de gestión documental electrónicos no solo deben enfocarse a partir de criterios técnicos y tecnológicos, sino que deben tenerse en cuenta aspectos jurídicos y organizacionales, relacionados a su vez con las competencias de los actores que van a participar en la formulación y ejecución de tales proyectos.

En este sentido, Bustelo (1997), plantea que dada la alta producción de documentos electrónicos, se debe prestar especial atención a la forma cómo se pueden gestionar, atendiendo a una serie de elementos, factores y variables, cuyo adecuado control y seguimiento, permitirán la efectividad e impacto de los proyectos de gestión documental electrónica en el ámbito organizacional.

De igual manera, es relevante el aporte que la administración, el derecho, la ingeniería y otras ciencias auxiliares de la archivística han realizado para consolidar una metodología que garantice el adecuado análisis, el diseño, la implementación y la evaluación de un proyecto de gestión documental electrónica, puesto que la información contenida en los diversos soportes reviste gran preocupación por su conservación y futura recuperación, ya sea por motivos administrativos, legales o culturales, dado que a corto, mediano y largo plazo muchos de estos documentos se transforman en evidencia de procedimientos, acciones o decisiones Gómez, (2003).

Además de la formulación de los proyectos, su adecuada gestión podría evitar la duplicidad de documentos y una cantidad significativa de versiones, formatos y contenedores de estos. Con ello, se estaría evitando la generación de dificultades al momento de almacenarlos en computadores, discos externos y servidores, por el gran volumen que allí ocupan. Por último, hay una continua inquietud sobre la recuperación de la información, en vista de que al existir volúmenes y versiones significativas de documentos almacenados sin un tratamiento archivístico idóneo y sin una descripción normalizada con un mínimo de metadatos, es posible que la información se pierda o no pueda recuperarse fácilmente frente a una eventual necesidad para un apoyo a la toma de decisiones.

Siguiendo estas ideas, Zapata (2011), enuncia que “si las entidades no implementan una adecuada gestión documental, exponen no solo el éxito de iniciativas electrónicas, sino la preservación de la memoria corporativa”. En este contexto, las entidades de carácter público no son ajenas a esta problemática; por tal motivo, el direccionamiento estratégico que tiene el Gobierno se enfoca a reducir la demanda de papel y contribuir con procesos administrativos más eficientes y transparentes, a través de flujos de información por sistemas electrónicos.

Para el Estado colombiano se hace necesario el adentrarse en el modelamiento de sistemas de gestión documental, el diseño y aplicación de políticas públicas de Transformación Digital - TD como lo aporta el Pulido (2011) en cuanto a la consideración los esfuerzos normativos y la aplicación de buenas prácticas internacionales que coadyuven a las entidades a resolver las problemáticas planteadas y apoyen la eficiencia administrativa para un buen gobierno. Algunas de estas normativas han sido aportadas por la Agenda de Conectividad, Gobierno en Línea, el plan Vive Digital, la iniciativa Cero Papel y la Circular 04 del 2010 de la Comisión Intersectorial de Políticas y Gestión de la Información (COINFO).

El aporte en la determinación y socialización de un modelo de buenas prácticas para el gerenciamiento de los proyectos de gestión documental o gestión documental electrónica contribuirá positivamente a la implantación de la política pública, así como a la transformación digital de las organizaciones, tema este que apoyará la consecución de mejores resultados

de las administraciones; es decir, impactará positivamente los costos administrativos y operativos, que permitirán el desarrollo de una cultura ciudadana eficiente y transparente como apoyo a la gestión y el acceso a la información en menor tiempo optimizando sus recursos.

## METODOLOGÍA

El estudio se basó en el enfoque cuantitativo de tipo descriptivo, utilizando la técnica de la encuesta, sustentada a su vez en un cuestionario de recolección de datos, dirigido a 100 gerentes de proyecto, los cuales son cursantes también de Programas de Maestría en Gestión de Proyectos. El cuestionario exploró aspectos técnicos de la formulación de los proyectos y también consideró las competencias blandas, necesarias por parte de los gerentes de proyectos, así como también las variables esenciales que se deben atender en los proyectos de gestión de documentos electrónicos, relacionadas con algunas de las áreas establecidas por la Guía PMBOK® del (Project Management Institute, Instituto de Gerencia de Proyectos) - PMI® y la Norma ISO 21500.

Las áreas utilizadas como base para la estructuración del instrumento fueron a). Comunicación en los proyectos de gestión documental electrónica. b.) Alcance del proyecto y su entendimiento c). Planeación de la calidad, d). Estimación del tiempo e). Costeo f). La gestión de los riesgos g). La documentación de proyecto h). El cómo se le entrego al cliente j). La presentación real al cliente.

El criterio fundamental para la selección de los informantes que participaron en el estudio fue que tuviesen experiencia comprobada de mínimo cinco años en gestión de proyectos y además tuviesen formación académica en este campo, para garantizar la identificación de las lecciones aprendidas, derivadas de cada una de las áreas exploradas. La información recolectada fue procesada con la herramienta SPSS, a partir de la cual se generaron gráficas de frecuencia, lo que permitió realizar el análisis estadístico descriptivo.

## ANÁLISIS DE LOS RESULTADOS

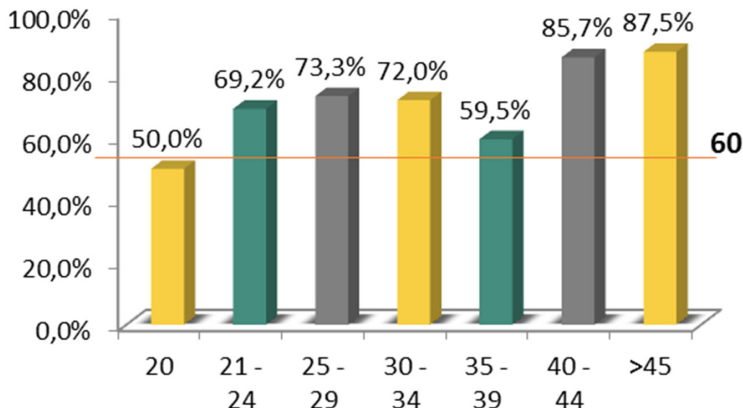
Los resultados que se presentan a continuación se analizan con base en la frecuencia e importancia que los informantes incluidos en el estudio le dieron a cada una de las áreas abordadas, de lo cual se generaron las lecciones que podían asumirse como aprendidas y luego convertirse en buenas prácticas en la gerencia de proyectos de gestión de documentos electrónicos a partir de conocimientos, habilidades, herramientas y/o técnicas y a las actividades propias de cada proyecto que busca alcanzar el cumplimiento de sus objetivos.

### LA COMUNICACIÓN EN LOS PROYECTOS DE GESTIÓN DOCUMENTAL ELECTRÓNICA

En cuanto a la comunicación, la mayor parte de los encuestados (87.5%), señalaron que esta área es de importancia estratégica, por cuanto constituye uno de los temas de mayor complejidad en los proyectos. Sin embargo, al considerar el promedio de los resultados en función de las edades de los informantes, se observa que tal promedio se ubica en el 60%, con lo cual se aprecia además que dependiendo de la edad, madurez y experiencia en la gestión de proyectos la percepción sobre la importancia de esta área varían con el paso de los años. De todos modos, la comunicación constituye un factor decisivo en el éxito de los proyectos. (Ver Gráfica 1)

La cantidad de involucrados en los proyectos de gestión documental electrónica debe ser proporcional a la cantidad de canales y mecanismos de intercambio de información y mensajes. La mayor parte de los proyectos de gestión documental necesita una fuerte carga operativa (manual), apoyada además en el análisis acerca de cómo se maneja la información y cómo se controla.

Gráfica 1 - Comunicación como factor decisivo de éxito de proyectos



Fuente: Elaboración propia de los autores

De lo anterior se deriva la necesidad de dejar claramente establecidas las formas cómo se van a gestionar los procesos de comunicación al interior del proyecto, así como los canales de comunicación con el cliente, lo cual implica a su vez determinar la frecuencia de las comunicaciones, los medios bajo los cuales se va a presentar y la responsabilidad del quien la elabora y de quien revisa. Es importante tener en cuenta que siempre hay personas que preguntan por la información directamente al cliente y esa información está desactualizada o no es lo que se informó sobre la ejecución del proyecto. (Ver gráfica 1).

El mayor problema en cualquier tipo de proyecto, y no solo en los proyectos de gestión documental electrónica, es la comunicación, en este ítem se encuentra el 90% de los problemas y por lo general la gerencia del proyecto no establece el tiempo suficiente para preparar reuniones o los responsables de las mismas, los proyectos de consultoría y de transformación digital en las organizaciones se desarrollan a través de reuniones y trabajos aplicados en grupo.

Una de las dinámicas de apropiación es la formación de equipos interdisciplinarios, de manera que estén presentes en las reuniones los que deben estar y los perfiles necesarios que puedan apoyar el logro de acuerdos, ustedes deben tener las competencias suficientes en su equipo para llevar a los técnicos que hablen con los técnicos, los funcionales con los funcionales, los ingenieros con los ingenieros, de esa forma ellos entienden y hacen acuerdos entre ambas partes, pero si usted dispone un funcional con un ingeniero los puntos de acuerdo no van hacer los mejores.

## **EL ALCANCE DEL PROYECTO Y SU ENTENDIMIENTO**

Los porcentajes relacionados con el alcance y entendimiento de los proyectos varió en función de la edad de los gerentes encuestados. En este sentido, se observó una variación que va desde el 50% hasta el 80%, considerado la variable edad; lo cual podría significar que la percepción sobre la importancia que tiene esta área, se refleja en consonancia con la madurez y experiencia en la formulación de los proyectos.

Los alcances de los proyectos de gestión documental electrónica podrían ser diversos, como por ejemplo: la organización física de un fondo documental, el levantamiento y análisis de procesos para un determinado tema, así como el diseño, personalización e implementación de una herramienta electrónica de gestión documental. Para clarificar el alcance y requerimientos de los proyectos, se debe tener clara la diferencia entre un requerimiento y un requisito. El requerimiento es una necesidad del negocio y el requisito corresponde a las condiciones para cumplir con un requerimiento.

Para la gerencia del proyecto, el requerimiento existe y normalmente debe lograr el resultado como lo solicita el cliente; pero lo que si debe precisar un acuerdo de voluntades que recoja las visiones del cliente y del ejecutor, para lograr homogenizar los dos equipos de trabajo hacia el mismo objetivo, considerando que el requisito es un aspecto puntual del requerimiento para alcanzar dicho objetivo.



## LA PLANEACIÓN DE LA CALIDAD

La planeación de la calidad fue otra de las áreas incluidas en el estudio con los gerentes de proyectos. Esta área obtuvo porcentajes que variaron entre el 30% y 60%, en función también de la edad de los gerentes. Con ello también se pone de relieve que la percepción con respecto a esta área varía de acuerdo con los grados de madurez y experiencia en los proyectos. El tema de la calidad es de los aspectos más complejos, porque ésta tiene un alto grado de subjetividad, de manera que se requiere usar unas métricas o unidades de medida que no generen conflicto y por ello se deben hacer acuerdos entre lo que entendemos por calidad y lo que el cliente entiende por lo mismo.

La calidad debe estar definida, entonces, por el cumplimiento de los estándares, adicionando lo que disciplinariamente contengan las normas y nos referimos a los decretos, resoluciones, acuerdos o cualquiera otro documento de política pública que fije los mínimos para la entrega del producto comprometido en el alcance de nuestro proyecto.

La calidad se basa en el cumplimiento del marco legislativo, regulatorio y normativo que aporte a la definición de cada una de las características de nuestra herramienta, teniendo en cuenta que el enfoque organizacional o socio cultural debe ser parte de la planeación para trabajar la transferencia de tecnología y el impacto organizacional al interior de la entidad y hacia el ciudadano.

Los criterios generales que se mencionan a continuación determinan la identificación de no conformidad de un producto dentro del proyecto y estarán acordes con los lineamientos generales del proyecto:

Tabla 1 – Identificación de producto no conforme

<b>Producto no conforme</b>	<p>Un producto no conforme se presenta cuando:</p> <ul style="list-style-type: none"> <li>• Tiene más de dos correcciones de un mismo punto tratado o analizado</li> <li>• Cuando se generan más de tres copias de un documento oficial</li> <li>• Desviaciones del alcance del proyecto</li> <li>• Retrasos o demoras en entrega de los informes</li> </ul>
<b>Manejo de producto no conforme</b>	<ul style="list-style-type: none"> <li>• Para evitar varias correcciones de un punto ya tratado, se debe velar por el diligenciamiento de formatos que garanticen una revisión juiciosa de los insumos entregados para el informe.</li> <li>• Aplica el manejo dado al punto anterior.</li> <li>• Dar seguimiento a los alcances detallados definidos para el proyecto.</li> <li>• Dar estricto cumplimiento al cronograma, para prever posibles retrasos</li> <li>• En caso de un producto no conforme se documentará en los formatos respectivos.</li> </ul>

Fuente: Elaboración de los autores (2016).

## ESTIMACIÓN DEL TIEMPO

El tiempo como variable crítica de los proyectos también se incluyó en el estudio dirigido a los gerentes. En este sentido, este aspecto obtuvo puntajes entre 40% y 60% de los informantes dependiendo de la edad. En una apreciación ligera de las cosas, siempre se piensa que un proyecto es similar a otro, o podría pensarse que es poco rentable destinar entre el 10% y el 20% del total del tiempo del proyecto para hacer una planeación detallada; Jeffrey (2014) señala que la mayoría de las veces, las preguntas que se hacen son: ¿Cuánto costó el proyecto? ¿Cuánto tiempo tomó su ejecución?

Además de las dos preguntas anteriores, la gestión de proyectos también debería considerar otras preguntas que permiten complementar el tema de los tiempos. Tales preguntas son: ¿Qué problemas tuvieron? ¿Qué recursos se necesitaron? ¿Cuántas jornadas tenían?

## **EL COSTEO**

Así como el tiempo se entiende como una de las áreas importantes de los proyectos, el costeo también es de suma importancia. El 65% de los gerentes incluidos en el estudio señaló que los costos se consideran de manera muy general en los proyectos, con lo cual se corre el riesgo de realizar una estimación del valor por encima o por debajo de la realidad, impactando el resultado final.

## **GESTIÓN DE LOS RIESGOS**

En cuanto a la gestión de los riesgos, dependiendo de sus edades, los informantes señalaron entre el 55% y 70% que este aspecto es poco considerado en los proyectos, siendo un área crítica que permite visualizar los posibles factores que generarían impactos positivos o negativos en la ejecución de los proyectos.

De acuerdo con los criterios señalados tanto en la Guía de PMBOK como en la Norma Iso 21500, la detección de los riesgos corresponden a un lista de posibles situaciones que tienen un impacto positivo o negativo para el normal desarrollo del proyecto, los cuales son posibles evidenciarlos cuando se está materializando, o cuando se está incumpliendo el alcance, el tiempo o superamos el presupuesto.

Es importante identificar si es un riesgo a favor o en contra mirar su ocurrencia y el tipo de impacto, que puede estar relacionado con aspectos técnicos, jurídicos y/o económicos que tiene se evidencia por medio de su materialización, la sugerencia es elaborar una estructura detallada de riesgos, construir una matriz de calor, clasificarlos por grupos, los técnicos, lo externos, los organizacionales y los de dirección del proyecto, evaluar su probabilidad de ocurrencia y su impacto en el proyecto.

## **LA DOCUMENTACIÓN DE PROYECTO**

Esta actividad, aunque de gran importancia para los proyectos, no se toma en cuenta; ello se evidencia en un 85% de respuestas positivas aportadas por los gerentes, en relación con la ausencia de esta área en el cierre de los proyectos. Esta área debe ser planificada y asignada a una persona que responda por ella desde el inicio, de esta forma se define una estructura física y electrónica de los documentos, requerimientos y entregables. La documentación es un proceso

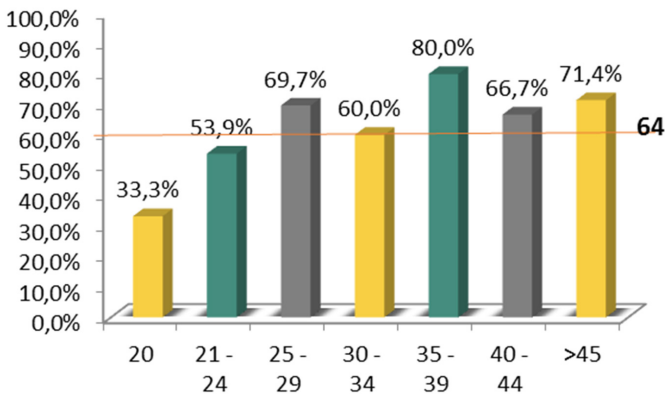
que permite sistematizar las historias de éxito, los nudos críticos y problemas confrontados en el proyecto, así como las formas y metodologías aplicadas para resolverlas y superarlas. Por ello, la documentación se concibe como un proceso vital de la gestión de los proyectos.

## ¿CÓMO SE ENTREGA AL CLIENTE?

La relación entre cliente y la dirección o gerencia del proyecto, debe mantenerse en el más alto nivel de confianza y con el mayor grado de asertividad en la comunicación porque si no manejamos estos elementos con el cuidado que ameritan, al final del proyecto la relación puede estar desgastada con el día a día, si el proyecto no ha sido exitoso, sino ha cumplido con las expectativas; las partes sienten que la relación debe darse por terminada y el ejecutor siente que debe marcharse o que su presencia es una molestia, entonces lo que se entrega al final, no es el mejor resultado.

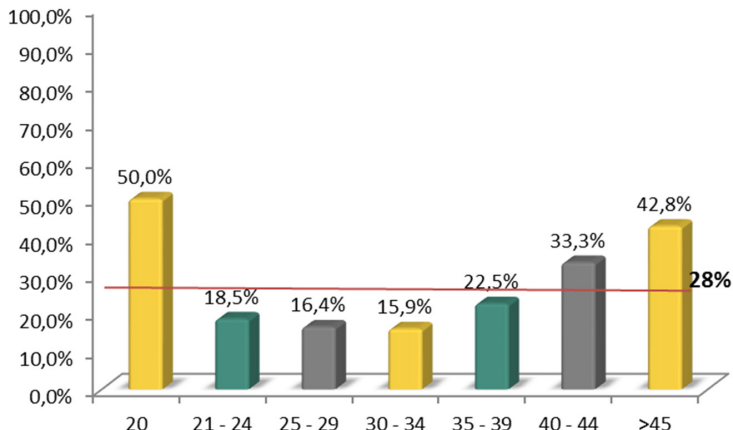
Para ello, las competencias de liderazgo y negociación son vitales en las interacciones que se establezcan con los clientes. En este sentido, estas dos competencias obtuvieron promedios entre 64% (la de liderazgo) y 28% (la de negociación), con lo cual se pone de relieve que los gerentes encuestados valoran más importante el liderazgo que la negociación, siendo esto última vital también para las relaciones que se establecen entre los gerentes de proyectos y sus clientes finales.

Gráfica 2 – Competencia de liderazgo en los proyectos



Fuente: Elaboración propia de los autores.

Gráfica 3 – Competencia de negociación en los proyectos



Fuente: Elaboración propia de los autores.

## LA PRESENTACIÓN REAL AL CLIENTE

Al finalizar un proyecto recogemos un cúmulo de experiencias y conocimientos que deben sistematizarse en los informes y actas del proyecto puesto que las lecciones aprendidas son tomadas con base en la decisión de las últimas semanas, ósea cuando ya se ha entregado cada producto y tenemos que redactar y consolidar la documentación, los manuales, informes de cierre, actas y no perder la historia; pareciera que es normal que los proyectos no se finalicen correctamente, los abandonen y se pierdan las experiencias vividas en ellos, en el desconocimiento del capital de buenas prácticas que conforman para los futuros proyectos.

## LECCIONES APRENDIDAS PARA LOS PROYECTOS DE GESTIÓN DOCUMENTAL ELECTRÓNICA

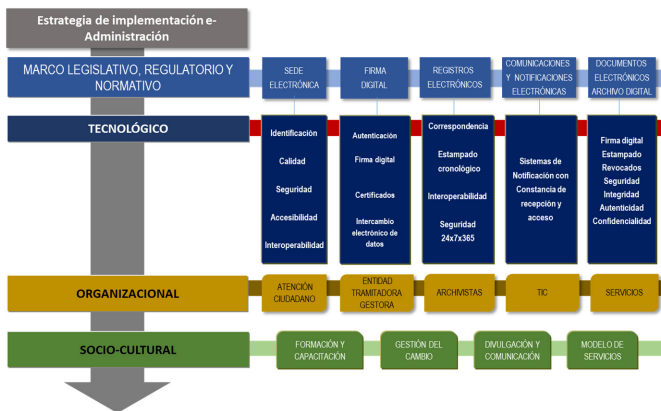
A partir de los resultados analizados, es posible proponer algunas lecciones aprendidas en la formulación y ejecución de proyectos para la gestión documental electrónica en las entidades públicas y privadas. En relación con la comunicación, la lección aprendida tiene que ver con la necesidad de

establecer planes que contemplen la forma cómo se van a generar los flujos de información y los mecanismos de intercambio de mensajes para asegurar procesos de comunicación efectiva y eficiente, en aras además de garantizar la homogeneización de los datos y de los ámbitos de responsabilidades de los actores involucrados en los proyectos.

En cuanto al alcance y entendimiento de los proyectos, es importante precisar hasta dónde se pretende llegar, en términos de tener claros los objetivos, productos entregables, responsabilidades, lo que está incluido en el proyecto, para de esta forma puntualizar el ser, deber ser y el horizonte funcional de las iniciativas para la gestión documental electrónica.

Sobre la planeación de la calidad, la lección aprendida apunta a la necesidad de incorporar una visión integral en la formulación y ejecución de los proyectos, visión que se construye sobre la base de considerar diversos elementos y no solo los técnicos y tecnológicos, sino los jurídicos, normativos, organizacionales y socio-culturales, los cuales se aprecian en la figura 1.

Figura 1 – Visión integral con perspectiva de calidad de los proyectos de gestión documental electrónica



Como parte de la estrategia de implementación de los proyectos de gestión documental electrónica, se deben considerar el marco legislativo, regulatorio y normativo, el componente tecnológico, del cual debe generarse los procesos de

apoyo a los procedimientos típicamente archivísticos, orientados a garantizar la maximización del acceso a la información. Igualmente, en el ámbito organizacional se deben tener muy presentes los propósitos de los servicios y los productos derivados de la gestión documental electrónica y dentro de los aspectos socio-culturales se deben considerar la formación, gestión del cambio, divulgación y generación de modelos de gestión de servicios.

Con respecto al tiempo, la lección aprendida que proponemos es profundizar con preguntas mucho más focalizadas, el tiempo real que se requiere para la ejecución de los proyectos. La recomendación es construir un esquema detallado de trabajo con los tiempos y los diccionarios de datos, donde se pueden establecer las condiciones que a nivel de actividad se deben desarrollar, eso permite unos mejores resultados que se ven en la consolidación de los objetivos de cada etapa.

Una lección aprendida para la estimación de los costos en un proyecto de gestión electrónica de documentos es estructurar dicha estimación con base en actividades o por paquetes de actividades, Miranda (2006) de manera que pueda involucrar variables de tiempo semanales, mensuales, trimestrales o semestrales y a éstas adicionalmente hacerles una reserva o contingencia; luego de esto puede adicionársele otra variable que puede ser el personal, dependiendo del tipo de necesidad y del perfil asociado.

La lección aprendida relacionada con la gestión de riesgos propone tomar en cuenta los siguientes elementos:

- » Planeación del Riesgo (Hace parte del desarrollo del proyecto)
  - √ Conocimiento general del Proyecto
  - √ Diseño del Plan de riesgos
    - Definir objetivos
    - Diseñar cronograma
    - Definir metodología a seguir
- » Valoración del Riesgo
  - √ Identificación del Riesgo
    - Qué, ¿Cómo, Por qué?

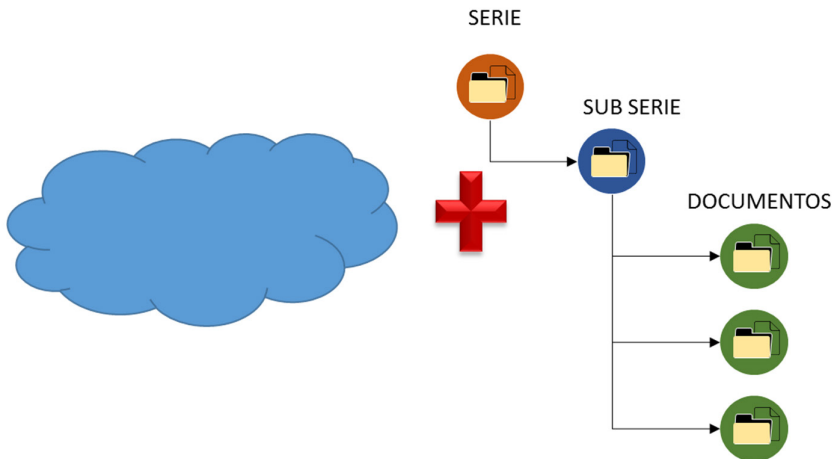
- Técnicas a desarrollar (Cuestionarios, Lluvia de ideas)
- Tipo: Internos/Externos, Financieros, del proyecto, Técnicos, Etc.
- √ Análisis del Riesgo
  - Probabilidad (Alto, Medio, Bajo)
  - Impacto (Alto, Medio, Bajo)
  - Calificación
- √ Determinación del Nivel del Riesgo
  - Evaluación
  - Semáforo
- » Manejo del Riesgo
  - √ Elaboración del Mapa de Riesgos
  - √ Implementación de Acciones
    - Evitar el Riesgo (Rediseño, Controles)
    - Reducir el Riesgo (Optimizar procesos, Controles)
    - Dispersar y Atomizar (Distribución distintos lugares)
    - Transferir el Riesgo (Pólizas, Subcontratar)
    - Asumir el Riesgo (Riesgos Bajos, Riesgos residuales, Riesgos Secundarios)
- » Monitoreo del Riesgo
  - √ Elaboración del plan de seguimiento
    - Plantilla de Control
  - √ Ejecución del seguimiento
    - Plantilla de Control
  - √ Presentación de resultados y propuestas
    - Informes de seguimiento

En cuanto a la documentación de los proyectos de gestión de documentos electrónicos se evidencia como lección aprendida la necesidad de recurrir a los recursos disponibles en la nube para sistematizar los nudos críticos



encontrados, así como otro tipo de problemas con sus correspondientes soluciones, con objeto de contar con un banco de historias de éxito en la formulación y ejecución de los proyectos. El modelo para la documentación de los procesos y la forma cómo se les dio respuesta a los problemas se enuncia en la Figura . 2

Figura 2 – Estructura de documentación en la nube



Fuente: Elaboración de los autores (2016).

Hay herramientas libres que apoyan la construcción de la documentación del proyecto de manera electrónica y en la nube que permite controlar los accesos y las modificaciones, el versionamiento de los documentos, hacer la trazabilidad de cualquier documento y regresar a rescatar una versión de meses atrás, sin perder su rastro, recuperar esos documentos y poder hacer los backup de correos que en los proyectos es gigantesca.

En cuanto a las relaciones entre los gerentes y clientes se deben son buenos aliados a la hora de cerrar el proyecto y entregar los resultados; es ideal que se recojan las lecciones aprendidas y que entreguemos los productos como se enunciaron en los requerimientos en muy buenos términos; para ello, es de buen recibo hacer un acta que recoja cada uno de los compromisos

y acuerdos que definían el proyecto, sin desconocer que el conflicto existe en los proyectos, lo que tenemos que determinar claramente es la causa y buscar una solución.

## CONCLUSIONES

Administrar las expectativas de los involucrados en un proyecto puede ser muy difícil ya que ellos a menudo tienen objetivos diferentes que pueden estar en conflicto y encontrar una solución apropiada a tales diferencias puede ser uno de los mayores retos de la gerencia de proyecto. Las competencias que deben reunir los líderes y los equipos, para gerenciar proyectos, se deben enmarcar dentro del contexto de cada organización y son únicas en cuanto a cultura, valores, relaciones, formas de trabajo en equipo y madurez corporativa.

En el ámbito de la gerencia de proyectos de gestión documental electrónica es importante trabajar con herramientas como las matrices de calor, para identificar visualmente los objetivos y los riesgos a los que nos enfrentamos; por ello deben tener cuidado con lo que este en rojo y lo que este en amarillo, con este tipo de posiciones el nivel de atención y cuidado es mayor, ya que la estrategia es generar las herramientas de control necesarias para una gestión cuidadosa.

Es obligatorio identificar y caracterizar a los usuarios, por lo que nos enfrentamos a riesgos innecesarios para atender cual es el interés e impacto de cada uno para con el proyecto y gestionar sus intereses. Una de las mejores prácticas en la gerencia de proyectos es la de generar un modelo de comunicación que fije claramente el esquema de comunicación de la información al interior del proyecto y al cliente; así como definir la frecuencia, el medio en que se va a presentar, la responsabilidad de quién elabora y de quien revisa, no podemos ser juez y parte para las entregas

La gerencia del proyecto debe lograr el resultado como lo solicita el cliente; con un acuerdo de voluntades que recoja las visiones del cliente y del ejecutor, de manera que se logren homogenizar los dos equipos de trabajo hacia el mismo objetivo. Dentro de las dinámicas de apropiación

es la formación de equipos interdisciplinares, de tal forma que estén presentes en las reuniones los que deben estar y los perfiles necesarios que puedan apoyar el logro de acuerdos.

La calidad en los proyectos está definida por la aplicación de estándares, adicionando lo que disciplinarmente contengan las normas y nos referimos a los decretos, resoluciones, acuerdos o cualquiera otro documento de política pública que fije los mínimos para la entrega del producto.

Para la estimación de los costos en un proyecto de gestión electrónica de documentos se recomienda realizarlo a nivel de actividades o por paquetes de actividades; es decir planeaciones de cuarenta (40) o máximo (80) horas de esfuerzo, de manera que pueda tener una planeación detallada de los costos asociados con cada actividad y sus recursos relacionados.



**REFERÊNCIAS**

- BUSTELO, C. Los sistemas de gestión electrónica de la documentación y la teoría del ciclo vital de los documentos en las organizaciones. *Scire*, v. 3, n. 2, p. 45-53, 2010.
- \_\_\_\_\_. *Serie ISO 30300: sistema de gestión para los documentos*. 2014. Disponible en: <<http://www.sedic.es/DT-n2-SEDIC-ISO30300.pdf>>. Acceso en: 13 jul. 2017.
- COLOMBIA. Ministerio de Tecnologías de la Información y las Comunicaciones. *Decreto 2609: disposiciones en materia de gestión documental para entidades públicas*. Colombia, 2012. Disponible en: <[http://www.mintic.gov.co/portal/604/articles-3528\\_documento.pdf](http://www.mintic.gov.co/portal/604/articles-3528_documento.pdf)>. Acceso en: 13 jul. 2017.
- COMISIÓN INTERSECTORIAL DE POLÍTICAS Y DE GESTIÓN DE LA INFORMACIÓN PARA LA ADMINISTRACIÓN PÚBLICA - COINFO. *Preguntas*. 2010. Disponible en: <<http://admonpublica.org/preguntas-coinfo/>>. Acceso en: 13 jul. 2017.
- GÓMEZ, D.D. et al. La gestión de documentos electrónicos: requerimientos funcionales. *El Profesional de la Información*, v. 12, n. 2, 2003. Disponible en: <<http://www.elprofesional-delainformacion.com/contenidos/2003/marzo/1.pdf>>. Acceso en: 13 jul. 2017.
- INSTITUTO COLOMBIANO DE NORMAS TÉCNICAS Y CERTIFICACIÓN – ICONTEC. *Guía técnica colombiana GTC-ISO 21500 2013-12-11: directrices para la dirección y gestión de proyectos*. Colombia, 2013.
- MIRANDA, J. J. *El desafío de la gerencia de proyectos*. Bogotá: MM Editores, 2006.
- \_\_\_\_\_. *Gestión de proyectos: identificación, formulación y evaluación*. 5. ed. Bogotá: MM Editores, 2009.
- PINTO, J.K. *Gerencia de proyectos: cómo lograr la ventaja competitiva*. Colombia: [s.n.], 2015.
- PROJECT MANAGEMENT INSTITUTE. *A guide to the Project Management Body of Knowledge*. PMBOK Guide 5<sup>o</sup> edition. Pennsylvania, 2012.
- PULIDO, N.J. *La gestión de documentos: modelos y características*. Cáceres, España: Fundación Ciencias de la Documentación, 2011.
- ZAPATA, C.A. *Gestión documental & gobierno electrónico*. Bogotá: Imprenta Nacional, 2011.

## ***Lições aprendidas no projeto de gestão de documentos eletrônicos***

### **RESUMO**

*Com o desenvolvimento e introdução de novas tecnologias em períodos cada vez mais curtos, percebe-se um aumento significativo na produção de documentos eletrônicos, que tem gerado trauma e dificuldades às organizações ao gerenciar, conservar e recuperá-los em ambientes tecnológicos. Assim, surge a necessidade de implementar novas metodologias e melhores práticas para gerenciar projetos de gestão de documentos eletrônicos nas diversas organizações, apoiando a transformação digital e a aplicação de princípios de arquivamento no contexto empresarial. Para desenvolver a pesquisa, a abordagem quantitativa foi assumida, por meio da técnica de enquete com base em questionário para coleta de dados, dirigido a 100 gerentes de projeto cursando o Programa de Mestrado em Gestão de Projetos. Os resultados mostraram que as principais deficiências da gestão de projetos de documentos eletrônicos ocorrem em duas áreas estratégicas: processos de comunicação, relacionados com o fluxo de informações e canais para troca de mensagens e ideias por parte dos atores envolvidos na formulação e execução, e a forma como é realizado o levantamento de requisitos por parte das organizações. Em conclusão, aponta-se a necessidade de desenvolver ações para alcançar a eficácia na concepção e execução de projetos de gerenciamento eletrônico de documentos, considerando o conjunto de lições aprendidas.*

**Palavras-chave:** *Projetos. Gestão. Desenvolvimento de projetos. Gerenciamento de projeto. Importância do gerenciamento de projetos. Gestão de registros eletrônicos. Sistemas de gestão de documentos. E-governo. Avaliação de projeto.*

## ***Lessons learned in electronic document management projects***

### **ABSTRACT**

*With the development and introduction of new technologies in shorter periods, there has been a significant increase in the production of electronic documents, which has caused traumas and difficulties to organizations when managing, preserving and recovering them within technological environments. Thus arises the need to apply new methodologies and best practices for managing electronic document management projects in a variety of organizations, to support the digital transformation and the application of archival principles in the business context. For the development of the research, the quantitative approach was adopted, using the survey technique, based on a data collection questionnaire, aimed at 100 project managers of the Master's Program in Project Management. The results showed that the main shortcomings of project management on electronic documents are in two strategic areas: communication processes, related to the fluidity of information and channels for the exchange of messages and ideas by the actors involved in formulation and implementation, and the way requirements gathering is performed within organizations. As a conclusion, the need to develop actions to achieve effectiveness in the design and execution of electronic document management projects is stressed, considering the set of lessons learned.*

**Keywords:** *Projects. Management. Project development. Project management. Importance of project management. Electronic records management. Document management systems. E-government. Project evaluation.*

# Valor probatorio de los documentos electrónicos: visiones desde Colombia y Venezuela

**Johann Pirela Morillo**  
**María Yaneth Álvarez**  
**Nelson Javier Pulido Daza**

### RESUMEN

Partiendo del concepto del documento electrónico, entendido como la representación del pensamiento, materializada en soportes magnéticos que garantizan el acceso inmediato a datos relevantes que a su vez también se pueden utilizar como medio de prueba, se realiza un análisis comparativo de la legislación y normatividad colombiana y venezolana, relacionada con este tipo de documentos. La metodología consistió en una investigación documental, que integró la comparación de principios conceptuales y jurídicos con algunas buenas prácticas que se aplican en Venezuela y Colombia en materia de gestión electrónica de documentos. Como resultados se evidencia que los documentos electrónicos están reconocidos como medios de prueba en las legislaciones de ambos países, lo cual genera un impacto significativo en el acceso a la información, la transparencia y los procesos de participación de los ciudadanos en la gestión pública y privada. Con ello se plantea como conclusión, la consideración de los documentos electrónicos, en cuanto evidencias probatorias, con posibilidades de uso en procesos legales, lo cual plantea además los esfuerzos realizados en los dos países, cuyas legislaciones se compararon, con miras a poner de relieve el alcance y las implicaciones de este tipo de documentos y sus implicaciones en la gestión de todo tipo de entidades. Los principios que deben guiar la aplicación de los documentos electrónicos como medios de prueba son: la inalterabilidad, conservación del mensaje, legibilidad, fiabilidad de los sistemas que lo producen y admisibilidad.

**Palabras clave:** Documentos electrónicos, Legislación venezolana, Legislación colombiana, Valor probatorio.

## INTRODUCCIÓN

Los cambios introducidos por las tecnologías digitales, así como también las profundas transformaciones suscitadas en las organizaciones de todo tipo, han estado generando importantes mutaciones en la producción, circulación y transferencia de la información: sus medios, fuentes y soportes. En este escenario, los documentos electrónicos se erigen como medios de prueba y ello ha obligado a los países a modificar las legislaciones que le dan marco normativo y jurídico, con lo cual se han estado configurando nuevos referentes acerca de la naturaleza, alcance e implicaciones de los documentos electrónicos dentro en el circuito de actividades productivas y de gestión académica, social, científica y tecnológica.

Se plantea, entonces, la conceptualización de documento electrónico y su valor probatorio, haciendo énfasis en las tendencias regulatorias que le dan validez legal, teniendo presente su carácter probatorio y algunos elementos de evaluación que como documento se pueden establecer desde el marco legal colombiano y venezolano. Igualmente, se destacan las incidencias de las tecnologías en el derecho privado, para concluir con los asuntos más relevantes en la relación de las tecnologías de la información con el derecho público.

En ese orden de ideas, el término valoración o evaluación documental se ha asociado con los de selección, descarte y expurgo, en distintos momentos de la evolución de la Archivística como disciplina y según las distintas tradiciones, siempre ha tenido un papel importante, las formulaciones desarrolladas por Theodoro Schellenberg en 1956 y que han sido recogidas y reinterpretadas en distintos contextos. Es así como Schellenberg es uno de los teóricos más influyentes en cuanto a la valoración que recoge el pensamiento de Sir Hilary Jenkinson, de los archivos como evidencia imparcial y de Karl Otto Muller, para quien lo que no merece preservarse, nunca debería ser admitido en un archivo. (RIVAS FERNÁNDEZ, 2011, 82).

Las legislaciones colombiana y venezolana tienen contemplado el documento electrónico como medio de prueba, para cuyo uso en procesos comerciales y jurídicos, se deben tener presentes un conjunto de criterios, entre los cuales figuran: la inalterabilidad, conservación del mensaje, legibilidad, fiabilidad de los sistemas que lo producen y admisibilidad.



Elementos que tendrán que ser considerados como criterios orientadores para el diseño y estructuración de sistemas que faciliten la producción y gestión de documentos electrónicos como medios de prueba.

## NATURALEZA Y CARACTERÍSTICAS DE LOS DOCUMENTOS ELECTRÓNICOS

Antes de conceptualizar el concepto de documento electrónico es pertinente presentar la definición de documento, el cual según Núñez Contreras (1993) alude a un objeto corporal, generado de la actividad humana, que puede ser utilizado como fuente de conocimiento o medio que permite demostrar algo. A partir de esta noción es posible identificar diversas significaciones del concepto de documento, con sus diferentes matices archivísticos, documentales e informacionales. En este sentido, los documentos son bienes muebles de carácter material, señalando entonces que la idea de documento es una representación del pensamiento, es así como todo objeto producido, directa o indirectamente, por la actividad del hombre y que, representa una cosa, un hecho o contiene una manifestación del pensamiento o de la voluntad de quien lo produce.

Otra perspectiva que resulta pertinente incluir en el análisis terminológico y conceptual de la noción de documento es la que aporta el International Council on Archives, según la cual los documentos se definen en términos de “unidades indivisibles de información, constituida por un mensaje fijado a un medio de manera sintácticamente estable. Un documento tiene por tanto siempre una forma fija y un contenido estable” (ICA, 2015).

En el contexto colombiano, el Código de Procedimiento Civil (Decreto 1400, 1970) establece en el artículo 251 que “son documentos los escritos, impresos, planos, dibujos, cuadros, fotografías, cintas cinematográficas, discos, grabaciones magnetofónicas, radiografías, talones, contraseñas, cupones, etiquetas, sellos y, en general, todo objeto mueble que tenga carácter representativo o declarativo y las inscripciones en lápidas, monumentos, edificios o similares”.

Ahora bien, aproximándonos a la conceptualización de los documentos electrónicos, se entienden como tales a los contenidos en soportes electrónicos o máquinas informáticas y cuyo contenido pueden ser “magnitudes físicas que representan en forma codificada unas declaraciones o representaciones y que son susceptibles de registro, proceso y transmisión, lo que traduce que este documento requiere de medios o instrumentos electrónicos como medio de vida y reproducción”. (PARRA QUIJANO, 2006).

Según Giménez-Chornet (2013), en cuanto a la gestión de este tipo de documentos, en la actualidad se cuenta con la Norma 16175 del 2012, la cual señala un conjunto de elementos funcionales y técnicos que se deben tener presentes en la creación de los documentos en su contexto, los metadatos durante la configuración del sistema, así como los soportes a la importación, exportación e interoperabilidad, como importación de la información de las pistas de autoría asociadas a documentos, o la exportación de documentos dentro del sistema de la organización o a una institución archivística que tenga un programa de conservación de documentos a largo plazo.

Los documentos electrónicos entonces han estado planteando nuevas configuraciones de las teorías archivísticas, uno de cuyos conceptos es el de preservación digital, el cual, en palabras de Alberch (2015), constituye un reto de gran relieve en la medida que transforma absolutamente los paradigmas de conservación de documentos en soporte papel y en un entorno legal y tecnológico en constante cambio. El experto español alude a la necesidad de garantizar una custodia a largo plazo de documentos auténticos, íntegros, fiables y accesibles, todo lo cual además requiere de un abordaje interdisciplinario entre los profesionales de archivística, (gestores de documentos), expertos en organización, juristas y administradores de sistemas.

La gestión de documentos electrónicos, entonces, supone la necesidad de generar nuevas conceptualizaciones y orientaciones para la práctica archivística, la cual debe generar diálogos con otras disciplinas, con objeto de ampliar la mirada y la comprensión de la naturaleza e impacto que tienen los nuevos soportes documentales en los escenarios organizacionales.

## **VALOR PROBATORIO DE LOS DOCUMENTOS ELECTRÓNICOS EN LA LEGISLACIÓN COLOMBIANA.**

Con la ley 527 de 1999, se constituye en Colombia, el régimen general de acceso y uso de los mensajes de datos, estableciendo una serie de directrices de la equivalencia funcional, del envío, recepción de datos y mensajes por medio electrónico. A partir de lo cual, la Superintendencia Financiera de Colombia, emitió el concepto No. 2007062815-001 (octubre 22 de 2007), en el que fija una postura frente al documento electrónico y se establece la posibilidad de archivar en medios magnéticos por ejemplo, los soportes de los cierres contables.

Es así como, en el artículo 7° de la Ley 527 de 1999 se establece la firma manuscrita como método que permite identificar el iniciador del mensaje de datos y su aprobación en el contenido del mensaje, siempre que dicho método sea tanto confiable como apropiado para el propósito por el cual el mensaje fue generado o comunicado. Mientras que el artículo 12, se establece como requisito el cumplimiento de las siguientes condiciones:

1. Que la información que contengan sea accesible para su posterior consulta.
2. Que el mensaje de datos o el documento sea conservado en el formato en que haya generado, enviado o recibido o en algún formato que permita demostrar que con exactitud la información generada, enviada o recibida, y
3. Que se conserve, de haber alguna, toda información que permita determinar el origen, el destino del mensaje, la fecha y la hora de envío o recibo del mensaje o documento.

De otra parte, en el Artículo 2 de la Ley 527 de 1999, establece el valor probatorio del documento electrónico, confirmado por medio del concepto 04044586 (junio, 2004) emitido por la Superintendencia de Industria y Comercio, en cuanto a los equivalentes funcionales establecidos por la ley, teniendo en cuenta para ello los requisitos de fiabilidad, inalterabilidad y rastreabilidad. (...) los documentos electrónicos están en capacidad de brindar similares niveles de seguridad que el papel y, en la mayoría de los casos, un

mayor grado de confiabilidad y rapidez, especialmente con respecto a la identificación del origen y el contenido de los datos, siempre que se cumplan los requisitos técnicos y jurídicos plasmados en la ley.

Siguiendo esta idea, la Corte Constitucional, en Sentencia C-662 de 2000, señaló que los mensajes de datos recibirían el mismo tratamiento que los documentos en soporte papel, estableciendo para ello características esenciales del mensaje de datos como prueba de la existencia y naturaleza de la voluntad de las partes, también se admite el almacenamiento e inalterabilidad en el tiempo, con lo cual se podrá revisar y auditar con fines contables o reglamentarios, según el caso.

En concordancia con el Artículo 64 del Código de Comercio, cualquier persona natural o jurídica, se encuentra facultada para implementar el sistema de archivo de información que considere más conveniente para su necesidad, condición de respeto de la Ley 527 de 1999, en particular las reglas sobre equivalencia funcional. Este sistema debería garantizar, de forma adecuada, la limitación en el acceso de la información, (excepto la reserva bancaria sobre las inversiones personales). Mientras que el concepto de la Corte Constitucional, deja claro que el mensaje de datos debe recibir el mismo tratamiento de los documentos establecidos en soporte papel, por cuanto contiene la misma eficacia jurídica del mensaje de datos recogiendo las características enunciadas en el Artículo 251 del Código de Procedimiento Civil, en lo referente al carácter representativo o declarativo del documento. (Código de Procedimiento Civil, 2015)

La mencionada Ley 527 de 1999, contiene una presunción de legalidad, al determinar la autenticidad y la integridad del documento electrónico cuando este contenga una firma digital, que señala específicamente que son “atributos jurídicos de una firma digital. Cuando una firma digital haya sido fijada en un mensaje de datos se presume que el suscriptor de aquella tenía la intención de acreditar ese mensaje de datos y de ser vinculado con el contenido del mismo”. (Ley 527, 1999, art. 28).

Por lo cual es posible señalar que la norma contiene unos condicionantes que validan la fuerza probatoria del documento y lo equiparan con el documento cartular o impreso bajo los siguientes términos:

- » El uso de una firma digital tendrá la misma fuerza y efectos que el uso de una firma manuscrita, incorporando para ello una serie de atributos:
  - √ Es única a la persona que la usa.
  - √ Es susceptible de ser verificada.
  - √ Está bajo el control exclusivo de la persona que la usa.
  - √ Está ligada a la información o mensaje, de tal manera, que, si éstos son cambiados, la firma digital es invalidada.
  - √ Está conforme a las reglamentaciones adoptadas por el Gobierno Nacional.

Como evidencia son las relaciones entre los ciudadanos y la administración pública son reguladas por el procedimiento administrativo, en el contexto del gobierno electrónico y el concepto de administración electrónica, como conjunto de procesos y sistemas ofrecidos a partir del uso de las tecnologías de la información y la comunicaciones, utilizando para ello los procesos democráticos, el refuerzo de las políticas públicas, el uso del Internet y la participación más activa de los ciudadanos.

Para el año 2000, se establece el gobierno en línea en Colombia, con el objeto de mejorar la eficiencia y transparencia del estado, que en un primer momento establecía la promoción de los servicios de información y del mismo estado a través del uso del Internet. Toda esta regulación se fundamentó a partir de la implementación de la ley 527 de 1999 y con los elementos mencionados anteriormente.

A partir de lo cual, es posible señalar que el documento electrónico es una evolución de los documentos en formato impreso, debido a que son medios de prueba que poseen unas características: tiene un continente (cuerpo): soporte material (cintas, disquetes, memorias), tiene un contenido, un “mensaje” “electrónico”, puede ser de texto, voz, datos, video, imágenes, está escrito bajo un código determinado, tiene grafía y puede ser atribuido (autenticidad) a una persona determinada.

Lo cual exige particularidades específicas de:

2. **La autenticidad** que se traduce en la identificación de la persona de quien emana;
- c. Que sea generado y conservado en condiciones que permitan garantizar **su integridad**, que puede significar el que no haya sido adulterado.

La evidencia probatoria del documento electrónico, se establece en dos grandes funciones o finalidades de carácter probatorio, que en principio le corresponde la que persigue con su existencia propiamente dicha y es de carácter sustancial (solemne) y la segunda corresponde a la prueba propiamente dicha (cuando prueba un hecho o un acto jurídico) es un medio de prueba, pero incontables veces puede ser objeto de prueba, cuando es necesaria la verificación de la autenticidad. Parra- Quijano (2006).

Por el mismo contexto de los avances, es necesario certificar la seguridad, teniendo como marco la legislación colombiana, es necesario hacer un análisis del documento electrónico como medio de prueba de la siguiente forma:

- Frente a la admisibilidad y fuerza probatoria de los mensajes de datos, la Ley 527 de 1999, en su Artículo 10 prescribe que “los mensajes de datos serán admisibles como medios de prueba y su fuerza probatoria es la otorgada en las disposiciones del Capítulo VIII del Título XIII, Sección Tercera, Libro Segundo del Código de Procedimiento Civil”. (Código de Procedimiento Civil, 2015);
- Respecto a la prueba documental, el Capítulo VIII, del Título XIII, Sección Tercera del Libro Segundo del Código de Procedimiento Civil, manifiesta que se entienden incorporados a la legislación en el aspecto referenciado y se apreciarán teniendo en cuenta las disposiciones, plasmadas en los artículos del 251 al 301 del Código de Procedimiento Civil. (Código de Procedimiento Civil, 2015).

Para la valoración de la fuerza probatoria de los mensajes de datos que se refiere la Ley 527 de 1999 en su Artículo 11, se establecen criterios para valorar probatoriamente los mensajes de datos y para ello se tendrán en cuenta las Reglas de la Sana Crítica y demás criterios reconocidos legalmente para la apreciación de las pruebas, por consiguiente, habrán de tenerse en cuenta:

- “La confiabilidad en la forma que se haya generado, archivado o comunicado el mensaje;
- La confiabilidad en la forma en que se haya conservado la integridad de la información;
- La forma en que se identifique a su iniciador y cualquier otro factor pertinente”.

Bajo este contexto es importante señalar que el ciudadano en la e-administración cumple un rol muy importante desde la consideración del ciclo de vida del documento en soporte electrónico, donde el registro es la presencialidad electrónica de la información, que en otros medios se conoce como fase de captura, propia de la creación o producción del documento, por parte de un ciudadano o una organización. La tramitación es la parte esencial del medio electrónico o la fase de uso una vez finalizado el aspecto administrativo cuando los documentos deberán estar disponibles y la conservación de los documentos en el largo plazo con relación específica a la utilidad en el ámbito administrativo, jurídico, histórico, de investigación entre otros aspectos.

En contexto colombiano, se ha hecho evidente el desarrollo de una política de gestión de documentos, a partir de la implementación de un sistema de gestión electrónica bajo la articulación del ciclo de vida de los documentos y la incorporación de normas, recursos, instructivos y diseños propios.

Caso particular es lo señalado por la ley 1712 de 2014 Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional, cuyo objeto es regulación del derecho de acceso a la información pública, los procedimientos para el ejercicio y garantía del derecho y las excepciones a la publicidad de información. La mencionada ley incluye en su contextos los principios de transparencia, facilitación, buena fe, no discriminación, gratuidad, celeridad, eficacia, calidad de la información, responsabilidad en el uso de la información y de la divulgación de la información que se constituyen en la interpretación del derecho de acceso a la información como derecho fundamental donde toda persona puede conocer y acceder a la información pública con excepciones contempladas en la misma ley y la Constitución Política. Igualmente, la ley incluye apartados específicos para el tema de los archivos y los sistemas de información contemplándose además la adopción de programas de gestión

documental hacia el establecimiento de procedimientos y lineamientos para la producción, distribución, organización, consulta y conservación de los documentos públicos como herramientas hacia el acceso a la información pública. (Ley 1712, 2014, art.3)

En cuanto al valor probatorio señalado por el gobierno colombiano se hace alusión directa a la política cero papel de la administración pública que señala que para exista esta condición se hace indispensable la implementación de sistemas de archivo que garanticen el mantenimiento y conservación de la autenticidad, fiabilidad y accesibilidad de los documentos en el largo plazo, elementos determinados de una u otra forma por el ciclo de vida de los documentos, para lo cual se ha implementado un modelo de gestión que incluye lineamientos, principios, políticas, funcionalidades, normativas y requerimientos funcionales como requisitos obligatorios, mínimos, optativos y de seguridad para los documentos electrónicos y que a su vez garanticen la gestión y preservación de los documentos establecidos en las Tablas de Retención Documental TRD (Ministerio de Tecnologías de la Información y las Comunicaciones, 2011).

Razón por la cual las buenas prácticas son establecidas como un fundamento esencial que aportan una mejora en la gestión de los archivos, cuyo rol hoy se establece en el acceso a la información pública desde la misma administración del Estado, con la implementación de formatos electrónicos, la optimización y reducción de plazos en las solicitudes información a partir del uso de peticiones, quejas y reclamos, así como en la reducción de costos y el uso de nuevas tecnologías (web 2,0 y web 3,0).

Ejemplos que se han establecido en los organismos del orden estatal, como ha sido el caso de Copnia, (El Consejo Profesional Nacional de Ingeniería – COPNIA, entidad pública que tiene la función de inspeccionar y vigilar el ejercicio de la ingeniería, de sus profesiones afines y de sus profesiones auxiliares en general, en el territorio nacional), quienes implementaron una sede electrónica para la solicitud, consulta, certificaciones y pagos de los trámites de las tarjetas profesionales, establecidos desde el registro y trazabilidad de los documentos, además de la implementación de un control para la entrada y salida de información. (COPNIA, 2014).



## VALOR PROBATORIO DE LOS DOCUMENTOS ELECTRÓNICOS DESDE LA LEGISLACIÓN VENEZOLANA

En el contexto legal venezolano, también se ha previsto que los documentos electrónicos pueden ser utilizados como medios de prueba en procesos administrativos y jurídicos. En palabras de Delgado (2016), las tecnologías de la Información y las Comunicación denominadas (TIC), han contribuido notablemente al desarrollo de la sociedad y a la generación de cambios a nivel social, educativo, económico, administrativos, así como también, en el impulso de proyectos científicos.

En este sentido, con la masificación de los medios tecnológicos, se han generado cambios significativos, que se apoyan en nuevos soportes electrónicos y su uso en procesos educativos, comerciales, legales, entre otros. Rico (2003), señala que las nuevas tecnologías representan “el conjunto de medios que permiten la adquisición, producción, almacenamiento, tratamiento, comunicación, registro y presentación de informaciones en forma de voz, imágenes y datos, contenidos en señales de naturaleza acústica, óptica o electromagnética...” (p. 2).

Con la denominada convergencia teletecnológica, se integran funciones, por medio de las cuales se producen innovaciones e impactos en todos los ambientes, lo que ha permitido importantes cambios. Uno de los campos en los cuales se ha sentido con mucha fuerza el impacto tecnológico, es en el comercio, y en este sentido, se han estado generando documentos en soportes electrónicos para facilitar las gestiones de compra-venta y las actuaciones legales.

Para Serra (2008), el documento electrónico es el que requiere de una máquina que funcione de forma electrónica, sea analógica o digital, para poder ser reproducido o visualizado” (p. 15). Considerando la importancia y todo cuanto implica el desarrollo de las TIC, Mireles (2003) sostiene que ellas están, en esa convergencia de las diferentes tecnologías, y estima han contribuido al cambio radical en conceptos y género, dada la forma tradicional de concebir, en este caso, el documento. Afirma en sentido general, que las TIC “transformaron dramáticamente las concepciones territoriales, espaciales, económicas, políticas y culturales de la sociedad durante el último siglo.” (p. 89).

Siguiendo estos planteamientos, Delgado (2016), se refiere al tratamiento informático generado a través de las TIC, que ha producido modificaciones importantes en la forma de concebir el documento, tanto a nivel de su estructura, como en sus soportes, pero sin perder el carácter que lo legitima como medio e instrumento probatorio y fuente esencial de información para ampliar el acceso al saber.

Con los documentos electrónicos se configuran nuevas posibilidades para la materialización del pensamiento, logrando producir transformaciones en su forma y fondo, y ello genera mutaciones importantes en los procesos de gestión documental, que alude a la producción, descripción, desincorporación y conservación, de conformidad con lo establecido técnicamente, para el cumplimiento de estos procesos, donde ha sido necesario revisar medios de almacenamiento, valoración hasta la disposición legal, sobre todo en lo referente a la desincorporación de documentos. Estos nuevos tipos de documentos también han impactado en los ámbitos comerciales globales, influyendo en la formalidad de establecer las relaciones y en la manera de operar los servicios, donde es necesario determinar la legislación y jurisdicción respectiva.

En el marco de estas consideraciones conceptuales y técnicas, se tiene lo planteado en la Constitución de la República Bolivariana de Venezuela (1999), artículo 110, en el cual se lee textualmente: “el Estado reconocerá el interés público de la ciencia, la tecnología, el conocimiento, la innovación y sus aplicaciones y los servicios de información necesarios por ser instrumentos fundamentales para el desarrollo económico, social y político del país...” (p. 115), con lo cual se le imprime especial interés a los avances tecnológicos, científicos y los servicios de información, para el desarrollo en esas áreas fundamentales.

De acuerdo con Delgado (2016), siguiendo a Peñaranda (2008), el documento electrónico se considera en el ordenamiento jurídico venezolano como medio de prueba, al expresar que este tipo de documentos...“es aquel que proviene de cualquier medio de Informática o que también sea formado o realizado por ésta...” (p. 8). De modo que el documento electrónico posee un lenguaje convencional (bits), que no se corresponden con el documento clásico en papel, y cuyo soporte es electrónico o informático.

No obstante, ambos constituyen un medio de prueba. Aun cuando, para el documento electrónico el valor probatorio, tal como lo establece Rico (2003) "... depende de su autenticidad, esto es de la posibilidad de identificar a su autor.

Esta situación nos hace considerar los extremos requeridos para que un documento electrónico sea considerado como auténtico y obligue legalmente a su emisor." (p. 99). Agrega la autora, que la forma de adquirir la autenticidad va a estar relacionada con la firma de quien suscriba el documento y comprende asimismo el nombre y apellido, los cuales se colocan al final del escrito, en respaldo de todo lo expresado. En función de lo anterior, las funciones jurídicas de una firma según Fuentes (2007), son: "a) La identificación de un sujeto con la misma; y b) la manifestación expresa del consentimiento de dicho sujeto con el contenido o texto al cual se suscribe su firma." (p. 250). La firma autógrafa, tal cual la conocemos, le imprime al documento la aprobación del contenido por parte de su autor, y ese aval le otorga responsabilidad y compromiso institucional, laboral o personal, por cuanto constituye prueba fehaciente de quien lo emitió.

En este sentido, Delgado (2016), sigue planteando que La Ley Decreto sobre Mensajes de Datos y Firmas Electrónicas en su artículo 16, se refiere al vínculo que se da entre el signatario y el mensaje de datos y la atribución de autoría, cuando determina que "...tendrá la misma validez y eficacia probatoria que la ley otorga a la firma autógrafa..." (p. 44) y según el artículo 17 de la misma ley, "la firma electrónica que no cumpla con los requisitos señalados en el artículo anterior no tendrá los efectos jurídicos que se le atribuyen en el presente Capítulo; sin embargo, podrá constituir un elemento de convicción valorable conforme a las reglas de la sana crítica." (p. 44).

En el contexto legal venezolano, se concede a las partes que intervienen en un proceso o juicio, cuando hay hechos que han sido narrados o alegados por ellas y en los que no están de acuerdo o existe contradicción, ciertos elementos para que puedan comprobarlos o demostrarlos, y estos son llamados medios de prueba o medios probatorios, que según La Roche (2004),..."constituyen el instrumento, el vehículo, es decir, los modos aceptados por la ley para trasladar al conocimiento del juez el resultado de la prueba..." (p.178) y los mismos permiten la verificación de los expresados

hechos, para que el juez encuentre los motivos de su convicción o certeza sobre ellos y proceda a dictar la decisión o sentencia que resuelva ese conflicto presentado entre las partes.

Los documentos electrónicos constituyen medios con los cuales se establecen también relaciones comerciales, que se apoyan en lógicas diferentes a las tradicionales y para determinar su valor como medio de prueba, es necesario considerar fundamentalmente su carácter de original o copia. Delgado (2016) insiste que ello plantea una dificultad inicial, por cuanto independientemente del soporte y el medio de transmisión, siempre podría pensarse que el original lo es, tanto el que está en la memoria del computador, como aquel que ha sido enviado por correo electrónico.

Para Rico (2003), el criterio de inalterabilidad es aplicable pero sólo a la grabación de un documento electrónico en un soporte informático distinto al que se utilizó en su generación. Lo que se pretende en el manejo de la inalterabilidad, es la conservación de este tipo de documento, en forma íntegra, lejos de manipulaciones, alteraciones que puedan presentarse debido a su vulnerabilidad. Sin duda el documento electrónico ha venido a transformar la forma global de hacer negocios.

En Venezuela, el Decreto Ley sobre Mensajes de Datos y Firmas Electrónicas, en su objeto y aplicabilidad, reconoce la eficacia y el valor jurídico de la firma electrónica. En el artículo 4, determina que “Los mensajes de datos tendrán la misma eficacia probatoria que la ley otorga a los documentos escritos. En esta fuente legal se afirma que dichos mensajes: “... gozan de tarifa legal y producen plena prueba entre las partes y frente a terceros de acuerdo a su naturaleza.

De modo pues que el documento electrónico, forma parte de las pruebas contenidas en el artículo 395 del Código de Procedimiento Civil”. (p.32). La ley es clara y contundente al hacer notar que si en algún momento es necesario el cumplimiento de solemnidades o formalidades, éstas se realizarán siguiendo los mecanismos descritos en el mencionado artículo 6 de dicho Decreto-Ley.

Delgado (2016), señala que el valor probatorio del documento electrónico en Venezuela, es de reciente data y se podría decir que, a pesar de los avances establecidos en la ley, todavía hay que cumplir con disposiciones que no

hacen tal fácil el procedimiento legal. Por otro lado, Salcedo (2007) hace una referencia en este sentido, apoyándose en el Código Orgánico Procesal Penal "...que consagró el principio de la libertad de la prueba, inspirado en el Código de Procedimiento Civil de 1986, que en su artículo 395, abrió, con la amplitud que proporcionan los avances científicos y técnicos, los medios probatorios ..." (p. 96). El documento electrónico puede ser un medio de prueba en caso de un delito informático o de otra naturaleza. En cuanto al Código de Procedimiento Civil Venezolano, se establece en el artículo 395: "Son medios de prueba admisibles en juicio aquello que determina el Código Civil, el presente Código y otras leyes de la República.

El Decreto Ley sobre Mensajes de Datos y Firmas Electrónicas (LMDFE), en el artículo 7 señala que se considerará un Mensaje de Datos para permanecer íntegro, debe cumplir con los siguientes elementos:

- a. inalterabilidad;
- b. conservación del mensaje;
- c. legibilidad;
- d. fiabilidad de los sistemas que lo producen;
- e. admisibilidad;

Adicional a lo anterior, RICO (2003) señala que los mensajes de datos deben promoverse, en principio, bajo la prueba documental. No obstante, al tratarse de un procedimiento que emplea la técnica en la generación del soporte documental, es necesario el auxilio de una experticia, apoyándose en los criterios señalados.

El documento electrónico tiene una particularidad, en virtud de que fue generado mediante un formato con características propias, posee un contenido, en tanto que manifestación de voluntad. Estos tipos de documentos al ser producidos a partir de actos humanos, y representaciones de un hecho, necesariamente implican medios probatorios frente a un proceso judicial.

Delgado (2016), expresa que en Venezuela, el valor probatorio del documento electrónico, requiere de procedimientos diferentes al convencional, pero igualmente la legislación venezolana en las leyes mencionadas, ha considerado

esta nueva tendencia, y es factible, que en las próximas reformas de códigos y leyes se amplíen los criterios y se establezcan otras maneras de concebir el documento electrónico y se reglamenten procedimientos, para proteger los derechos y deberes en esta etapa de relaciones virtuales. Más aún, cuando su uso ya forma parte de la nueva manera de hacer negocios y de interactuar, cada vez más, con operaciones comerciales a través de la Internet.

Ovalle (2008), plantea que en Venezuela la admisión del documento electrónico no representa ningún problema como medio de prueba, pero dada la forma y facilidad con la que puede ser creado o modificado, es necesario que se combinen los procedimientos técnicos con las normas jurídicas, a fin de que los intervinientes, puedan tener prueba de la operación electrónica realizada y que ésta sea similar al negocio cuando se plasma en papel.

A pesar de los avances logrados mediante el Decreto-Ley, todavía hay que cumplir con disposiciones que no hacen tan fácil el procedimiento legal para el manejo del documento como medio de prueba. Si se amplía las posibilidades de promover las pruebas que se puedan obtener del avance de la ciencia y la tecnología, a través de medios de pruebas legales, que estuvieran determinados expresamente, se lograría promoverlas directamente en el proceso y contribuir con su realización al esclarecimiento de los hechos y a la aplicación de la justicia, determinando expresamente el procedimiento que se debe seguir para su promoción y evacuación, con toda la seguridad requerida para otorgarles confiabilidad, inalterabilidad, conservación de su contenido y la posibilidad de su recuperación.

## ANÁLISIS DE RESULTADOS

Para facilitar la comparación entre las dos legislaciones y marcos normativos estudiados, se elaboró un cuadro, en el que se visualizan las diferencias, semejanzas y complementariedades de los países incluidos en el estudio, cuyas leyes que rigen los documentos electrónicos fueron objeto de comparación.

Cuadro 1 – Comparativo entre las legislaciones de Colombia y Venezuela en materia de documentos electrónicos. (Continúa)

Categorías de comparación	Legislación colombiana	Legislación venezolana
Premisas que apoyan la validez de los documentos electrónicos	Integración digital, diversificación de formatos tras la evolución de las tecnologías y su impacto en los procesos comerciales y productivos.	Las tecnologías de la Información y las Comunicación denominadas (TIC), han contribuido notablemente al desarrollo de la sociedad y a la generación de cambios a nivel social, educativo, económico, administrativos, así como también, en el impulso de proyectos científicos.
Condiciones o requisitos para los documentos electrónicos sean considerados como medios de prueba.	Los requisitos que debe cumplir los documentos electrónicos son: fiabilidad, inalterabilidad y rastreabilidad. (...) los documentos electrónicos están en capacidad de brindar similares niveles de seguridad que el papel y, en la mayoría de los casos, un mayor grado de confiabilidad y rapidez, especialmente con respecto a la identificación del origen y el contenido de los datos, siempre que se cumplan los requisitos técnicos y jurídicos plasmados en la ley. (Artículo 2 de la Ley 527).	Los documentos electrónicos para poder considerados medios de prueba deben cumplir los siguientes requisitos y condiciones: a. Inalterabilidad, b. conservación del mensaje, c. legibilidad, d. fiabilidad de los sistemas que lo producen, e. admisibilidad. (Ley sobre mensajes de datos y firmas electrónicas, 2011).

Cuadro 1 – Comparativo entre las legislaciones de Colombia y Venezuela en materia de documentos electrónicos.

Categorías de comparación	Legislación colombiana	Legislación venezolana
Uso y características de la firma digital.	Cuando una firma digital haya sido fijada en un mensaje de datos se presume que el suscriptor de aquella tenía la intención de acreditar ese mensaje de datos y de ser vinculado con el contenido del mismo". (Ley 527, 1999, art. 28).	La firma electrónica... "tendrá la misma validez y eficacia probatoria que la ley otorga a la firma autógrafa..." (p. 44)... "la firma electrónica que no cumpla con los requisitos señalados en el artículo anterior no tendrá los efectos jurídicos que se le atribuyen en el presente Capítulo; sin embargo, podrá constituir un elemento de convicción valorable conforme a las reglas de la sana crítica." (p. 44). Decreto de Ley sobre Mensajes de Datos y Firmas Electrónicas (2011).

(Conclusión)

Fuente: Elaboración propia.

Luego de presentar los elementos conceptuales y jurídicos que orientan la concepción, aplicación y fuerza probatoria de los documentos electrónicos en Colombia y Venezuela, y considerando tres categorías esenciales de la comparación, a saber: las premisas que apoyan el uso de este tipo de documentos, los requisitos y condiciones que deben tener los documentos electrónicos y el uso y características de las firmas electrónicas, es posible encontrar las siguientes semejanzas y diferencias.

En cuanto a las premisas que orientan y soportan el uso de los documentos electrónicos en ambos países, se reconoce que el haber llegado a la utilización de este tipo de soportes como medios probatorios es producto de la evolución de las tecnologías digitales y de los procesos de convergencia interactiva, que generaron las condiciones favorables para que los procesos comerciales, productivos y jurídicos se puedan soportar en el uso de soportes y dispositivos tecnológicos, en los cuales se expresen situaciones y acciones que tengan carácter de evidencia.



Sobre las condiciones que deben poseer los documentos electrónicos para que puedan ser considerados como medios de prueba, ambas legislaciones toman en cuenta el carácter de inalterabilidad y fiabilidad, por lo que es preciso establecer sistemas que garanticen en ambos países estas condiciones y con ello contribuir con el uso masificado de los documentos electrónicos. Solo la legislación colombiana le da importancia a la condición de rastreabilidad, como un requisito esencial de los documentos electrónicos, entendiendo este requisito como la posibilidad de que se pueda constatar la procedencia y autenticidad de los contenidos expresados, en virtud de los sujetos o instituciones generadoras. La legislación venezolana, por su parte, habla de admisibilidad, entiendo como tal condición el carácter de fidedigno y propio de las entidades productoras de dichos documentos.

En cuanto a las condiciones y requisitos, se puede hablar de relaciones de complementariedad entre las dos legislaciones, cuyos principios al ser considerados de forma compartida, es posible que tales legislaciones aporten a una mayor comprensión sobre la aplicación e instrumentalización del uso de los documentos electrónicos como medios de prueba.

Finalmente, en relación con las firmas electrónicas, la legislación colombiana las plantea como firmas digitales, y se asumen como válidas en tanto que se puede presumir que tales firmas están vinculadas con los sujetos que produjeron sus contenidos. Para la legislación venezolana, las firmas electrónicas tendrán la misma eficacia y fuerza probatoria, con base también en la relación de la firma con quien la produjo.

Se considera que Colombia tiene mayores referentes legales, políticos y normativos que le dan marco al uso de los documentos electrónicos, por cuanto no solo se cuenta con leyes que fundamentan su uso, sino que también se cuenta con políticas generadas por el Estado en lo que concierne a otros ámbitos de aplicación de los medios y recursos tecnológicos, no solo en contextos productivos y comerciales, sino educativos y culturales, debido a que los planes formulados por los entes ministeriales que se ocupan de las tecnologías de información y comunicación y su apropiación por parte de los estamentos sociales, así lo han definido.

## CONCLUSIONES

La conceptualización de los documentos en general apunta a reconocerlos como bienes muebles, productos de la representación del pensamiento o de la voluntad de quien los genera. Ahora bien, en el caso de los documentos electrónicos, son contenidos en soportes electrónicos o máquinas informáticas y cuyo mensaje puede ser “magnitudes físicas que presentan en forma codificada unas declaraciones o representaciones y que son susceptibles de registro, proceso y transmisión, lo que traduce que este documento requiere de medios o instrumentos electrónicos como medio de vida y reproducción.

Las legislaciones venezolana y colombiana contemplan la eficacia y fuerza probatoria de los documentos electrónicos, como medios que soportan gestiones comerciales, financieras y como evidencias en procesos jurídicos. Para hacer efectivo el uso de los documentos electrónicos como medios de prueba, se deben considerar un conjunto de criterios, uno de los cuales tienen que ver con la inalterabilidad y la conservación, en forma íntegra, lejos de manipulaciones, alteraciones que puedan presentarse debido a su vulnerabilidad. Sin duda el documento electrónico ha venido a transformar la forma global de hacer negocios, de participar activamente en actuaciones legales, y en suma en diversas áreas de la gestión pública y privada.

En Colombia se aprecian mayores referentes legales y políticos que le otorgan un marco de acción claro al tema de los documentos electrónicos, su adecuada gestión y su pertinencia en el contexto de la construcción de un Estado orientado hacia acercar a los ciudadanos y ciudadanas con los procesos de participación efectiva, transparente y ágil en la gestión pública.



## REFERENCIAS

- ALBERCH I FUGUERAS, R. Gestión de documentos: entre la alianza interdisciplinar y la teoría del rompecabezas. *Acervo*, [S.l.], v. 28, n. 2 jul-dez, p. 110-120, nov. 2015. Disponível em: <<http://revista.arquivonacional.gov.br/index.php/revistaacervo/article/view/625/645>>. Acesso em: 22 Set. 2017.
- ASAMBLEA NACIONAL. Decreto Ley de Mensaje de Datos y Firmas Electrónicas de Venezuela. Caracas. In. *Gaceta Oficial* N° 37.148 del 28 de febrero, 2011.
- ASAMBLEA NACIONAL. Constitución de la República Bolivariana de Venezuela. Caracas. Venezuela. In. *Gaceta Oficial* N° 5.453. 24 mar. 2000.
- COLOMBIA. \_\_\_\_\_. Código de Comercio. [Código]. Bogotá: Legis, 2015
- \_\_\_\_\_. Código de Procedimiento Civil. In: *Gaceta oficial* N° 4209 15 set.1990.
- \_\_\_\_\_. *Código de procedimiento Civil*. Bogotá: Legis, 2015
- \_\_\_\_\_. *Código de Procedimiento Civil*, (Decreto 1400, 1970).
- \_\_\_\_\_. Congreso de Colombia. (6 de marzo de 2014) Artículo 3. [Título I]. *Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional* [Ley 1712 de 2014]
- \_\_\_\_\_. Congreso de Colombia. (18 de agosto de 1999). Artículo 7. [Parte I, Título II]. Reglamenta el acceso y uso de los mensajes de datos, del comercio electrónico y de las firmas digitales, y se establecen las entidades de certificación [Ley 527 de 1999]
- \_\_\_\_\_. Congreso de Colombia. (18 de agosto de 1999). Artículo 11. [Parte I, Título II]. Reglamenta el acceso y uso de los mensajes de datos, del comercio electrónico y de las firmas digitales, y se establecen las entidades de certificación [Ley 527 de 1999]
- \_\_\_\_\_. Congreso de Colombia. (18 de agosto de 1999). Artículo 10. [Parte I, Título II]. Reglamenta el acceso y uso de los mensajes de datos, del comercio electrónico y de las firmas digitales, y se establecen las entidades de certificación [Ley 527 de 1999]
- \_\_\_\_\_. *Corte Constitucional*. Sentencia C-662 de 2000.
- \_\_\_\_\_. Ministerio de Tecnologías de la Información y las Comunicaciones. Dirección de Gobierno en Línea Guía No. 6. Programa gobierno en línea, 2011. Disponible en: < [http://programa.gobiernoenlinea.gov.co/apc-aa-files/Cero\\_papel/guia-6-sistemas-de-gestion-de-documentos-v1.pdf](http://programa.gobiernoenlinea.gov.co/apc-aa-files/Cero_papel/guia-6-sistemas-de-gestion-de-documentos-v1.pdf).
- \_\_\_\_\_. Superintendencia de Industria y Comercio. (18 jun. 2004). Concepto 04044586.
- COPNIA. Disponible en: < <https://copnia.gov.co/>. Acceso en: 12 jun. 2016.

DELGADO, F. El documento electrónico como medio de prueba en la legislación venezolana. In: Las ciencias de la información y documentación. Miradas desde Venezuela y Colombia. 2016. Disponible en: < <http://www.cidtec.luz.edu.ve/images/stories/Libros/las%20ciencias%20de%20la%20informacion%20y%20documentacion-miradas%20desde%20venezuela%20y%20colombia.pdf>. Acceso en: 12 jun. 2016.

FUENTES, P. F. Marco Legal de la Informática y la Computación. Caracas: Vadell Hermanos Editores. 2007. Disponible en: < <http://www.ciscra.org/mat/mat/term/1676>. Acceso en: 12 jul. 2017.

GIMÉNES-CHORNET, V.

INTERNATIONAL COUNCIL ON ARCHIVES (ICA). Multilingual Archival Terminology. 2015

LA ROCHE, A. Anotaciones del Derecho Procesal Civil. Procedimiento Ordinario. Maracaibo: Centro de Estudios Jurídicos (CEJUZ.) Mireles, 2004.

NÚÑEZ CONTRERAS, L. (1999, p. 104). Concepto de documento, en Archivística. Estudios básicos. Sevilla: Diputación Provincial, 1983, p.31.

OVALLE, N. Valor probatorio del documento electrónico en Venezuela y otros países de Latinoamérica, 2008. Dissertação (Mestrado em Ciência da Informação - tese não publicada). Universidad del Zulia. Maracaibo.

PARRA QUIJANO, J. El documento electrónico y su alcance probatorio. Trabajo presentado en I Convención internacional de derecho informático, documentación y documento electrónico Universidad Externado de Colombia. Bogotá: U. Externado, 2006.

PEÑARANDA QUINTERO, H. El documento Electrónico (Comercio Electrónico, Actos cibernéticos procesales). Venezuela, Maracaibo: Editorial Universidad del Zulia (EDILUZ), 2008.

RIVAS FERNÁNDEZ, J.B. La experiencia de Costa Rica en los procesos de valoración, selección y eliminación documental. In: \_\_\_\_\_ Códices. Bogotá, v. 7, n.1 p.81-109, 2011.

RICO CARRILLO, M. Comercio Electrónico, Internet y Derecho. Venezuela: Editorial Legis. 2003.

RUIZ, W.; RUIZ, D. Medios de Prueba y Criminalística. Venezuela, Barquisimeto: Editorial Horizonte C.A., 2009.

SALCEDO CÁRDENAS, J. La Prueba Documental. 2ª.ed. Caracas: Editorial Gráficas León, C. A., 2007.

SERRA SERRA, J. Los Documentos Electrónicos. Qué son y cómo se tratan. España: Ediciones Trea, 2008.

## **Valor probatório dos documentos eletrônicos: pontos de vista de Colômbia e Venezuela**

### **RESUMO**

*Partindo do conceito de documento eletrônico, entendido como a representação do pensamento, materializado em suportes magnéticos que garantem o acesso imediato a dados relevantes que também podem ser utilizados como meio de teste, uma análise comparativa da legislação e normatividade colombiana é realizado e venezuelano, relacionado a este tipo de documentos. A metodologia consistiu em uma pesquisa documental, que integrou a comparação de princípios conceituais e legais com algumas boas práticas aplicadas na Venezuela e na Colômbia no gerenciamento de documentos eletrônicos. Os resultados são evidências claras da evidência nas leis de ambos os países, que tem um impacto significativo no acesso à informação, transparência e processos de participação pública na gestão pública e privada. O escopo e as implicações deste tipo, bem como a evidência do uso de documentos legais nos processos judiciais, de documentos e suas implicações na gestão de todos os tipos de entidades. Os princípios que devem orientar a aplicação de documentos eletrônicos como meio de prova são: inalterabilidade, preservação de mensagens, legibilidade, confiabilidade dos sistemas que o produzem e admissibilidade.*

**Palavras-chave:** documentos eletrônicos, legislação venezuelana, legislação colombiana, valor probatório.

## ***Probative value of electronic documents: views from Colombia and Venezuela***

### **ABSTRACT**

*Starting from the concept of the electronic document, understood as the representation of the thought, materialized in magnetic supports that guarantee the immediate access to relevant data that in turn can also be used as means of test, a comparative analysis of the Colombian legislation and normativity is realized And Venezuelan, related to this type of documents. The methodology consisted of a documentary research, which integrated the comparison of conceptual and legal principles with some good practices that are applied in Venezuela and Colombia in electronic document management. As results it is evident that electronic documents are recognized as evidence in the laws of both countries, which has a significant impact on access to information, transparency and processes of citizen participation in public and private management . As a conclusion, the consideration of electronic documents, as evidential evidence, with possibilities of use in legal proceedings, which also raises the efforts made in the two countries, whose legislation were compared, with a view to highlighting The scope and implications of this type of documents and their implications in the management of all types of entities. The principles that should guide the application of electronic documents as means of proof are: inalterability, message preservation, legibility, reliability of the systems that produce it and admissibility.*

**Keywords:** *electronic documents, Venezuelan legislation, Colombian legislation, probative value.*

# Ontologia digital arquivística

Charlley Luz<sup>1</sup>

### RESUMO

Capítulo que aborda como a representação da informação em arquivos pode colaborar na criação de ontologias e como estes podem apoiar a interoperabilidade e a preservação dos documentos. Trata da Descrição Arquivística e a interoperabilidade de dados digitais e como os contextos podem ser representados em Ontologias, respeitando-se a cadeia de custódia documental, necessária para a garantia de autenticidade documental em Repositórios Digitais Confiáveis - RDC-Arq. Como resultados, apresenta as ontologias digitais arquivísticas que podem ser usadas para representar padrões de descrição de documentos de arquivo e que podem contribuir para a manutenção de contexto em sistemas de ambientes digitais.

**Palavras-chave:** Arquivística. Ontologias. Descrição arquivística. Repositórios digitais confiáveis.

---

<sup>1</sup> Arquivista, mestre em Ciências da Informação pela ECA-USP em São Paulo/Brasil.

## INTRODUÇÃO

Parte da memória humana está sendo registrada em meio eletrônico no mínimo pelos últimos 40 anos, se considerarmos a época do processamento dos cartões perfurados. Todas as definições em relação às opções tecnológicas de estruturação da informação terão um impacto nos documentos que serão acessados no futuro; é como se vivêssemos, agora, o primitivismo digital (LUZ, 2015). Parte da nossa história, ou seja, o que geramos hoje de informações digitais corre o perigo de se perder, ficar sem leitura no futuro, resultado da obsolescência tecnológica ou de decisões equivocadas.

É necessário o uso de estratégias de preservação digital que poderão ajudar nossos descendentes a entenderem os dias de hoje e acessarem essas informações como fontes de prova e pesquisa num futuro. Para tanto, vamos abordar neste capítulo como a representação da informação em arquivos pode colaborar na criação de ontologias e como eles podem apoiar a interoperabilidade e a preservação dos documentos.

A informação institucionalizada, orgânica no caso dos documentos arquivísticos, mantida em sistemas de arquivos só é fidedigna e autêntica se estruturada em ambientes de confiança. Parte desta informação precisa evoluir e migrar durante o tempo, mantendo todos os dados acerca dos documentos e de seu contexto original para garantir a autenticidade e fidedignidade, o que ocorre por meio de metadados. Este processo está digitalizado e mediado por sistemas e softwares e as linguagens devem ser entendidas por agentes inteligentes computacionais.

O tratamento informacional é a organização dos documentos para descoberta e uso e, assim, atendem às necessidades dos usuários. No contexto digital e de redes, esses documentos informáticos, também identificados como informação digital, precisam passar por processos que garantem sua padronização, o que facilita as operações de busca e revocação em plataformas e sistemas. Portanto, a organização da informação pressupõe, também, a criação de padrões de interoperabilidade.

A descrição arquivística, função matricial da arquivologia segundo Lopes (1996), cria a representação da informação presente (e sobre) os arquivos, e que gera os instrumentos de pesquisa e parte do que é identificado como



informação arquivística. Este deve ser considerado o padrão de representação para arquivos, aplicando-se, para tanto, normas nacionais ou internacionais. A informação arquivística pode trazer representações da estrutura da informação orgânica e dados sobre a estrutura funcional das instituições pois utiliza metadados para apresentar os indícios de fidedignamente da informação que representa, o que estimula a confiança na autenticidade por meio da manutenção de cadeia de custódia e estimula a preservação em longo prazo.

Para Duranti (1997), a descrição arquivística evidencia o vínculo de arquivo com seu contexto documentário, e pode ser considerada a principal forma de perpetuar e autenticar o significado dos documentos. Em tempos de migrações de ambientes informáticos é, provavelmente, o melhor método para garantir a autenticidade no longo prazo. E isto atua para garantir a integridade da cadeia de custódia documental e dos ambientes e plataformas digitais que a compõem e onde a interoperabilidade teria, portanto, um papel em sua manutenção.

Ao preparar a interoperabilidade, aplicando a descrição, também estamos facilitando a preservação em longo prazo. A preparação da interoperabilidade engloba a aplicação de significado aos objetos informacionais, aos registros e recursos *on-line* de modo que tanto os humanos quanto os agentes computacionais entendam. Para Lopes (2009, p. 359), “vive-se em uma nova etapa, em que a presença da informação registrada mensurável em bytes tem uma significação muito maior do se possa, a priori, imaginar”. E a informatização, resultando na informação digital em rede, trouxe para os arquivos a realidade dos sistemas de gestão documental presentes no universo corporativo e na estrutura estatal.

Para o Conselho Nacional dos Arquivos (Conarq, 2011), considera-se os seguintes tipos de sistemas informatizados de arquivo: o Sigad, que é o modelo brasileiro definido pelo e-Arq e o RDC-Arq - Repositório Digital Arquivístico Confiável. A condição para a existência do Sigad, responsável pela guarda de documentos correntes e intermediários, é a utilização de instrumentos de gestão, enquanto o RDC-Arq é o arquivo permanente, aquele ambiente que contém a custódia dos documentos autênticos em ambiente confiável.

Quem lidar com sistemas de informação de arquivo está buscando tornar a informação digital gravada acessível, como descreve Quisbert (2006). Para o autor, os documentos arquivísticos são entidades lógicas, em vez de entidades físicas, e afirma ainda que a interoperabilidade é possível por meio de atitude e uso de padrões abertos. A acessibilidade e a preservação em longo prazo devem ser capazes de esticar a longevidade dos registros no decorrer do tempo-espaço.

## ONTOLOGIAS DIGITAIS ARQUIVÍSTICAS

O processo de uniformização e a padronização contribuem para a interoperabilidade da informação digital. No caso dos arquivos, a informação arquivística pode ser definida como toda informação gerada pelas instituições que disponibilizam arquivos e dados sobre os fundos documentais que administram, possuem, descrevem e disponibilizam em redes.

Os sistemas informatizados de gestão e guarda documental geram dados sobre seu acervo por meio de metadados que tornam possíveis seu processamento e disponibilização. São informações dos contextos dos fundos (as respectivas estruturas de organização deste acervo, ou seja, a instituição, as seções e séries documentais) que são contidas em instrumentos descritivos como guias, em perfis de metadados ou em *links* semânticos ancorados em ontologias.

Sabe-se que uma ontologia é a “soma de uma série de conceitos relevantes que representam o conhecimento compartilhado em um domínio”, como descreve Moreiro Gonzáles (2011, p. 77). Segundo o autor, a ontologia é uma especificação explícita, em que a conceitualização é testada por meio das múltiplas relações entre os conceitos e validados pelos axiomas (assertivas sobre o domínio) destes conceitos. Trata-se, portanto, de uma conceitualização formal de um domínio específico ou uma parcela da realidade. O domínio de uma ontologia pode ser uma área temática ou uma área de conhecimento, ou uma parcela da realidade, o que pode também representar o contexto e a informação arquivística.

Além disso, uma ontologia pode ser uma aplicação tecnológica aplicada a ambientes digitais. Por exemplo, as ontologias também são artefatos de camadas para a Web semântica, como destacam Berners-Lee, Hendler e Lassila (2001), que utilizam-se de ontologias e metadados com o intuito de estruturar e dar significado ao conteúdo das páginas Web. Gilchrist (2003) afirma que duas áreas onde o uso de ontologias está sendo promovido são a gestão do conhecimento e a Web semântica. As ontologias possuem, nestas aplicações, uma abordagem informacional e representam fontes de estudo para a organização da informação e do conhecimento.

Há no *corpus* da área, de modo não unânime, a comparação de ontologias com as linguagens documentárias, tal como fizeram Pedraza-Jiménez, Codina e Rovira (2007, p. 574), ao afirmarem que “Las ontologías pueden considerarse lenguajes documentales con distintos niveles de estructura, pero a diferencia del tesoro tradicional están elaboradas con una sintaxis comprensible para los ordenadores”. Porém esta visão poderia ser relativizada ao incluirmos as ontologias como mais uma KOS (Knowledge Organization System). Para a área da ciência da informação isto pode representar mais uma frente de atuação. Hodge (2000) observa que ontologia é o mais novo rótulo ligado a alguma KOS ou sistema de organização do conhecimento.

Partindo da definição de que as ontologias representam realidades, podemos então verificar que o contexto arquivístico trata das realidades da geração dos documentos (ambiente de geração orgânica) e, com base nessa afirmação é possível deduzir que as ontologias servem, também, na descrição de realidades institucionais que geram os arquivos, e assim poderiam permitir acesso em longo prazo ao contexto de geração dos documentos, inclusive os arquivísticos ao representarem este domínio arquivístico.

Podemos identificar como ontologia digital arquivística um produto de linguagem, da descrição de um contexto, de uma estrutura e de um registro ao longo de um tempo, realizado num ambiente digital por meio de plataformas tecnológicas utilizando aplicações de ontologias leves para tal. Isto pode auxiliar no processo de perenizar o contexto, importante na preservação da informação digital e para validar a cadeia de custódia, o que é primordial na garantia da autenticidade documental, missão essencial do arquivista.

O documento arquivístico serve como testemunho da ação que o gerou, para tanto, ele precisa ser confiável e autêntico. Ele precisa trazer as informações de contexto e o conjunto de metadados que o descrevem para sua transmissão entre custodiadores. Uma ontologia pode ser capaz de fornecer informações contextuais a cada objeto informacional identificado como documento arquivístico, complementando assim sua estrutura de metadados administrativos, descritivos e estruturais (NISO, 2004) resultantes do processo de organização e representação da informação de arquivo, como veremos a seguir.

## **ORGANIZAÇÃO, REPRESENTAÇÃO E PRESERVAÇÃO**

A gestão da informação realiza a constituição e a preservação da memória por meio da organização e seleção de documentos e registros. As principais funções que envolvem esta atividade são: identificar a existência dos documentos assim que estiverem disponíveis, agrupar sistematicamente esses documentos para estruturar coleções e dossiês, além da produção de sistemas para disponibilizar os documentos, oferecendo metadados úteis para sua identificação e guarda.

Na transferência de custódia de documentos arquivísticos de um acervo corrente ou intermediário organizado para um acervo permanente, é necessário realizar o processamento técnico para custódia e, portanto, sua avaliação. Neste caso, o processamento para material de suporte físico, não digital, frequentemente inclui ações como deixar os documentos em quarentena, limpar e reembalar, além da avaliação da classificação para criação do arranjo.

O processamento técnico visando à organização também ocorre no meio digital, mas os métodos serão diferentes. Para o projeto Pericles (2014), a quarentena será uma verificação de vírus, e reembalagem vai significar mover da mídia de armazenamento do formato original, observando a fixidez. O planejamento para a preservação em longo prazo do material digital vai exigir extenso conjunto de metadados sobre a composição do arquivo digital, sobre formatos de arquivo, sistemas operacionais, tamanhos de arquivo e informações de fixidez para o eficiente planejamento de

preservação evidenciados por metadados. A descrição arquivística, a parte final do processamento técnico, que cria a representação, gera também um conjunto de metadados. Devemos considerar que “os metadados fazem parte do processo de preservação dos documentos arquivísticos digitais e que os mesmos devem ser preservados juntamente com os documentos arquivísticos digitais” (INNARELLI, 2015, p. 209). Aqui temos os metadados administrativos, descritivos e estruturais.

A preservação digital engloba ações visando manter a integridade e a acessibilidade dos documentos digitais ao longo do tempo. “Devem alcançar todas as características essenciais: físicas (suporte), lógicas (software e formato) e conceituais (conteúdo exibido)”, conforme destaca o Conarq (2016). Uma estratégia de preservação deve ser aplicada desde o planejamento da gestão da informação arquivística, definindo-se prazos de guarda e eliminação e aplicado diretamente no objeto digital via sistemas de gerenciamento, desde sua gênese. Para o conselho, as ações de preservação digital devem ocorrer desde o início do ciclo de vida do documento.

Para o órgão, esse é um processo de manutenção de aspectos que estimulam a confiabilidade, em que “Tradicionalmente, a preservação de documentos arquivísticos se concentra na obtenção da estabilidade do suporte da informação” descreve o Conarq (2011, p. 35). O Conselho destaca que, para realizar a preservação, é necessário realizar diversas operações, entre elas “mudanças de suporte e formatos, bem como atualização do ambiente tecnológico. A fragilidade do suporte digital e a obsolescência tecnológica de hardware, software e formato exigem essas intervenções periódicas” (Conarq, 2011, p. 35).

## SISTEMAS DE INFORMAÇÃO ARQUIVÍSTICOS

Para Quisbert (2006), um sistema de informação, no sentido amplo, é a totalidade de todas as atividades formais e informais de representação de dados e de processamento dentro de uma organização, incluindo a comunicação vinculada, tanto internamente quanto com o mundo exterior, e sistema de informação em sentido estrito são subsistemas baseados em computador, destinados a fornecer gravação e apoio a serviços de operação e gestão organizacional.

Para o autor, um sistema de informação de arquivo é um tipo especial de sistema de informação. Como cada tipo especializado de sistema de informação, o desenvolvimento de sistemas de informação de arquivo deve ser precedido de uma forma especializada de pensar. Quisbert (2006) propõe uma forma especializada de pensar, um quadro para o desenvolvimento de sistemas de informação de arquivo. O quadro proposto por Quisbert tem como base o Modelo australiano *Record Continuum* e o Modelo de Referência OAIS como principais fundamentos.

Para o autor, a interoperabilidade e a preservação em longo prazo devem encaminhar a longevidade dos registros no curso do tempo-espço. Os conceitos de sistemas informatizados de arquivo, sistemas de gestão documental, sistema de informação de arquivo e sistemas de gerenciamento de arquivos são como sinônimos, pois todas são aplicações de tecnologia da informação que apoiam, por intermédio da “informatização”, as operações de gestão arquivística de documentos.

A informação é conhecimento que pode ser trocado. A intenção de um sistema de informação de arquivo, portanto, não deve ser apenas o de armazenamento, mas de garantir a autenticidade, a perenidade da cadeia de custódia, além de estimular o uso de seu acervo por meio difusão e da disponibilização de metadados aos agentes computacionais por meio da Web semântica, conforme afirma Quisbert (2006). O Conarq considera os tipos de sistemas informatizados de arquivo o Sigad e o RDC-Arq. Para a gestão de arquivos correntes e intermediários utilizam-se os ambientes de Sigad e para a guarda permanente de informação digital em documentos o ambiente adequado é o RDC-Arq.

## O SIGAD E O MODELO DE REQUISITOS E-ARQ

Um sistema informatizado de gestão arquivística de documentos (Sigad) permite ter uma compreensão da organização como um todo ao estruturar os instrumentos que propiciem a existência da gestão de documentos arquivísticos, principalmente o plano de classificação e a tabela de temporalidade.

Trata-se de um sistema desenvolvido para produzir, receber, armazenar, dar acesso e destinar documentos arquivísticos digitais. Pode compreender um software particular, determinado número de softwares integrados, adquiridos ou desenvolvidos por encomenda, ou uma combinação desses recursos. Deve realizar todas as operações técnicas da gestão arquivística (da criação até a destinação final - a eliminação ou a guarda permanente).

O funcionamento de um Sigad depende de implementação de procedimentos e políticas de gestão de documentos como, por exemplo, um plano de classificação e de destinação. Eles carregam as regras do sistema que fazem a gestão dos documentos arquivísticos digitais. Estes são os requisitos que veremos mais adiante, no e-ARQ (Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos).

O conjunto de requisitos de um Sigad está determinado pelo e-ARQ, proposto pelo Conarq (2011). Logo, esses sistemas tornam os profissionais da informação definidores de requisitos funcionais. Afinal, como aborda Luz (2010, p. 26), “a tecnologia da informação sempre será um conjunto de requisitos que são atendidos de formas diversas por uma interface de sistema, sendo assim um mecanismo de interação do humano com o computador”. Com isso, o papel de definidor de requisitos no qual atua o profissional da informação num projeto de implantação tecnológica como um site ou um sistema está cada vez mais claro, pois ele:

[...] não precisa programar um sistema, mas deve ter claros os requisitos de seu “negócio” (esta é a linguagem utilizada), e também deve saber o que é arquitetura tecnológica, arquitetura de informação e arquitetura de software, por exemplo. Como um cliente assíduo das áreas de TI (Tecnologia da Informação), ele deverá ser capaz de saber solicitar requisitos, de validar documentações como especificações funcionais e até acompanhar o gerenciamento de projetos. (LUZ, 2010, p. 26).

O Sigad recupera todos os componentes digitais (objeto físico e lógico) que formam o documento arquivístico (objeto conceitual), atendendo aos requisitos de gestão dos documentos a partir do plano de classificação para manter a relação orgânica entre os documentos. Ele gerencia o ciclo de vida dos documentos arquivísticos, que significa classificá-los de acordo com as funções e atividades da organização, e aplica prazos de guarda específicos para cada classe.

Todavia, para preservação em longo prazo de documentos digitais, é necessária uma abordagem de encaminhamento de custódia por meio dos sistemas de gerenciamento de arquivos e de preservação em longo prazo, partindo da gestão documental (que ocorre no Sigad) rumo ao RDC-Arq (repositórios digitais confiáveis), que são os “arquivos permanentes digitais”. Assim, o Sigad já trata na gênese os procedimentos para a preservação de longo prazo dos documentos arquivísticos, garantindo a preservação do vínculo arquivístico e à manutenção da cadeia de custódia. O documento ali gerado e utilizado é encaminhado para a avaliação e posterior arranjo definitivo no repositório digital confiável.

### **REPOSITÓRIOS ARQUIVÍSTICOS DIGITAIS CONFIÁVEIS (RDC-ARQ)**

Com base na resolução número 43 do Conarq (2015), chamada “Diretrizes para a Implementação de Repositórios Arquivísticos Digitais Confiáveis - RDC-Arq”, existem algumas considerações acerca do cenário para garantias de confiança de um sistema informático de arquivo. Esses elementos englobam a responsabilidade pelo repositório, o processamento técnico, a independência de repositórios e as garantias de interoperabilidade. Assim, verifica-se que a interoperabilidade dos documentos e da informação arquivística já é garantida a partir de seu repositório, que será capaz de preservar a sua integridade e fixidez.

A mesma norma estabelece alguns princípios de preservação digital para garantir o acesso em longo prazo aos documentos arquivísticos autênticos, o que implica a adoção de alguns princípios e requisitos de um repositório digital confiável. Deve-se atender a esses requisitos pressupondo que a autenticidade dos documentos arquivísticos digitais está sob ameaça no momento da transmissão em rede (entre pessoas e sistemas) e em relação ao tempo (atualização e/ou substituição de hardware ou software usados para armazenar, processar e comunicar os documentos). Por isso, é necessário considerar que a preservação digital é um processo contínuo, que começa na concepção do documento (CONARQ, 2015).



A autenticidade dos documentos arquivísticos digitais tem por base os procedimentos de sua gestão e preservação, além da confiança tanto no repositório como no órgão responsável pela guarda desses documentos. A elaboração de manuais e os procedimentos de preservação realizados pelo repositório digital apoiam a presunção de autenticidade dos documentos, assim como o registro em metadados das intervenções de preservação apoiam a presunção de autenticidade desses documentos (CONARQ, 2015). Essa operação pode ser acompanhada por uma ontologia que evolui no espaço-tempo (acompanha o processo de avaliação e eliminação) e registra esses tipos de alterações até a definição de um quadro final de arranjo, que represente as funções e atividades de um arquivo.

Os repositórios confiáveis focam nos documentos arquivísticos autênticos, e não genericamente em objetos digitais, pois é necessário arbitrar o que se considera como documento original, uma vez que a preservação digital implica a necessidade de conversão de formatos e atualização de suportes. Outro fator destacado pelo Conarq é quanto a distinção da autenticidade, que é a qualidade de o documento ser verdadeiro, e a autenticação, que é uma declaração dessa qualidade, feita, em dado momento, por uma pessoa autorizada para tal. A autenticidade dos documentos digitais deve ser avaliada e presumida no momento de sua submissão ao repositório, ou seja, na fase de captura do pacote de submissão de informação.

Como vemos, os princípios almejados pelo Conarq para serem aplicados pelo Sinar (Sistema Nacional de Arquivo) representam as condições mínimas para considerar uma plataforma digital confiável e de que forma devem ser tratados os registros. Essas plataformas podem englobar ontologias, representantes do domínio específico que podemos identificar como fundo arquivístico (e suas relações orgânicas multinível), e também a gestão de metadados e registros, do estado de integridade e de processamento da informação declarada como arquivística.

A confiabilidade é também concedida pela relação de segurança e credibilidade, conforme destacam Santos e Flores (2015, p. 205), pois “o repositório digital deve ser o ambiente autêntico para a preservação em longo prazo, dispondo, por exemplo, de ferramentas para a implementação das estratégias de preservação e inserção de padrões de metadados”. Para esse

repositório, todas as ações realizadas em relação aos documentos digitais (migrações, por exemplo) devem ser registradas detalhadamente, criando um histórico de cada objeto digital armazenado, acrescentando confiabilidade aos conteúdos. Ainda, destacam a interoperabilidade como fermenta de gestão e preservação, associada às políticas e ao plano de preservação para servirem de artifícios para a criação de um repositório arquivístico digital confiável em longo prazo (SANTOS; FLORES, 2015, p. 204-205).

Portanto, considera-se que uma ontologia tem importante função no quesito confiabilidade, principalmente ao analisarmos o RDC-Arq preconizado pelo Conarq, no qual podemos verificar seu papel ao longo do tempo nas migrações entre ambientes de custódia documental. O uso de ontologias em repositórios confiáveis pode acompanhar a manutenção da cadeia de custódia de documentos e trazer dados do contexto de criação, garantindo preservação, acesso e confiabilidade documental.

## CADEIA DE CUSTÓDIA E CONTEXTO ARQUIVÍSTICO

A manutenção da autenticidade em toda a cadeia de custódia visa o acesso em longo prazo aos documentos e a consequente preservação da confiança nos ambientes que o geraram, independentemente se for um Sigad (focado em documento corrente e intermediário), um RDC-Arq (focado em arquivo permanente) ou sistema integrado (CONARQ, 2015).

A norma e-Arq Brasil contempla plataformas digitais para a gestão documental, isto é, no uso corrente e intermediário quando apoiam a administração e as atividades das instituições e famílias. Após o término da fase da gestão de documentos, com a alteração da cadeia de custódia, passa-se para a fase de administração de arquivos permanentes e, para isso usa-se os RDC-Arq (Conarq, 2015) que recebe os pacotes de informação para arquivamento (*archival information package* – AIP) que tratam do acondicionamento e armazenamento dos documentos digitais e seus metadados associados.

Para Flores (2014), esse processo deve contemplar os processos descritivos de acervos permanentes realizados nas etapas de arranjo, descrição, digitalização, difusão e acesso de documentos e não mais permitindo ações

ou operações da gestão de documentos nos documentos ali custodiados. O autor também destaca que o documento arquivístico nasce orgânico, mas para assim permanecer precisa estar classificado em um plano de classificação documental (PCD). Podemos entender que, se o documento não permanecer vinculado a outros (o PCD servindo como um seriadador, agregando os documentos em séries), tem-se o risco de perder seu vínculo, o que pode indicar a interrupção da cadeia de custódia.

Quando destinados para guarda permanente, há uma alteração na cadeia de custódia, passando a responsabilidade pela preservação dos documentos dos produtores para a instância de guarda. O processo é migrar de um Sigad para um RDC-Arq. Para o Conarq (2015), os documentos digitais em fase permanente são dependentes de um bom sistema informatizado que apoie o tratamento técnico adequado, incluindo arranjo, descrição e acesso, a fim de assegurar a manutenção da autenticidade e da relação orgânica desses documentos.

Os arquivos devem dispor de repositórios digitais confiáveis para a gestão, a preservação e o acesso de documentos digitais. A preservação dos documentos arquivísticos digitais, nas fases corrente, intermediária e permanente, deve estar associada a um ambiente de gestão e um repositório digital confiável. As normas e-ARQ e a Resolução 39 e 43 para RDC-Arq do conselho apresentam os requisitos para a garantia da integridade da cadeia de custódia, que deve ser ancorada num contexto arquivístico com segurança e que garanta a confiança nos registros (CONARQ, 2015).

O documento arquivístico serve como testemunho da ação que o gerou e, por isso, ele precisa ser confiável e autêntico. Traz, assim, junto aos documentos as informações de contexto e metadados quanto à sua transmissão e custodiadores. Uma ontologia de aplicação digital e arquivística, combinada com a já estabelecida gestão de metadados e os pacotes de submissão, arquivamento e difusão do modelo OAIS podem centralizar as informações de cada documento arquivístico e também trazer informações do contexto que o gerou e a informação arquivística.

## CONTEXTOS REPRESENTADOS EM ONTOLOGIAS

O contexto arquivístico, que espelha as relações orgânicas, são todos os fatores ambientais que combinados determinam como documentos são criados, estruturados, gerenciados e interpretados. Os fatores ambientais que determinam diretamente os conteúdos, formas e estrutura dos registros podem ser diferenciados em contexto de proveniência, contexto administrativo e contexto de uso (THOMASSEN, 2006, p. 10). Esses fatores são determinados pelo contexto sociopolítico, cultural e econômico.

Para Thomassen (2006, p. 10), o contexto de proveniência é o contexto de produção dos documentos, e está referido a como o ente produtor dos documentos é organizado (contexto organizacional), como suas funções estão estruturadas (o contexto funcional) e como seus processos de trabalho são delineados (contexto de procedimentos administrativos). Isso naturalmente é apoiado pelo Plano de Classificação de Arquivo, porém com a evolução tecnológica a representação do contexto pode incorporar também outras formas, como as ontologias.

O contexto de administração para Thomassen (2006), que é o contexto de gestão de documentos e sua conservação, é constituído de todos procedimentos, métodos, conhecimento, meios e documentos com os quais o produtor assegura a disponibilidade e integridade dos arquivos. O autor relaciona o contexto de administração ao chamado contexto documentário, o qual identifica como estrutura do arquivo. Esse contexto pode ser descrito por ontologias e axiomas capazes de representar as principais funções de uma instituição ou empresa, a história arquivística e os principais personagens, por exemplo.

Ademais, o contexto de uso, para Thomassen (2006), consiste nos usuários, suas competências, as perguntas que fazem, e as maneiras pelas quais tentam respondê-las. O ponto destacado pelo autor aproxima a temática aos sistemas que podem suportar tais inferências deste contexto de uso, que são os sistemas semânticos capazes de responder diretamente a inferências diretas. Por fim, o autor ainda lista o contexto sociopolítico, cultural e econômico, e tudo que influencia os fatores ambientais, determinando o conteúdo, forma e estrutura dos documentos arquivísticos. O autor conclui afirmando

que arquivos não podem ser interpretados corretamente sem informação relacionada aos seus contextos. Informação contextual deve, portanto, ser incluída no sistema de informação do qual os arquivos formam parte.

Um projeto italiano que já aplica esses princípios, só que vinculado à Web semântica, é o Repository for Linked Open Archival Data (ReLoad), desenvolvido pelo Archivio Centrale Dello Stato e Istituto per I Beni Culturali dell'Emilia Romagna Regesta, que promove o uso de normas técnicas de descrição de arquivo e a Web Semântica, no qual dispõe que a definição de uma ontologia de descrição arquivística é necessária a fim de testar o potencial da Web de dados para descrições arquivísticas.

Para o projeto, a descrição arquivística visa a representação de uma unidade de descrição por coletar, analisar, organizar e registrar as informações necessárias para identificar, gerir, localizar e explicar o contexto e o material documental, além dos sistemas de armazenamento que os produziram. (Repository for Linked Open Archival Data - RELOAD, 2014). O projeto divulga a ontologia de descrição arquivística (OAD - Ontology for Archival Description), que tem como objetivo a representação formal de contextos resultantes do processo de descrição de arquivo.

Os contextos condicionam a formação dos chamados domínios de conhecimento, ou domínios da realidade, onde os documentos são gerados, estruturados, administrados e interpretados. Cada contexto possui características de funções e atividades que são únicas. A manutenção da cadeia de custódia para estabelecer a autenticidade documental e o contexto de criação descrito e registrado, podendo estar ancorados em ontologias, podem fazer parte de uma política de preservação da informação em longo prazo.

## **ALGUMAS CONSIDERAÇÕES POSSÍVEIS**

A confiança de um repositório arquivístico digital, considerando seu contexto característico e requisitos que qualificam esse tipo de sistema de informação de arquivo, assim como a relação da cadeia de custódia e do contexto arquivístico, são essenciais para garantir, em longo prazo, a confiabilidade do ambiente de gênese documental, pois segundo Thomassen (2006, p. 7)

os “arquivos confiáveis contêm evidência confiável a respeito de decisões tomadas, direitos adquiridos e compromissos assumidos”. Logo, sistemas informatizados de arquivos definem a forma que a informação é preservada e acessada por meio de sistemas ao longo do tempo.

A fidedignidade e autenticidade do documento arquivístico é garantida pela credibilidade em sua cadeia de custódia por meio de sistemas de informação de arquivo confiáveis, tanto na gestão como na guarda permanente. Tal tarefa requer que as informações geradas no documento, além de informações sobre seu contexto de criação sejam preservadas, desde a gênese documental por meio de diferentes técnicas de empacotamento e identificação dos conjuntos documentais.

Na preservação também devemos considerar as duas dimensões da essência informacional, de acordo com Café e Sales (2010), como sendo “a dimensão descritiva, voltada aos elementos relativos à forma dos documentos (como na catalogação descritiva) e a dimensão temática, voltada aos conteúdos informacionais (como na catalogação de assuntos, na classificação, na indexação e na análise documental)” e ambas podem ser consideradas em conjunto para a preservação em longo prazo, representadas por metadados e por ontologias.

As ontologias podem ser usadas para representar padrões de descrição arquivísticas e contribuir para a manutenção de contexto em sistemas de ambientes digitais. A estrutura do contexto se dá em diferentes tipos de sistemas de informação de arquivo e também pode servir como uma maneira de identificar informações orgânicas na rede mundial. O uso de ontologias em repositórios confiáveis pode acompanhar a manutenção da cadeia de custódia de documentos e conservar dados do contexto de criação, garantindo a preservação, acesso e a confiabilidade documental.

Aos arquivistas significa, segundo Quisbert (2006), devido à sua responsabilidade de guarda e preservação, dar acesso a informações usando diferentes gerações de softwares. Todo software é substituível e atualizado, mas a informação deve permanecer original. No caso de arquivos, os documentos arquivísticos são preservados do mesmo modo: por meio de diferentes gerações de softwares. Os registros devem permanecer

como originais, a fim de cumprir suas exigências como prova. Por isso, a interoperabilidade garante o uso e a encontrabilidade dos metadados estruturados dos objetos informacionais. A necessidade de interoperabilizar a informação é básica para os tempos de comunicação em rede.

Parte do processo de organização da informação digital nada mais é do que evidenciar a essência informacional de cada registro, estabelecendo o processo de descrição e de indexação por meio de metadados e assim garantindo sua interoperabilidade e sua preservação. Esse tratamento da informação é objeto de trabalho dos profissionais da informação, incorporando, assim, estas duas dimensões no processo e pode ser apoiado por ontologias aplicadas já na gênese documental e nos diferentes tipos de sistemas de informação arquivísticos.

Uma ontologia é um produto de linguagem e pode representar um domínio ou parcela de uma realidade. A descrição de um contexto, de uma estrutura e de um registro ao longo de um tempo pode ser expressa numa ontologia. Ela, como produto de linguagem, auxilia no processo de perenizar o contexto, importante para certificar a cadeia de custódia e que é basilar para a garantia da autenticidade documental, missão básica do arquivista.

Considera-se, portanto, a ontologia gerada da aplicação do padrão arquivístico descritivo uma ontologia digital arquivística, pois trata-se de uma parcela da realidade disponibilizada na rede e representando um arquivo, objetivando sua interoperabilidade, pois esta informação traz o contexto orgânico da geração informacional intrínseco ao objeto informacional, permitindo assim sua migração entre sistemas o que pode garantir sua preservação em longo prazo. Com esse vínculo orgânico, vê-se que a ontologia carrega em sua estrutura também o esclarecimento acerca desse organismo que acumula a informação arquivística.

O acesso em longo prazo para as instituições representa acumulação de arquivos digitais originados em períodos diferentes, em diversos formatos e tecnologias, mas que devem ser recolhidos e preservados em uma forma fixa e acessível para ser fonte de prova confiável. A autenticidade e o acesso a esse acervo só se tornam possíveis com a manutenção da cadeia de custódia do documento e, nesse aspecto, as ontologias podem

ser elementos constitutivos de sistemas informatizados e da informação arquivística na rede, contribuindo com o registro do contexto arquivístico e acompanhando os pacotes de submissão, arquivamento e difusão preconizados pelo modelo OAIS. É necessário aprofundar esses estudos para assegurar-se, desse modo, que a informação arquivística (documento + contexto) e sua representação semântica em ontologias evoluam na cadeia de custódia, entre plataformas e sistemas, até chegar ao repositório digital confiável (RDC-Arq) para guarda definitiva.





- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, May 2001. Disponível em: <[http://ldc.usb.ve/~yudith/docencia/UCV/ScientificAmerican\\_FeatureArticle\\_TheSemanticWeb\\_May2001.pdf](http://ldc.usb.ve/~yudith/docencia/UCV/ScientificAmerican_FeatureArticle_TheSemanticWeb_May2001.pdf)>. Acesso em: 09 abr. 2017.
- CAFÉ, L.; SALES, R. Organização da informação: conceitos básicos e breve fundamentação teórica. In: ROBREDO, J.; BRÄSCHER, M. (Org.). *Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento*. Brasília: IBICT, 2010. P. 115-129. Disponível em: <<http://repositorio.ibict.br/bitstream/123456789/36/1/eroic.pdf>>. Acesso em: 09 abr. 2017.
- CONSELHO NACIONAL DE ARQUIVOS - CONARQ. *NOBRADÉ*: Norma Brasileira de Descrição Arquivística. Rio de Janeiro: Arquivo Nacional, 2006. Disponível em: <[http://www.conarq.arquivonacional.gov.br/images/publicacoes\\_textos/nobrade.pdf](http://www.conarq.arquivonacional.gov.br/images/publicacoes_textos/nobrade.pdf)>. Acesso em: 09 abr. 2017.
- \_\_\_\_\_. *Resolução nº 39, de 29 de abril de 2014*. 2014. Disponível em: <<http://www.conarq.arquivonacional.gov.br/legislacao/resolucoes-do-conarq/281-resolucao-n-39,-de-29-de-abril-de-2014.html>>. Acesso em: 09 abr. 2017.
- \_\_\_\_\_. *Resolução nº 20, de 16 de julho de 2004*. Disponível em: <<http://www.conarq.arquivonacional.gov.br/legislacao/resolucoes-do-conarq/262-resolucao-n-20,-de-16-de-julho-de-2004.html>>. Acesso em: 09 abr. 2017.
- \_\_\_\_\_. Câmara Técnica de Documentos Eletrônicos. *E-ARQ Brasil*: modelo de requisitos para sistemas informatizados de gestão arquivística de documentos. Rio de Janeiro, 2011. Disponível em: <<http://www.siga.arquivonacional.gov.br/images/publicacoes/e-arq.pdf>>. Acesso em: 09 abr. 2017.
- \_\_\_\_\_. *Glossário*. Rio de Janeiro, 2009. Disponível em: <[http://www.conarq.gov.br/images/ctde/Glossario/2009glossario\\_v5.0\\_final.pdf](http://www.conarq.gov.br/images/ctde/Glossario/2009glossario_v5.0_final.pdf)>. Acesso em: 09 abr. 2017.
- \_\_\_\_\_. *Perguntas mais frequentes*: web site, 2016. Disponível em: <<http://www.conarq.arquivonacional.gov.br/perguntas-mais-frequentes.html>>. Acesso em: 09 abr. 2017.
- \_\_\_\_\_. *Diretrizes para a implementação de repositórios arquivísticos digitais confiáveis*: RDC-Arq 2015. Rio de Janeiro, 2015. Disponível em: <[http://www.conarq.arquivonacional.gov.br/images/publicacoes\\_textos/diretrizes\\_rdc\\_arq.pdf](http://www.conarq.arquivonacional.gov.br/images/publicacoes_textos/diretrizes_rdc_arq.pdf)>. Acesso em: 09 abr. 2017.
- DURANTI, L. The archival bond. *Archives and Museum Informatics*, v. 11, n. 3-4, p. 213-218, 1997.
- FLORES, D. *Ambiente autêntico de preservação e acesso em longo prazo de documentos arquivísticos digitais - RDC-Arq*. Apresentação, UNB, dez. 2014. Disponível em: <<https://pt.slideshare.net/dfloresbr/unb-nov2014-ambiente-autntico-de-preservao-e-acesso-em-longo-prazo-de-documentos-arquivsticos-digitais-rdc-arq-prof-dr-daniel-flores-ufsm>>. Acesso em: 09 abr. 2017.
- GILCHRIST, A. Thesauri, taxonomies and ontologies: an etymological note. *Journal of Documentation*, v. 59, n. 1, p. 7-18, 2003. Disponível em: <<http://www.emeraldinsight.com/doi/abs/10.1108/00220410310457984>>. Acesso em: 09 abr. 2017.

- HODGE, G. *Systems of knowledge organization for digital libraries: beyond traditional authority files*. Washington, DC: Council on Library and Information Resources, 2000. 37 p. Disponível em: <<http://www.clir.org/pubs/reports/reports/pub91/pub91.pdf>>. Acesso em: 09 abr. 2017.
- INNARELLI, H. C. *Gestão da preservação de documentos arquivísticos digitais: proposta de um modelo conceitual*. 2015. 348 f. Tese (Doutorado)- Escola de Comunicações e Artes, Universidade de São Paulo. São Paulo, 2015. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/27/27151/tde-27052015-101628/pt-br.php>>. Acesso em: 09 abr. 2017.
- LOPES, L. C. *A informação e os arquivos: teorias e práticas*. Niterói: Eduff, 1996.
- \_\_\_\_\_. *A nova arquivística na modernização administrativa*. Brasília: Projecto Editorial, 2009.
- LUZ, C. *Arquivologia 2.0: a informação digital humana: excertos de um arquivista 2.0 no mundo digital*. Florianópolis: Bookess, 2010.
- \_\_\_\_\_. *Primitivos digitais: uma abordagem arquivística*. Salvador: 9Bravos, 2015.
- MOREIRO GONZÁLEZ, J. A. *Linguagens documentárias e vocabulários semânticos para a web: elementos conceituais*. Salvador: EDUFBA, 2011.
- NATIONAL INFORMATION STANDARDS ORGANIZATION - NISO. *Understanding metadata*. Bethesda, MD, 2004. P. 1. Disponível em: <<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>. Acesso em: 09 abr. 2017.
- PEDRAZA-JIMÉNEZ, R.; CODINA, L.; ROVIRA, C. Web semántica y ontologías en el procesamiento de la información documental. *El Profesional de la Información*, v. 16, n. 6, p. 569-578, 2007.
- PERICLES PROJECT. *Blog do projeto Pericles: art & media: born digital archive scenario*. Disponível em: <<http://pericles-project.eu/blog/post/born%2C%20digital%2C%20archive%2C%20scenario>>. Acesso em: 09 abr. 2017.
- QUISBERT, H. *A framework for the development of archival information systems*. 2006. Tese (Doutorado)- Luleå Tekniska Universitet, 2006.
- REPOSITORY FOR LINKED OPEN ARCHIVAL DATA- RELOAD. *ReLoad repository for linked open archival data Istituto per i beni artistici culturali e naturali Regione Emilia-Romagna*. 2014. Disponível em: <<http://labs.regesta.com/progettoReload/>>. Acesso em: 09 abr. 2017.
- SANTOS, H. M.; FLORES, D. Repositórios digitais confiáveis para documentos arquivísticos: ponderações sobre a preservação em longo prazo. *Perspectivas em Ciência da Informação*, v. 20, n. 2, p. 198-218, abr./jun. 2015. Disponível em: <<http://dx.doi.org/10.1590/1981-5344/2341>>. Acesso em: 09 abr. 2017.
- THOMASSEM, T. Uma primeira introdução à Arquivologia. *Arquivo & Administração*, v. 5, n. 1, p. 5-16, jan./jun. 2006.

## **Digital archival ontology<sup>2</sup>**

### **ABSTRACT**

*Chapter that addresses how the representation of information in archives can collaborate in the creation of ontologies and how these can support the interoperability and preservation of documents. Archival Description and the interoperability of digital data and how contexts can be represented in Ontologies, respecting the chain of documentary custody, necessary for the guarantee of documentary authenticity in Reliable Digital Repositories - RDC-Arq. As results, it presents the archival digital ontologies that can be used to represent document description patterns of archives and that can contribute to the maintenance of context in systems of digital environments.*

**Keywords:** Archival. Ontologies. Archival description. Reliable digital repositories.

## **Ontología digital archivística<sup>3</sup>**

### **RESUMEN**

*Capítulo que aborda cómo la representación de la información en archivos puede colaborar en la creación de ontologías y cómo éstos pueden apoyar la interoperabilidad y la preservación de los documentos. Trata de la Descripción Archivística y la interoperabilidad de datos digitales y cómo los contextos pueden ser representados en Ontologías, respetando - la cadena de custodia documental, necesaria para garantizar la autenticidad documental en los repositorios digitales confiables - RDC-Arq. Como resultados, presenta las ontologías digitales archivísticas que se pueden utilizar para representar patrones de descripción de documentos de archivo y que pueden contribuir al mantenimiento de contexto en sistemas de ambientes digitales.*

**Palabras clave:** Archivística. Ontologías. Descripción archivística. Repositorios digitales de confianza.

---

<sup>2</sup> Archivist, Master in Information Sciences from ECA-USP in São Paulo/Brazil.

<sup>3</sup> Archivista, maestro en Ciencias de la Información por la ECA-USP en São Paulo/Brasil.

**Dados de pesquisa:**

o que são, impactos do grande volume produzido, como organizá-los e quais preservar

**Maíra Murrieta Costa**

**Murilo Bastos da Cunha**

**Sonia de Assis Boeres**

**RESUMO**

Discute aspectos sobre a ciência colaborativa do século XXI, a internacionalização e a virtualização da ciência que culminaram com a explosão de dados de pesquisa coletados on-line, dando origem ao fenômeno de big data e cyberinfrastructure, também denominado e-Science. Expõe os fatos que trouxeram à tona a ciência de uso intensivo de dados de pesquisa, dentre eles o desenvolvimento tecnológico dos instrumentos de coleta e análise de dados. Explica conceitualmente o que é e-Science e cyberinfrastructure. Ao mesmo tempo, apresenta os termos dados de pesquisa e dados científicos e argumenta que, por se tratar de um tema novo, ainda não há consenso na literatura sobre qual termo deve ser utilizado. Externa o conceito de big data, e como ele trabalha com a utilização de dados de redes sociais para o desenvolvimento de aplicativos. Evidencia a e-Science como uma parte do big data, que lida com dados em larga escala no âmbito científico. Procura definir o termo dado para então apresentar a expressão dados de pesquisa, suas peculiaridades, formas de coleta, tratamento e preservação. Discorre sobre os aspectos de armazenamento e preservação digital dos dados de pesquisa, abordando aspectos de infraestrutura tecnológica. Finaliza o capítulo com a apresentação de reflexões sobre a gestão de dados de pesquisa no Brasil.

**Palavras-Chave:** *Big data*. Ciência orientada a dados. Curadoria de dados. *Cyberinfrastructure*. Dado de pesquisa. *E-Science*. Gestão de dados de pesquisa. Preservação de dados de pesquisa.

## INTRODUÇÃO

O que é a ciência orientada ao uso intensivo de dados? O que é *big data*? O que o *big data* e a denominada *e-Science* têm em comum? De maneira didática, este capítulo procura, por meio de alguns conceitos introdutórios, responder às questões acima mencionadas para o leitor. Também procura elucidar como se dá o uso das tecnologias da informação para a produção de grandes volumes de dados, seja no aspecto comercial (*big data*), ou no universo científico (*e-Science*).

Dentre os objetivos deste capítulo está o de apresentar o que é a *e-Science/ cyberinfraestrutura*, bem como o que são os dados de pesquisa, por vezes denominados dados científicos ou dados de pesquisa. Além disso, argumenta-se que a terminologia na área, por estar em plena ebulição, ainda não está completamente consolidada. Por fim, ele traz reflexões sobre como o Brasil está se preparando para gerenciar os dados de pesquisa gerados nas universidades, nos institutos e nos seus centros de pesquisa, a fim de dar-lhes acesso e garantir-lhes a preservação de longo prazo. Os desafios são muitos, principalmente quando se compreende a variedade de dados produzidos pelas pesquisas, as nuances de cada campo, o comportamento das áreas mais internacionalizadas, como, por exemplo, a energia nuclear, a biodiversidade ou mesmo a área espacial.

O texto pretende trazer contribuições para pesquisadores que objetivem realizar a gestão de seus dados, assim como para profissionais da informação que já começam a se deparar com esse recente desafio. Além disso, traz reflexões para os profissionais de agências de fomento, associações de pesquisa e, ainda, para estudantes e pesquisadores que querem se preparar para essa nova realidade.

Por se tratar de tema extremamente novo, é relevante que o leitor não confunda conceitualmente o que é *e-Science/cyberinfraestrutura* e o que são os dados de pesquisa, algumas vezes também denominados dados científicos.

*E-Science* e *cyberinfrastructure* são termos guarda-chuva que se referem à infraestrutura tecnológica necessária para apoiar a pesquisa científica do século XXI, como, por exemplo, a computação em grid<sup>1</sup> e bancos de dados que suportem petabytes de dados não estruturados, com fluxo constante.

Os dados de pesquisa e/ou científicos, por sua vez, são aqueles coletados em grande volume, por sensores, telescópios, satélites, dentre outros instrumentos e que exigem a infraestrutura tecnológica já comentada para processamento e análise.

Isto posto, cabe ressaltar que ainda não há um consenso na literatura quanto ao uso da expressão *dados científicos* ou *dados de pesquisa*. Os autores Hey e Hey (2006), Bell (2011) e Rodrigues et al. (2010) utilizam *dados científicos*. Por outro lado, Borgman (2015), Sales (2014), Sayão e Sales (2014) utilizam o termo *dados de pesquisa (data scholarship)*.

Também não há consenso quanto ao uso do termo *e-Science* ou *cyberinfrastructure*. Tal situação já havia sido observada na indexação das bases de dados Library and Information Science Abstracts (LISA) e Library and Information Science & Technology Abstracts (LISTA), em estudo bibliométrico sobre a literatura referente ao tema, realizado por Costa e Cunha (2015). Nesse estudo, os autores observaram que o termo *e-Science* foi mais utilizado como indexador, em detrimento do termo *cyberinfrastructure*.

---

<sup>1</sup> A computação em grid é um modelo computacional capaz de alcançar uma alta taxa de processamento de dados dividindo as tarefas entre diversas máquinas. Os grids são compostos por recursos heterogêneos, reunindo desde clusters e supercomputadores, até desktops e dispositivos móveis. Essas máquinas podem estar em uma rede local ou em uma rede de longa distância, o que, por sua vez, forma uma máquina virtual. O processamento de dados pode ser executado no momento em que as máquinas não estão sendo utilizadas pelo usuário, assim evitando o desperdício de processamento da máquina utilizada. De acordo com Buyya (2005) “algumas destas aplicações estão relacionadas ao termo *e-science*, que denota a pesquisa realizada de forma colaborativa em escala global. Este ambiente de *e-science* envolve o compartilhamento de instrumentos científicos, dados distribuídos, visualização remota e interpretação colaborativa de dados e resultados, se adequando perfeitamente às características de uma infraestrutura de computação em grade”. Dentre as iniciativas nacionais de computação em grid destaca-se o LNCC – <http://www.portalgrid.lncc.br>.

A compreensão da ciência com o uso intensivo de dados de pesquisa produzidos em larga escala por sensores especializados, sem dúvida, perpassa o entendimento das origens da ciência moderna. Suas origens, por sua vez, se encontram na Inglaterra do século XVII (ALFONSO-GOLDFARB, 1994; SOLLA PRICE, 1976). Nesse período, a ciência não precisava de grandes justificativas. Quando sofria ataques, sua resposta estava sempre voltada para o futuro, e não para o passado.

A quantidade de publicações em cada campo do conhecimento, após a Segunda Guerra Mundial, cresceu exponencialmente, duplicando a cada dez ou quinze anos (SOLLA PRICE, 1976). Esse fenômeno deu origem à chamada *Big Science*<sup>2</sup>. Solla Price (1976) teorizou sobre a pequena ciência e a grande ciência (*little science* e *big science*), argumentando que a transição de uma para a outra foi gradual. O autor (1976, p. 3) defendeu a ideia de que “se um seguimento suficientemente amplo da ciência for medido de alguma forma razoável, o modo normal de crescimento é exponencial”.

A literatura científica revela que a evolução da ciência está altamente relacionada com o aprimoramento do instrumental tecnológico, que permitiu a realização de observações de diversos fenômenos. Para Bell (2011), as teorias científicas do século XX foram baseadas em dados geralmente disponíveis em cadernos científicos pessoais. Já no início do século XXI, emergiu, de modo crescente, uma questão: os dados oriundos de pesquisas são coletados por meio de sensores especializados, telescópios, satélites e ensaios de laboratórios. Há autores, como Green (2011), Fox e Hendler (2011), que destacam a transformação pela qual passará a pesquisa científica em razão da criação e disponibilidade de grande volume de dados *on-line*.

O desenvolvimento tecnológico de instrumentos de coleta de dados em larga escala e a facilidade de troca de informações sobre determinada pesquisa, por meio da internet, permitiu que cientistas interagissem mais *on-line*. Nas palavras de Castells (2003) a internet tornou-se a espinha dorsal da sociedade contemporânea, “a base tecnológica para a forma organizacional da era da informação – a rede” (CASTELLS, 2003, p. 7).

---

<sup>2</sup> Representa um momento da ciência após as grandes guerras, marcado pelo alto grau de investimentos em C&T.

Esse conjunto de fatores permitiu o desenvolvimento de uma ciência colaborativa *on-line*, que produz dados em larga escala. A ciência colaborativa tem como um de seus marcos iniciais o Projeto Genoma Humano. Outros bons exemplos de colaboração são: a) as iniciativas do European Organization for Nuclear Research (CERN) para descobrir a partícula da vida – uma partícula subatômica que poderia ser o bóson de Higgs; b) o Projeto Netuno do Observatório Oceânico EUA-Canadá; c) o Projeto de Celeste Digital Sloan, dentre outros. Todos esses projetos têm em comum o enorme volume de dados coletados por sensores especializados.

Tapscott e Williams (2007) consideram que o Projeto Genoma Humano representou um divisor de águas. Afinal, as indústrias farmacêuticas pararam com as suas tentativas isoladas de mapear o genoma e passaram a apoiar colaborações abertas (*open-science*). A experiência desse projeto representa o resultado final de forte concentração de esforços públicos e privados em prol da informação genética do ser humano.

Essas iniciativas que envolvem o compartilhamento de recursos e a infraestrutura tecnológica acabam por ser realizados em diferentes instituições que, por sua vez, podem estar em distintos países. Esse fato, associado à facilidade de acesso à informação, contribuiu para uma internacionalização e virtualização da ciência.

A pesquisa colaborativa, produzida por uma equipe multidisciplinar, que coleta grande quantidade de dados, em diversos lugares, fomentou o chamado dilúvio de dados. É nesse cenário que surgem os fenômenos de *big data* e da *e-Science*.

O dilúvio de dados, quando abordado na perspectiva de coleta de dados científicos por sensores, telescópios, radares, satélites, dentre outros, originou a *e-Science*, também denominada *cyberinfrastructure*.

A contemporaneidade do tema traz à tona questões conceituais que ainda não passaram pelo processo de reflexão necessário ao seu amadurecimento. Por exemplo, merece ser comentado que dentre as denominações utilizadas para *e-Science*, também se destacam na literatura os termos ciência orientada por dados (*data-driven science*), computação fortemente orientada a dados (*data-intensive computing*), ciberinfrastructure (*cyberinfrastructure*), ciência com



uso intensivo em dados, quarto paradigma da ciência (*fourth paradigm of science*), dentre outros (ALVARO et al., 2011; CESAR JÚNIOR, 2011; MARCUM, GEORGE, 2010; GRAY, 2007; HEY, TREFETHEN, 2003).

A diferença básica entre o *big data* e a *e-Science* parece estar no fato de a *e-Science* tratar de grande volume de dados no âmbito científico. Poder-se-ia dizer que a *e-Science* é um aspecto particular do *big data* que, por sua vez, está voltado para o âmbito comercial do uso de grandes volumes de dados.

Ao retomar as questões sobre a evolução da ciência, tem-se que nos primórdios da ciência moderna a sociedade se preocupou com o armazenamento dos dados de pesquisa primários, registrados em cadernos pessoais, bem como com a preservação dos resultados das pesquisas, publicados em artigos de periódicos e livros. As atividades de armazenamento e preservação foram exercidas com primazia pelas bibliotecas. O momento atual é propício para a sociedade se preocupar com a gestão dos dados digitais e, conseqüentemente, com a preservação dos dados de pesquisa coletados *on-line*, a fim de garantir o acesso às futuras gerações de pesquisadores.

## O DILÚVIO DE DADOS: *BIG DATA*, *E-SCIENCE*, DADOS DO GOVERNO E DADOS ABERTOS

A pesquisa colaborativa presente no século XXI é descrita como aquela que tem a “capacidade de gerar e armazenar dados em uma escala sem precedentes e muito além da capacidade humana de análise” (CESAR JÚNIOR, 2011). Suas características deram origem aos termos *big data* e *e-Science*.

*Big data* é um termo mais amplo, refere-se a grande volume de dados e ao conjunto de soluções tecnológicas para tratar esses dados digitais. Relaciona-se com a percepção e compreensão de informações analisadas em larga escala, utilizadas geralmente em aplicações comerciais (como, por exemplo, na Amazon para sugerir qual livro o usuário deve comprar), na prospecção de cenários futuros, em campanhas publicitárias, em campanhas de eleição, dentre outros. Para Mayer-Schonberger e Cukier (2013), o *big data* representa “uma nova fonte de valor econômico e informação”. A filosofia do *big data* é deixe os dados falarem.

Mayer-Schonberger e Cukier (2013) exemplificam o conceito de *big data* lembrando o surgimento do vírus H1N1 em 2009. Os autores relatam que pesquisadores da empresa Google analisaram os 50 milhões de termos de busca mais comuns entre os americanos e os compararam com a lista do Centers of Disease Control (CDC). A pesquisa nos termos de busca utilizados no *search engine* Google revelou o local onde o vírus estava se espalhando com mais velocidade que o sistema de informações do CDC.

A literatura indica que a definição de *big data* pode apresentar variações conforme a área de aplicação, por exemplo, na ciência da computação, na análise de finanças e até mesmo no caso de um empresário que está lançando uma ideia para um empreendimento capitalista. Entretanto, há um consenso de que o *big data* se refere à crescente capacidade tecnológica para captar, agregar e processar um volume cada vez maior de dados, que dificilmente seriam processados com as aplicações de tecnologia da informação tradicionais existentes. (BOLLIER, 2010; MAYER-SCHÖBERBER; CUKIER, 2013; UNITED STATES, 2014)

São exemplos desses dados os *posts* das redes sociais, sejam elas Facebook, Twitter ou algum outro aplicativo social. Os dados postados e coletados podem ser pela tecnologia de RFID, dados de localização geográfica de um usuário de aplicativo de mapas da empresa Google – disponibilizados na rede por meio do seu telefone celular ou do aparelho GPS do automóvel, dados de compras *on-line* realizadas com cartão de crédito, dados dos programas de televisão e filmes assistidos na *smart TV* por meio do Netflix ou Youtube, dentre tantos outros exemplos. Esses dados podem ser utilizados em benefício de políticas públicas na área de saúde e educação. Também têm aplicação no conceito de *smart cities* e têm sido frequentemente utilizados por empresas de comércio eletrônico para aprimorar suas estratégias de vendas.

De acordo com Davenport (2014, p. 3-7), o conceito é revolucionário e começou a ganhar força no quarto trimestre de 2010. Para o autor, o *big data* é caracterizado por grande volume de dados desestruturados, provenientes de diversas fontes e com uma necessidade de análise

constante (*streaming data*<sup>3</sup>). Tecnicamente, o autor procura clarificar a diferença entre os conceitos do *big data* e os conceitos do *analytics tradicional*, conforme exposto no quadro 1.

Quadro 1 - Diferença entre os conceitos do *big data* e o *analytics tradicional*

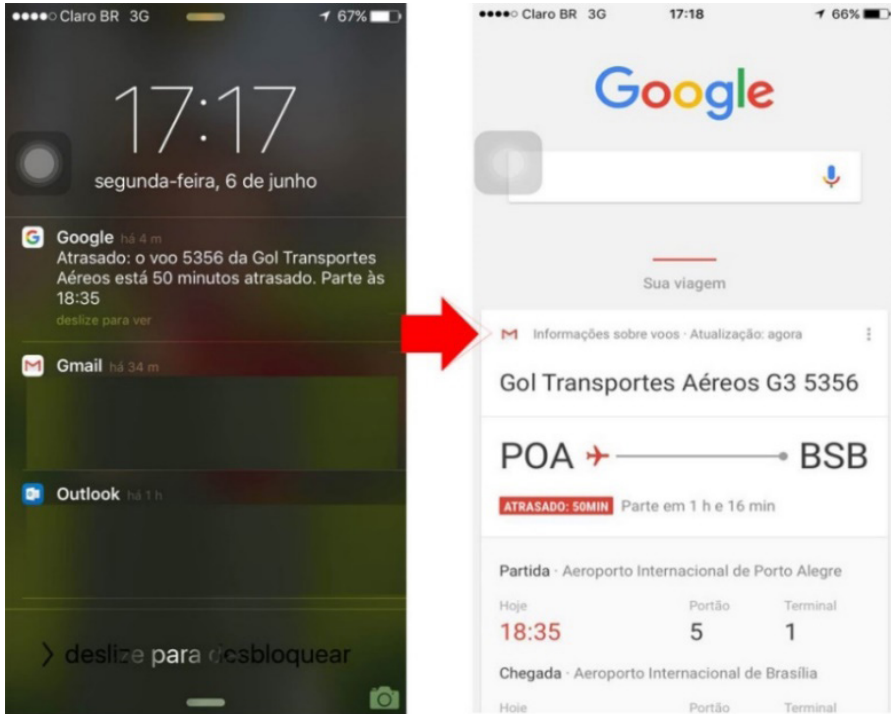
	<b>Big data</b>	<b>Analytics tradicional</b>
<b>Tipos de dados</b>	Formatos não estruturados	Dados formatados em linhas e colunas
<b>Volume de dados</b>	100 terabytes a petabytes	Dezenas de terabytes ou menos
<b>Fluxo de dados</b>	Fluxo constante de dados	Pool estático de dados
<b>Métodos de análise</b>	Aprendizado de máquina	Baseados em hipóteses
<b>Objetivo principal</b>	Produtos baseados em dados	Suporte ao processo decisório

Fonte: Davenport (2014, p. 4)

Desde 2015, pelo menos, já se pode dizer que é uma realidade o fato de que os usuários da empresa Google são avisados por meio dos aplicativos em seu celular sobre atrasos no horário do voo que pegarão, o tempo estimado para chegar em casa devido ao trânsito e até mesmo recebem felicitações quando fazem aniversário. A figura 1 demonstra a utilização de dados pela Google para avisar o atraso de um voo no Brasil.

<sup>3</sup> São dados gerados por inúmeras fontes de dados de forma contínua, que geralmente enviam os registros de dados simultaneamente. Possuem ampla variedade de dados, como, por exemplo, arquivos de log gerados por clientes usando seus aplicativos móveis ou da web, compras de e-commerce, atividade de jogador durante o jogo, informações de redes sociais, pregões financeiros ou serviços geoespaciais, como também telemetria de serviços conectados ou instrumentação em datacenters.

Figura 1 – Big data – a utilização de dados pessoais pela empresa Google.



Fonte: Costa (2016).

Na área da saúde, Davenport apresenta como exemplo o fato de que algoritmos conseguirão prever a possibilidade de que “pessoas tenham um ataque cardíaco” (DAVENPORT, 2014, p. 11) e, conseqüentemente, paguem mais por um plano de saúde. Outros exemplos apresentados referem-se a algoritmos para monitorar a condição financeira das pessoas, bem como seu histórico de ‘comportamento’ e problemas com a polícia local.

Esse poder de uso de dados pessoais, disponíveis na *web* para empresas comerciais e até mesmo para o governo, tem suscitado discussões sobre a privacidade individual que envolvem aspectos éticos tais como – quem permitiu a utilização dos ‘meus dados pessoais’? Qual o limite para a utilização desses dados? Quais as regras para reutilização dos dados?

Outro ponto polêmico, onde o *big data* mostra seu lado perverso, refere-se à utilização de dados pessoais em larga escala em prol da segurança nacional de um país. O exemplo de maior repercussão, até 2016, foi o de Edward Snowden, que revelou o programa de vigilância da National Security Agency (NSA). Snowden revelou que o programa acessava vários tipos de dados<sup>4</sup> de usuários dos serviços de acesso à internet fornecidos pelas empresas AOL, Apple, Facebook, Google, Microsoft, Paltalk, Skype, Yahoo! e YouTube. As empresas negaram que tenham oferecido acesso aos dados para o governo americano.

Se o mundo terá dados suficientes para estabelecer uma tendência geral para delinear o perfil de consumo, perfil de saúde e até mesmo o perfil de atitudes pessoais, o que será do ser humano quando as áreas de inteligência e a polícia de determinado país resolverem utilizar informações para evitar a criminalidade ou ações antiterrorismo? De repente nos vemos no cenário do filme de ficção científica *Minority Report*, dirigido por Steven Spielberg, lançado em 2002, que descreve a Washington de 2054. O filme aborda a redução da criminalidade a partir da possibilidade do crime em questão vir a ser executado. Pessoas são presas por pensarem em cometer um crime!

A ironia é que em apenas 12 anos da data de lançamento, a ficção se tornou realidade, ou seja, muito antes do cenário de 2054 relatado no filme. A única diferença é que a divisão pré-crime do filme determinava suas ações por meio de *um possível futuro* visualizado pelos paranormais e clarividentes *precogs*. Em contrapartida, a realidade de 2014 é a possibilidade de prever comportamentos a partir do dilúvio de dados disponibilizados *on-line* pelo próprio usuário em seu *post* ou *tweet*.

A respeito do tema, Mayer-Schöberber e Cukier (2013, p. 105) alertam sobre o risco da punição com fundamento nas probabilidades oferecidas pela análise do *big data*. Para os autores, “a possibilidade de usar previsões de *big data* sobre pessoas para julgá-las e puni-las antes mesmo que elas ajam, [...] renega a ideia de justiça e livre arbítrio”.

---

<sup>4</sup> São exemplos dos dados acessados: conteúdo de *e-mail*, conversas nos aplicativos de mensagens, vídeos e fotos baixados na internet, conversa telefônica, dados de transações bancárias, dentre outros.

Já no que diz respeito ao grande volume de dados produzidos no âmbito científico, conforme mencionado no primeiro tópico deste capítulo, há vários termos sendo utilizados para tratar as mudanças ocorridas na condução da ciência contemporânea. Dentre esses termos, destacam-se *e-Science* e *cyberinfrastructure*. Aparentemente surgem como termos sinônimos, mas em países com iniciativas diferentes no tratamento do grande volume de dados científicos *on-line*. Por esse motivo, faz-se necessário contextualizar o surgimento de ambos.

Para Jankowski (2007), o termo *cyberinfrastructure* está extremamente relacionado às iniciativas dos cientistas americanos, em 2003, de obter patrocínio da National Science Foundation. Essa iniciativa resultou na publicação do Atkins Report. Nas palavras do relatório: “(...) se a infraestrutura é necessária para a economia industrial, então a *cyberinfrastructure* é necessária para a economia do conhecimento” (ATKINS, 2003, p. 5).

Já o termo *e-Science* surge de iniciativas europeias, especialmente no Reino Unido, onde John Taylor – diretor geral do Escritório de Ciência e Tecnologia do Reino Unido cunha o termo em 1999<sup>5</sup>, durante o lançamento de um programa de financiamento para pesquisas (JANKOWSKI, 2007). Na perspectiva de Hey e Trefethen (2003), Marcum e George (2010) e Gray (2007) a *e-Science* faz referência à coleção de instrumentos e tecnologias necessárias para apoiar a pesquisa científica do século XXI, e amparar o grande volume de dados produzidos que precisam estar em rede, com a característica intrínseca da colaboração e da multidisciplinaridade.

Para Gray<sup>6</sup> (2007), a *e-Science* é o ponto onde a tecnologia da informação encontra os cientistas. Ele explica que a coleta de dados de pesquisa é realizada por instrumentos (satélites, telescópios, sensores) ou é gerada por máquinas de simulação. Os dados capturados, ou obtidos por meio de simulação, são processados por um *software*, que providenciará o armazenamento da

---

<sup>5</sup> Há referências de que o termo *e-science* foi criado no ano 2000, dentre elas Gray (2007).

<sup>6</sup> Jim Gray foi vencedor do Prêmio Turing de 1998. É considerado um dos pioneiros em aplicações e técnicas computacionais para o tratamento de grandes quantidades de dados gerados por cientistas de outras áreas. FONTE: CORDEIRO, D.; BRAGHETTO, K.R.; GOLDMAN, A.; KON, F. Da ciência à e-ciência: paradigmas da descoberta do conhecimento. Revista USP, n. 97, p. 71-80, março 2013.

informação em bancos de dados. Ao comentar que um telescópio é operado por 20 a 50 pessoas e que há milhares de pessoas escrevendo códigos para lidar com a informação coletada pelo instrumento, ele utiliza o campo da astronomia para defender a sua tese.

Na perspectiva de Jankowski (2007, p. 549), *e-Science* é um termo guarda-chuva, utilizado para as iniciativas de computação em *grid*, a colaboração global de pesquisadores e internet baseada em instrumentos. São esses dados, produzidos por esses instrumentos, que precisam passar por um processo de curadoria, armazenamento, divulgação, reutilização e preservação digital. Falaremos sobre as características desses dados no próximo tópico.

Em função do resultado do estudo bibliométrico sobre o tema, realizado por Costa e Cunha (2015), optou-se por utilizar, neste trabalho, o termo *e-Science* em detrimento dos demais. Cabe ressaltar que os autores não identificaram nas bases de dados LISA e LISTA o termo *cyberinfrastructure* indexado na ocasião em que os metadados<sup>7</sup> da base foram analisados. Por fim, os autores inferem que o termo criado no Reino Unido (*e-Science*) tenha ganhado mais adeptos, levando-o a ser um termo indexador.

No que diz respeito ao grande volume de dados produzidos por instituições, no caso do Brasil, não podemos nos esquecer das principais bases de dados do governo, como, por exemplo, as produzidas pelo IBGE, as coletadas pelo Datasus, as recolhidas e geradas pelo Ipea, ou mesmo os dados financeiros do governo federal disponíveis no Siafi<sup>8</sup>. Dentre esses dados, alguns classificam-se como dados abertos e serão abordados a seguir.

Ao analisar a questão dos dados produzidos pelo governo, Sayão e Sales (2015, p. 9) defendem que “embora estes dados não tenham sido originalmente coletados para fins de pesquisa, eles se tornam dados de pesquisa uma vez que tenham sido modificados ou expandidos”. Os autores observam que a

---

<sup>7</sup> Os dados foram coletados nas bases de dados LISA e LISTA entre o período 19/03/2013 a 19/06/2013.

<sup>8</sup> O Sistema Integrado de Administração Financeira do Governo Federal consiste no principal instrumento utilizado para registro, acompanhamento e controle da execução orçamentária, financeira e patrimonial do governo federal. Disponível para acesso em < <http://www.tesouro.fazenda.gov.br/siafi> > .

partir do momento que os dados produzidos pelo governo são utilizados por alguma área de pesquisa e sofrem alguma modificação, eles passam a ser dado de pesquisa.

Na realidade, o que se tem nesse caso é a utilização de dados governamentais abertos, que podem não ter sido produzidos para uma pesquisa acadêmica/científica, mas certamente foram gerados para a avaliação de programas de governo (saúde, educação, indústria e comércio, desenvolvimento tecnológico etc.), ou ainda, dados referentes ao orçamento do governo, como os disponíveis no Siasi ou mesmo no Portal da Transparência. A manipulação desses dados abertos por pesquisas científicas gera dados secundários.

Assim, entende-se que o *big data* é composto pelos diversos tipos de dados que muitas vezes são recombinaados de forma a gerar novas análises e produtos. A figura 2 ilustra o conceito de *big data*, *e-Science*, dados de governo, dados abertos e a relação entre todos eles.

Figura 2 – Aspectos conceituais do *big data*.



Fonte: Costa (2016).



## OS DADOS DE PESQUISA

De acordo com Borgman (2015, p. 4),

a questão não declarada a fazer é: o que são dados” [grifo da autora]. Para a autora, “o único consenso sobre as diferentes definições é que nenhuma definição única será suficiente para definir o termo, uma vez que eles têm muitos tipos de valor, sendo que valor dos dados pode não ser aparente até muito tempo depois dos mesmos terem sido coletados, criados ou mesmo perdidos.

Em continuidade ao assunto, a autora defende que valor dos dados varia muito ao longo do tempo, lugar e contexto. Além disso, enfatiza que ter os dados corretos é geralmente melhor do que ter mais dados. Entretanto, é importante destacar que os dados não têm nenhum valor ou significado quando estão isolados.

Nesse sentido, Borgman (2015, p. 17) nota que conceituar o termo dado não é algo trivial, e aponta que a proposta de Machlup e Mansfield (1983) de dividir em três partes – dado, informação e conhecimento simplifica as relações complexas entre esses conceitos. A autora também recorda a colocação de Meadows (2001) de que “o que nós consideramos ser dados básicos tem sempre um elemento de arbitrariedade nele”.

Davenport, na área de administração, em 2001, apresentou a diferença entre *dado*, *informação* e *conhecimento*. Esses conceitos foram exaustivamente trabalhados e discutidos na famosa pirâmide informacional, predominantemente nas áreas de administração, gestão do conhecimento, inteligência competitiva, dentre outras. Essa discussão em torno do termo *dado*, de certo modo, já demonstra sua complexidade.

Dentro do contexto empresarial, o autor entende que o dado é uma “simples observação sobre o estado do mundo” (DAVENPORT, 2001). Além disso, apresenta como características do dado o fato de ele ser facilmente estruturado, obtido por máquinas, ser frequentemente quantificado e facilmente transferível.

Aproximadamente seis anos depois, com a explosão de dados produzidos e transmitidos por máquinas nos ambientes de pesquisa, Gray (2007, p. 35) propõe uma nova pirâmide informacional, trazendo mais uma vez o *dado*

na base da pirâmide e a literatura no topo. A questão é que Gray tem como projeto deixar todos os dados de pesquisa *on-line*, para assim contribuir com o desenvolvimento da ciência de jeito mais célere. A proposta de Gray (2007) é ilustrada na figura 3.

Figura 3 – Todos os dados científicos *on-line*.



Fonte: Gray (2007, p. 25).

Na perspectiva de Borgman (2015, p. 17), pela sua complexidade, o termo dado, por si só, é digno de um livro. A autora defende que “a questão o que é dado é melhor abordada como quando são dados” fundamentando-se em definições sobre o termo no *Dicionário de Oxford de 1646*, que traz o uso da palavra na teologia, bem como no estudo de Rosenberg, sobre o termo dado, no século XVIII. Além disso, a autora relembra que diversos<sup>9</sup> autores da ciência da informação já discutiram sobre o fato de o dado ser uma forma de informação.

Ao abordar os diferentes tipos de dados, Simberloff et al. (2005, p. 18) fazem uma metáfora com o universo financeiro ao argumentarem que “assim como a moeda na esfera financeira assume diferentes formas, o dado digital também assume diferentes formas no universo de coleção de dados”. Os autores vão além, defendendo que as diferenças dos dados incluem a natureza do mesmo, sua reprodutibilidade, bem como o nível de processamento ao qual o dado é submetido [grifo nosso]. Na percepção de Simberloff et al. (2005, p. 18) “cada uma dessas diferenças traz importantes implicações políticas”.

<sup>9</sup> Blair (2010), Brown e Duguid (2009), Burke (2000, 2012), Day (2001), Ingwersen e Javelin (2005), Liu (2004), Meadows (2001), Buckland (1991) e outros autores.

Simberloff et al. (2005, p. 18/19) argumentam que o dado, quanto à sua natureza, em uma coleção, pode ser diverso. Dentre os exemplos, citam números, imagens, vídeo, arquivos de áudio, *software*, informações sobre a versão de um *software*, equações, animações, algoritmos, ou mesmo, modelos/simulações. Os autores também alertam que os dados podem ser diferenciados em função das suas origens. Nesse aspecto, eles podem ser “observacionais, computacionais ou experimentais”. Além disso, eles enfatizam que a distinção é fundamental para as escolhas a serem feitas sobre o arquivamento e a preservação digital desses dados.

A questão que está em plena ebulição para os gestores de informação, cientistas de dados, assim como para os pesquisadores é: *quais dados devem ser armazenados e por quanto tempo*. Sem sombra de dúvida, a arquivologia traz importantes contribuições nesse aspecto, pois já é tarefa rotineira para esses profissionais a elaboração de Tabelas de Temporalidade Documentais no âmbito de documentos orgânicos de origem primária e secundária.

Pois bem, no que diz respeito a essas questões no âmbito dos dados de pesquisa coletados em larga escala, na percepção de Simberloff et al. (2005, p. 19): “dados de observação, tais como observações diretas de temperatura do oceano em uma data específica, a atitude dos eleitores antes de uma eleição, ou fotografias (...) são registros históricos que não podem ser recoletados”. Logo, para os autores, os dados observacionais são geralmente arquivados indefinidamente. Ou, utilizando a terminologia arquivista – fariam parte do arquivo permanente.

Em continuidade ao assunto, Simberloff et al. (2005, p. 19): argumentam que

(...) um diferente conjunto de considerações aplica-se aos dados computacionais, tais como os resultados da execução de um modelo pelo computador ou por uma simulação. Se a informação detalhada sobre o modelo (incluindo uma descrição completa do *hardware*, *software* e dados de entrada) está disponível, a preservação em um repositório [de dados] de longo prazo pode não ser necessária. Pois, os dados em questão podem ser reproduzidos. Assim, embora os resultados de um modelo possam não necessitar passar pelo processo de preservação, o arquivamento do próprio modelo e de um conjunto robusto de metadados pode ser essencial.

Já no que diz respeito aos dados experimentais, Simberloff et al. (2005, p. 19) defendem que “em princípio os dados de experimentos, que podem ser reproduzidos com precisão, não precisam ser armazenados por tempo indeterminado”. Porém, os autores revelam que, na prática, pode não ser possível reproduzir com precisão todas as condições experimentais, particularmente quando algumas condições e variáveis não podem ser conhecidas. Além disso, há situações em que os custos de reprodução da experiência são proibitivos e nestes casos, em específico, a preservação de longo prazo deve ser garantida para essa categoria de dados. Em síntese, Simberloff et al. (2005) ponderam que as questões de custo e a capacidade de reprodutibilidade são a chave ao considerar-se políticas para a preservação de dados experimentais.

Fox e Harris (2013, p. 10) incluem em sua definição para dados os qualitativos e os estatísticos, conforme descrito a seguir:

(...) inclui, no mínimo, observações digitais, acompanhamento científico, dados de sensores, metadados, cenários e modelos de saída, dados comportamentais observados ou qualitativos, visualizações e dados estatísticos coletados para fins administrativos e comerciais. Dado normalmente é visto como um *input* no processo de pesquisa.

A variedade na tipologia de dados exposta pelos autores anteriormente mencionados (MACHLUP, MANSFIELD, 1983; DAVENPORT, 2001; MEADOWS, 2001; SIMBERLOFF et al., 2005; GRAY, 2007; FOX e HARRIS, 2013) corrobora a percepção de Borgman (2015) sobre a dificuldade de definir o que é dado.

A respeito do assunto, Sayão e Sales (2015, p.7) argumentam que a “noção de dados pode variar consideravelmente entre pesquisadores e, ainda mais, entre áreas do conhecimento”. Para explicar o ponto de vista, os autores teorizam “a constatação de que os dados são gerados para diferentes propósitos, por diferentes comunidades acadêmicas e científicas e por meio de diferentes processos intensifica ainda mais essa percepção de diversidade”.

Em seu guia de pesquisa, Sayão e Sales (2015) propõem como formas de classificação de dados: **a) quanto à sua origem** (observacionais, computacionais e experimentais), **b) quanto à sua natureza** e **c) quanto à fase da sua pesquisa**. A proposição desses autores assemelha-se à proposta

de Simberloff (2005) nos itens **a** e **b**. E ainda trazem uma proposta nova ao proporem uma classificação quanto à fase da pesquisa (dados brutos, crus ou preliminares; dados derivados; dados canônicos ou referenciais).

Do ponto de vista prático, pode-se dizer que o sensor que está implantado nas tartarugas do Projeto Tamar<sup>10</sup>, ou mesmo o sensor que estava implantado no Leão Cecil<sup>11</sup> geram dados de biodiversidade. E quando se trata de informação georreferenciada, pode-se ter a latitude e a longitude indicando a posição de uma espécie de bromélia rara na Floresta Amazônica. Em ambos os casos os dados são armazenados em grandes bancos de dados.

Além dos sistemas já citados, têm-se os dados geodésicos, os dados provenientes da área de energia nuclear, tais como os dados de monitoramento das simulações e das operações de um reator nuclear, ou mesmo os dados sobre mudanças climáticas. Também são dados aqueles produzidos por um laboratório e registrados manualmente em cadernos, como, por exemplo, os dados produzidos pelo Laboratório de Membranas Poliméricas do Instituto de Energia Nuclear.

É preciso explorar semelhanças e diferenças na maneira como os dados são criados, utilizados e compreendidos nas comunidades acadêmicas (BORGMAN, 2015). A partir da observação da autora, é pertinente pensar no caso de um dado, como, por exemplo, da tartaruga do Projeto Tamar coletado pelo sensor. Quem é o autor desse dado? É o líder do projeto de pesquisa? É o pesquisador responsável pelo monitoramento daquela tartaruga específica? Como a comunidade acadêmica entende esses dados? Identificar o autor do dado traz à tona a resposta de como citar o dado. Identificar o modo como o dado foi gerado permite classificá-lo de acordo com as propostas já existentes, como, por exemplo, a de Simberloff et al. (2005). Ao classificar esse dado, por consequência, sabe-se o período de temporalidade dele no

---

<sup>10</sup> Um projeto de 35 anos que representa uma das mais bem-sucedidas experiências de conservação marinha desenvolvidas no Brasil e serve de modelo para outros países.

<sup>11</sup> Leão africano que vivia no Parque Nacional de Hwange localizado no Zimbábue. Era monitorado por cientistas da Universidade de Oxford, no Reino Unido, que estudavam a longevidade e a conservação de leões no Zimbábue. O leão foi morto, aos 13 anos de idade, no ano de 2015, por turista americano em caçada de lazer, abrindo a discussão sobre esta prática e a sobrevivência de animais selvagens.

repositório de dados. Essa cadeia de atividades alimenta o ciclo de gestão de dados de pesquisa que será abordado a seguir, no tópico sobre preservação dos dados. Outras reflexões que impactam na gestão de dados de pesquisa são ilustradas na figura 4.

Figura 4 – Reflexões sobre a gestão de dados de pesquisa.



Fonte: Costa (2016).

Dentre essas reflexões, merecem ser analisadas com cuidado, do ponto de vista da tecnologia da informação, as tecnologias necessárias para armazenar dados oriundos da *e-Science*. É preciso analisar se a área de tecnologia da informação da instituição possui a infraestrutura tecnológica necessária, se há profissionais capacitados para implementar as rotinas de *back-up* de dados e até mesmo restauração, caso seja necessário. Esses aspectos serão discutidos em mais profundidade no tópico sobre a preservação de dados.

## A PRESERVAÇÃO DOS DADOS

Tratar o objeto digital, dentre tantas atividades, implica viabilizar a sua preservação em longo prazo (*long-term-preservation*). Garantir essa preservação, por sua vez, envolve vários aspectos, pois o objeto digital pode sofrer, ao longo de sua vida, várias alterações. Essas alterações não devem impedir que *hardware* e *software*, no futuro, possam transformar os dados armazenados em informação legível para o usuário.

Um marco internacional no contexto da preservação digital (PD) foi a carta submetida a 32ª Sessão da Conferência Geral da Unesco, que ocorreu em Paris em 2003. Na ocasião, a Unesco apresentou, durante a conclusão da conferência, uma carta que ficou conhecida como a *Carta sobre a Preservação Digital*<sup>12</sup>, que trouxe à tona conceitos sobre a importância de estabelecer princípios para a preservação e contínua acessibilidade ao patrimônio digital mundial.

Merece destaque o fato de que a Unesco justificou a importância da PD devido aos “recursos culturais, educacionais, científicos, públicos e administrativos e a informação técnica e médica estarem cada vez mais sendo produzidas, distribuídos e acedidos apenas em formato digital” (UNESCO, 2003). Levando-se em conta que a informação digital está sujeita a suscetibilidades de cair em desuso e decadência física, a carta teve como objetivo levantar um manifesto em prol de um compromisso de longo tempo, para assegurar o contínuo acesso aos conteúdos e à funcionalidade dos objetos digitais.

Segundo a Carta da Unesco (2003, p. 2-4), o desaparecimento do patrimônio, não importa em que forma esteja, é um empobrecimento das nações. Para a entidade, o patrimônio digital são recursos de informação e expressão criativa produzidos, distribuídos, acessados e mantidos em forma digital, e sua preservação é um benefício para a presente e para as futuras gerações.

---

<sup>12</sup> A Carta sobre a Preservação Digital foi publicada em 2003 pela Unesco.

A organização sem fins lucrativos Portico (2015), a maior comunidade mundial de arquivos digitais, que provê serviços de preservação digital, em seu site<sup>13</sup>, apresenta a definição de PD como uma série de políticas e atividades de gestão necessárias para assegurar a duradoura usabilidade, autenticidade, descoberta e acessibilidade dos conteúdos em longo prazo.

Em consonância com a proposta, Rosenthal et al. (2005) argumentam que o objetivo do sistema de preservação digital é que a informação que ele contém permaneça acessível ao usuário por longo do tempo.

Márdero Arellano (2008) trouxe novas contribuições ao apresentar três pontos básicos para a preservação digital, sendo eles: autenticidade, confiabilidade e integridade. O autor resumiu como sendo a autenticidade dos dados a certeza de quem é seu criador (p. 135); já a confiabilidade é ligada à certificação e segurança dos dados digitais (p. 277). A integridade, por sua vez, representa a inteireza/confiabilidade do conteúdo, representada pela não alteração ou modificação para permitir o acesso continuado.

Já na perspectiva da organização Blue Ribbon Task Force (2010, p. 6), a preservação digital possui quatro grandes contextos: o educacional (scholarly discourse), os dados de pesquisa (research data), os conteúdos de internet (collectively produced web content) e o conteúdo comercial e cultural (commercially owned cultural content).

A preservação digital pode estar voltada para a digitalização de documentos em formatos não digitais, ou mesmo voltada para a recuperação de objetos digitais que já se tornaram obsoletos, como, por exemplo, o disquete. O importante é ter a compreensão de que a preservação envolve o uso de técnicas (por exemplo, migração, emulação, espelhamento) e a aplicação de políticas e de gestão de um projeto que tenha como objetivo dar acesso àqueles objetos de modo que eles permaneçam confiáveis, acessíveis e disponíveis para uso ao longo do tempo para quem deles precisar.

---

<sup>13</sup> Site - <http://www.portico.org/>.

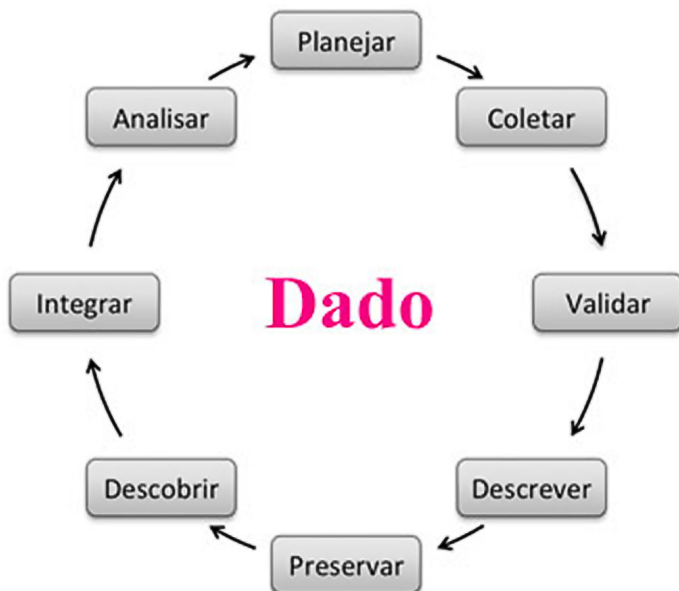


No que diz respeito à preservação de dados brutos de pesquisa oriundos da *e-Science*, esses dados já nascem digitais, são produzidos por equipamentos específicos (satélites, sensores etc.) e, em larga escala, trazem peculiaridades específicas para o seu tratamento e, conseqüentemente, para a preservação. A respeito do assunto, Hey e Hey (2006, p. 515) comentam que:

A fim de explorar os muitos *petabytes* de dados científicos que surgirão a partir dos experimentos científicos de última geração, tais como as simulações em supercomputadores, as redes de sensores e os levantamentos feitos por satélite; os cientistas necessitarão do auxílio de motores de busca especializados e de poderosas ferramentas de mineração de dados. Para criar essas ferramentas, os dados primários deverão ser registrados com os seus metadados relevantes de forma a ter algumas informações quanto à proveniência, o conteúdo e as condições em que os dados foram produzidos. Ao longo dos próximos anos, os cientistas criarão vastos repositórios digitais de dados científicos, o que exigirá serviços de gestão semelhantes aos das bibliotecas digitais mais convencionais, bem como outros serviços específicos de dados.

Nos Estados Unidos, o Projeto Data Observation Network for Earth (DataONE), da National Science Foundation, com sede em Albuquerque, Novo México, tem envidado esforços para a preservação digital de dados de pesquisa. O projeto DataONE tem uma missão ambiciosa: “Fornecer o acesso universal aos dados sobre a vida na Terra e o ambiente que o sustenta [o acesso], bem como as ferramentas que os pesquisadores necessitam para tanto”. Assim, o DataONE tem desenvolvido um framework distribuído e uma ciberinfraestrutura sustentável que atenda às necessidades da ciência aberta. A iniciativa vai ao encontro do movimento de ciência aberta e acata a diretriz do governo americano de aumentar o acesso aos resultados da investigação científica financiada pelo governo federal, conforme ilustra a figura 5.

Figura 5 – Ciclo de vida do dado na perspectiva do pesquisador.



Fonte: DataONE Project<sup>14</sup> (2016).

Na visão do DataONE, o dado tem vida própria. A figura 5 ilustra as etapas de sua criação e utilização. A gestão do dado começa quando o pesquisador ainda está planejando sua etapa de coleta. Os próximos três estágios (coletar, validar, descrever) são a base para o acesso do dado em longo prazo. Enquanto isso, os três últimos representam a descoberta e o uso dos dados.

A filosofia do modelo DataONE parte da pergunta – “se você compartilhar seus dados com um cientista ou colega que não está envolvido com seu projeto de pesquisa, eles estarão aptos a ver sentido nos dados? Será que eles vão ser capazes de usá-los de forma eficaz e adequadamente?”

Nesse sentido, o Projeto DataONE divulgou uma cartilha – *Primer on Data Management: What you Always wanted to Know*. A cartilha descreve algumas práticas de gestão de dados fundamentais, trazendo contribuições para

<sup>14</sup> Melhores Práticas - <https://www.dataone.org/best-practices>.

desenvolver-se um plano de gestão de dados, bem como sugestões para se criar o dado de modo eficaz, organizá-lo, gerenciá-lo, descrevê-lo, preservá-lo e compartilhá-lo, conforme retratado no quadro 2.

Quadro 2 –Visão geral do ciclo de vida DataONE

Atividade	Descrição
Plan	Fase de descrição dos dados que serão compilados, e como os dados serão administrados e tornados acessíveis ao longo da sua vida útil.
Collect	Observações são feitas à mão, ou com sensores, ou outros instrumentos, e os dados são colocados em um formato digital.
Assure	A qualidade dos dados é assegurada por meio de controles e inspeções.
Describe	Os dados são descritos com precisão e são usados os padrões de metadados apropriados.
Preserve	Os dados são submetidos a um arquivamento de longo prazo adequado.
Discover	Dados potencialmente úteis estão localizados e são obtidos junto com as informações relevantes sobre os dados.
Integrate	Dados de fontes diferentes são combinados para formar um conjunto homogêneo de dados que podem ser facilmente analisados.
Analyze	Os dados são analisados.

Fonte: Strasser et al. (2012).

Além do Projeto DataONE, merece ser comentado que na visão de Borgman (2015, p. 20), entre os princípios mais conhecidos para arquivamento de dados tem-se o documento Reference Model for a Open Archival Information System<sup>15</sup> (OAIS). A autora comenta que esse documento apresenta um consenso sobre a prática originada na comunidade de Ciências Espaciais para tratamento e arquivamento de dados. A autora observa que essas orientações também têm sido amplamente adotadas nas ciências e ciências sociais como diretrizes para o arquivamento de dados.

<sup>15</sup> Consultative Committee for Space Data Systems.

A respeito do Modelo OAIS, de acordo com Borgman (2015, p. 22), “ao definir dados, em termos gerais, o modelo usa o termo dados de forma transformadora – conjunto de dados, unidade de dados, formato de dados, banco de dados, objeto de dados, entidade de dados, e assim por diante”. Dentre os exemplos, para a definição de dado tem-se

uma representação de múltiplas interpretações de informações de um modo organizado, adequado à comunicação, compilação ou processamento. Exemplos de dados incluem uma sequência de *bits*, uma tabela de números, os caracteres em uma página, a gravação dos sons feitos por uma pessoa ao falar, ou uma amostra de rocha da Lua coletada durante uma expedição (livre tradução).

Borgman (2015, p. 21) defende que “entre as categorias mais discretas dos dados estão os níveis de processamento definidos pelo Sistema de Informação de Dados sobre a Terra da Nasa”. Nesse sistema, dados com uma origem comum se distinguem pela forma como eles são tratados, conforme demonstra o quadro 3. De acordo com a Nasa (2016)

Produtos de dados da EOSDIS<sup>16</sup> são processados em diversos níveis, variando do **Nível 0** ao **Nível 4**. Os produtos de **Nível 0** são dados brutos na maior resolução do instrumento. Em níveis mais elevados, os dados são convertidos em parâmetros e formatos mais úteis. Todos os instrumentos da EOS devem gerar produtos de **Nível 1**. A maior parte gera produtos de **Nível 2** e **3**, e muitos geram produtos de **Nível 4**.

---

<sup>16</sup> Earth Observing System Data and Information System – em português - Sistema de Informação de Dados sobre a Terra.

Quadro 03 – Níveis de processamento de dados<sup>17</sup>

Nível do Dado	Descrição
<b>Nível 0</b>	Dados de instrumentos e de carga em resolução total, reconstruídos e não processados, com qualquer e todos os artefatos de comunicação removidos (por exemplo, quadros de sincronização, cabeçalhos de comunicação, dados duplicados).  Na maioria dos casos, o Sistema de Operação de Dados EOS (EDOS) fornece esses dados para os Data Centers como conjuntos de dados de produção para processamento pelo Departamento de Ciência de Processamento de Dados ou por um SIPS (Science Investigator-led Processing Systems – Sistema de Processamento liderado por Investigador Científico) para produzir resultados de níveis superiores).
<b>Nível 1A</b>	Dados de instrumentos em resolução total, reconstruídos e não processados, com referência ao tempo e com informações auxiliares anotadas, incluindo coeficientes de calibração geométricos e radiométricos e parâmetros de georeferenciamento (por exemplo, Plataforma Ephemeris), computados e anexados, mas não aplicados ao Nível 0 de dado <sup>18</sup> .
<b>Nível 1B</b>	Dados no Nível 1A que foram processados por unidade do sensor (nem todos os instrumentos possuem dados de origem para o Nível 1B <sup>19</sup> ).
<b>Nível 2</b>	Dados derivados de variáveis geofísicas na mesma resolução e posição que os dados de origem para o Nível 1.
<b>Nível 3</b>	Variáveis mapeadas em grades de escala uniforme do espaço-tempo, geralmente com alguma integridade e consistência.
<b>Nível 4</b>	Modelos derivados ou resultados da análise de dados de níveis inferiores (por exemplo, variáveis derivadas de múltiplas medições).

Fonte: Borgman (2015, p. 22); Feldman (2016), NASA (2016) – Livre tradução com fundamento nas fontes citadas.

<sup>17</sup> O quadro original pode ser visualizado no site da Nasa – <http://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products/>.

<sup>18</sup> De acordo com Feldman (2016) o Nível 1A de dados de arquivo é o preferido pelos cientistas da Nasa, pois se houver mudanças de calibração do sensor, os dados não precisam ser recoletados.

<sup>19</sup> De acordo com Feldman (2016), os dados de nível 1B são dados Nível 1A que tiveram calibrações de instrumentos / radiométricos aplicadas.

Ao se analisar o ciclo de vida dos dados proposto pelo Projeto DataONE, bem como os níveis de processamento de dados do Sistema de Informação de Dados sobre a Terra da Nasa, percebe-se a complexidade do tratamento dos dados coletados pela *e-Science* e, conseqüentemente, sua preservação. Nesse cenário, não é exagero afirmar que a formação do profissional da informação precisa ter pontos revistos à luz das novas necessidades de tratamento da informação pelos usuários. Além disso, a equipe multidisciplinar vai, gradativamente, impondo-se às bibliotecas que desejam enfrentar esse novo desafio.

No âmbito da preservação de dados de pesquisa, para Simberloff et al. (2005, p. 19) “inicialmente, os dados podem ser recolhidos na forma bruta, por exemplo, como um sinal digital gerado por um instrumento ou sensor. Estes dados não processados são frequentemente sujeitos a subseqüentes etapas de refinamento e análise, dependendo dos objetivos da investigação”. Logo, para o autor, o dado pode apresentar uma série de versões. Nesse sentido, Simberloff et al. (2005, p. 19) argumentam que (...) “embora os dados não processados possam não representar a forma mais completa, os dados derivados podem ser mais facilmente utilizáveis por outros [pesquisadores]”. Assim, na visão do autor, a preservação dos dados em múltiplas formas pode ocorrer em muitas circunstâncias. [grifo nosso].

Strasser (2015) defende que enfrentar os desafios inerentes à pesquisa do século XXI exige uma boa gestão de dados de pesquisa (GDP). Para o autor, ao se planejar com cuidado a documentação e preservação dos dados, os objetivos de ter-se dados de pesquisa reprodutíveis e transparentes são muito mais fáceis de alcançar. Além disso, dados bem geridos são mais fáceis de se utilizar e viabilizar sua reutilização, o que se traduz em maior colaboração para pesquisadores e o máximo de retorno do investimento para as agências de fomento.

Bell (2011, p. 13), simplificando o processo de gestão dos dados de pesquisa, argumenta que “a ciência com uso intensivo de dados consiste em três atividades básicas: captura, curadoria e análise”.

Curadoria de dados pode ser entendida como a gestão e a preservação de dados em longo prazo, incluindo-se nesse contexto o fato de agregar valor aos dados digitais, bem como viabilizar a criação de novos dados, de maneira colaborativa, a partir dos já existentes.

Além disso, a atividade de curadoria também pode propiciar a redução dos riscos de obsolescência digital (DIGITAL CURATION CENTER, 2016; HEY, TANSLEY, TOLLE, 2011; ABBOTT, 2008; GIARETTA, 2004).

Para o Digital Curation Centre (2016), a curadoria digital “envolve a manutenção, a preservação e a agregação de valor aos dados da pesquisa digital em toda sua vida útil. A gestão ativa dos dados de pesquisa, por sua vez, reduz as ameaças ao seu valor de pesquisa de longo prazo e reduz o risco de obsolescência digital”. A instituição vai além ao comentar sobre o compartilhamento e a reutilização de dados – os dados curados disponíveis em repositórios digitais de confiança podem ser compartilhados entre a comunidade mais ampla de pesquisa do Reino Unido.

Conway (1997), Sayão (2012) e Sales (2014) salientam que a teoria da curadoria digital traz, no contexto da preservação digital, o diferencial de que a informação não “apenas” deve ser preservada digitalmente, mas também de passar pelo processo de curadoria digital, o que envolveria o tratamento da informação desde a coleta dos dados de pesquisa até o reuso da informação por outros integrantes do fluxo informacional. [grifo nosso]

Em se considerando o já exposto a respeito da curadoria, deve ser comentado que essa atividade não aparece de maneira explícita no Modelo do DataONE, mas pode-se inferir que está implícita nas atividades de descrição e preservação.

## ASPECTOS TECNOLÓGICOS

Os sistemas de gerenciamento de bancos de dados (SGBD) adequados para o processamento de grandes quantidades de dados não são os tradicionais (MySQL, PostgreSQL, Oracle, SQLServer etc.), até mesmo em função do custo de armazenamento, como será demonstrado no quadro 4. A respeito do assunto, Davenport (2014, p. 113) argumenta que “esses dados volumosos não podem ser bem manipulados por um software de banco de dados tradicional ou com servidores individuais (...) dessa forma uma nova geração de software de processamento de dados foi desenvolvida para resolver esse problema”.

A Google lançou o *framework* MapReduce, que distribui o processamento de dados por grande nó de computadores interligados. Na sequência, a Yahoo lançou o Hadoop, uma plataforma de *software* em Java voltada para *clusters* e processamento de grandes massas de dados.

O Hadoop é um projeto de *software* livre desenvolvido pela Apache Software Foundation e por esse motivo às vezes é chamado de Apache Hadoop. A plataforma de computação distribuída do *software* Hadoop é em Java, voltada para *clusters* e processamento de grandes massas de dados. Possui alta escalabilidade, forte confiabilidade e tolerância a falhas. Para Davenport (p. 58), “O Hadoop é um ambiente de armazenamento e processamento de big data unificado em vários servidores”. De acordo com o autor, “um *cluster* Hadoop com cinquenta nós e oitocentos núcleos de processamento é capaz de processar 1 *petabyte* de dados” (DAVENPORT, 2013, p. 59). A respeito da plataforma, Chechia (2013) comenta que os maiores colaboradores para o seu aprimoramento são o Facebook, a Google, o Yahoo e a IBM.

Quadro 4 - Custos de armazenamento de dados

Volume de dados	Custo de armazenamento por 1 ano		
	Banco de dados relacional tradicional	Appliance de dados	Cluster Hadoop
1 Terabyte	US\$ 37 milhões	US\$ 5 milhões	US\$ 2 milhões

Fonte: Davenport (2014, p. 58).

Davenport (2014, p. 111) argumenta que o “big data é mais que apenas grande volume de dados não estruturados. Ele também inclui as tecnologias que possibilitam seu processamento e análise”. No intuito de expor as tecnologias utilizadas no *big data*, o autor elaborou uma síntese, conforme demonstra o quadro 5.



Quadro 5 – Visão geral das tecnologias de *big data*

Tecnologia	Definição
Hadoop	<i>Software</i> de código aberto para o processamento de <i>big data</i> em uma série de servidores paralelos.
MapReduce	Um framework arquitetônico no qual o Hadoop se baseia
Linguagens de Script	Linguagens de programação adequadas ao <i>big data</i> (por exemplo, Python, Pig, Hive).
Aprendizado de Máquina	<i>Software</i> para identificar rapidamente o modelo mais adequado ao conjunto de dados.
Visual Analytics	Apresentação dos resultados analíticos em formatos visuais ou gráficos.
Processamento de Linguagem Natural (PLN)	<i>Software</i> para análise de texto – frequências, sentido etc.
In-memory analytics	Processamento de <i>big data</i> na memória do computador para obter mais velocidade.

Fonte: Davenport (2014, p. 112).

Retomando as colocações de Hey e Hey (2006) de que os cientistas vão precisar de novos mecanismos de buscas, novas ferramentas de mineração de dados especializadas e que criarão repositórios digitais de dados de pesquisa, faz-se necessário analisar esse cenário no contexto brasileiro.

Por todo o exposto, deve-se refletir sobre o sucateamento pelo qual muitas unidades de pesquisa têm passado no Brasil. Se o pesquisador tem dificuldades para obter financiamento para a execução da pesquisa, a realidade da organização à qual o pesquisador está vinculado não é diferente. O orçamento necessário para que as unidades de tecnologia da informação se preparem para tratar o volume de dados imposto pela *e-Science* requer alto investimento. Além disso, como se discutiu anteriormente, novas tecnologias estão sendo adotadas para facilitar o processamento desse grande volume de dados.

Logo, há que se investir tanto na compra de equipamentos e, por vezes, em licenças de *software*, como na capacitação do profissional de tecnologia da informação. Por fim, não se pode negligenciar o fato de os dados serem um ativo institucional que, portanto, precisam passar pelo ciclo de preservação de dados de longo prazo, seja o proposto pelo Modelo OAIS, ou pelo Modelo DataONE, ou mesmo um modelo customizado adequado à realidade da instituição, o que novamente requer investimento em capacitação profissional.

Sem dúvida nenhuma, o importante é dispor de um ambiente em que o dado seja preservado de forma a ser reutilizado em pesquisas futuras. Mas como oferecer esse ambiente sem a infraestrutura tecnológica adequada e sem a capacitação profissional necessária para viabilizar a gestão dos dados de pesquisa? Os próprios pesquisadores precisam entender a importância da preservação de longo prazo do dado produzido pela sua pesquisa, para a partir de aí sensibilizarem o alto nível estratégico das instituições de pesquisa e assim obterem apoio financeiro e institucional para os projetos em questão.

É possível afirmar que enquanto as tecnologias digitais permitem que os dados de pesquisa sejam criados, manipulados, disseminados, recuperados e armazenados com uma facilidade cada vez maior, a preservação de longo prazo dos conjuntos de dados produzidos pela e-Science (datasets) apresentam desafios significativos. A não ser que as estratégias de preservação de dados sejam empregadas tempestivamente, esses dados tendem a se tornar inacessíveis muito rapidamente. O profissional que tiver sob sua responsabilidade a gestão desse dado, seja ele o pesquisador, ou o profissional da informação, ou o cientista de dados, deverá estar atento para selecionar um método de tratamento e preservação que observe a natureza do material (dados) produzido, pois é a natureza desse material que revelará quais aspectos precisam ser conservados.

No presente, muitas das ações ligadas à biblioteca e/ou ao repositório digital envolvem a digitalização do material existente, como, por exemplo, livros e fotografias. Infelizmente, poucos projetos dessas bibliotecas digitais consideram a preservação além da digitalização inicial. A ação de copiar a informação sem alterá-la oferece uma solução de curto prazo para a preservação do acesso aos objetos digitais. Isto faz com que a informação seja armazenada em uma nova mídia antes que a mídia antiga se deteriore.

Porém, em longo prazo, essa simples migração nem sempre funciona. Aqui entra, portanto, a necessidade de implantar-se uma política de preservação digital de longo prazo que leve em consideração todos os outros aspectos relacionados com a informação digital, bem como aspectos relacionados aos dados de pesquisa produzidos em larga escala.

Merece ser ressaltado que as bibliotecas brasileiras ainda têm uma atuação tímida e isolada na preservação de documentos. Pode-se se dizer o mesmo sobre a realidade brasileira no que diz respeito à preservação de dados de pesquisa. Sob esse aspecto, os profissionais da informação precisam estar atentos às mudanças de necessidades de informação do usuário a fim de preencher esse espaço profissional, caso contrário, corre-se o risco de as atividades de curadoria, preservação e outros aspectos, inerentes ao tratamento desse grande volume de dados, serem realizadas pelos especialistas de tecnologia da informação, ou ainda por uma nova categoria de profissional que nasceu para atender às demandas do big data – o cientista de dados.

## OS REPOSITÓRIOS DE DADOS DE PESQUISA NO BRASIL

No Brasil, a percepção de Sales (2014; p.49) é de que: “os pesquisadores, as instituições acadêmicas e as agências de fomento à pesquisa começam a entender que estes dados, se preservados e bem gerenciados, constituem uma excelente fonte de recursos informacionais que podem ser compartilhados e reutilizados como insumo para novas pesquisas”.

Corroborando a percepção de Sales (2014), já é possível constatar que o Brasil possui os seguintes repositórios: a) Repositório de Dados do Programa de Pesquisa de Biodiversidade da Amazônia Ocidental (PPBIO), b) Repositório de Dados do Programa de Pesquisas Ecológicas de Longa Duração (PELD), Portal GEOINFO de infraestrutura de dados espaciais da Embrapa (com 1.081 itens *catalogados*) dentre outros que serão comentados.

Além desses, merece ser comentado o Portal da Biodiversidade<sup>20</sup> (SISBio) lançado pelo Ministério do Meio Ambiente e pelo Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio) em 26 de novembro de 2015. Seu desenvolvimento teve o auxílio de pesquisadores da Escola Politécnica da USP, que conseguiram reunir em uma única interface de busca as informações de bancos de dados mantidos pelo ICMBio e pelo Jardim Botânico do Rio de Janeiro. O Portal oferece buscas textuais e geoespaciais, visualização e *download* de registro de ocorrências de espécies. Além disso, “já conta com mais de um milhão de registros (coordenadas geográficas) de espécies, resultantes da integração de nove bases de dados mantidas pelo ICMBio” (BRASIL. ICMBio, 2015).

No que diz respeito às preocupações do ICMBio com uma política de gestão dos dados de pesquisa, deve ser ressaltada a publicação da Instrução Normativa nº 03, de 01 de setembro de 2014 que, dentre outros, “(...) regulamenta a disponibilização, o acesso e o uso de dados e informações recebidos pelo Instituto de Informações Chico Mendes de Conservação e Biodiversidade por meio do SISBio”. Além dessa, o ICMBio ainda publicou a Instrução Normativa nº 2 de 25 de novembro de 2015 que “Institui a política de dados e informações sobre biodiversidade do Instituto Chico Mendes de Conservação da Biodiversidade e dispõe sobre sua disponibilização, acesso e uso”.

Outros órgãos **iniciaram** o desenvolvimento de seu repositório de dados, dentre eles merece ser citado como exemplo o Instituto de Energia Nuclear que já criou a plataforma CarpeDIEN<sup>21</sup> (Dados e Informações em Engenharia Nuclear) e aos poucos vem alimentando dados de pesquisas de energia nuclear do instituto. Em sua página inicial, já se observa que o sistema oferece a busca pelo autor do dado, por assunto e data de publicação.

A respeito das iniciativas do IEN no âmbito da gestão de dados de pesquisa, merece destaque a publicação do *Guia de Gestão de Dados de Pesquisa*, em novembro de 2015, por Sayão e Sales (2015).

---

<sup>20</sup> O Portal da Biodiversidade representa a interface web do Sistema de Autorização e Informação em Biodiversidade (SISBIO). Disponível para consulta em <<https://portaldabiodiversidade.icmbio.gov.br/portal/>>

<sup>21</sup> Disponível em <<http://carpedien.ien.gov.br/>>.

No que diz respeito à curadoria das informações (altamente técnicas) a serem inseridas nesses repositórios, é importante ressaltar que no Brasil a profissão bibliotecário é de graduação. Portanto, o profissional que tiver como objetivo trabalhar como bibliotecário de dados, ou cientista de dados, terá que se aprofundar no tema que escolher trabalhar, seja ele energia nuclear, infraestrutura para os dados espaciais, ou biodiversidade, por exemplo. Nesse sentido, parece ser prudente que as instituições reflitam sobre o modelo de organização do Centro de Informação Nuclear (CIN) da Comissão Nacional de Energia Nuclear (CNEN) que já na década de 1970 trabalhava com uma equipe multidisciplinar na biblioteca.

## REFLEXÕES SOBRE A GESTÃO DE DADOS DE PESQUISA NO BRASIL

O cenário exposto evidencia a emergência do tema dados de pesquisa e sua complexidade. O Brasil carece de uma política explícita que norteie as ações do Estado em termos de gestão e preservação dos dados de pesquisa (GDP), bem como diretrizes para reutilização dos dados em questão. Também precisa posicionar-se quanto à necessidade de acesso aberto aos dados de pesquisas financiadas por agências de fomento brasileiras.

Inúmeros governos e agências de fomento, segundo Katheleen Shearer (2015, p. 4), começam a elaborar políticas públicas relacionadas com a GDP. Geralmente essas políticas visam ampliar a eficiência da pesquisa, motivar a reutilização de dados, acelerar as ações cooperativas entre pesquisadores e suas entidades. Para a autora:

As jurisdições com os ambientes de políticas mais abrangentes são o Reino Unido, os Estados Unidos, a Austrália e a União Europeia. Detalhes de políticas variam entre regiões, agências e domínios, mas eles também têm uma série de coisas em comum. Os componentes políticos mais frequentes são os requisitos em torno de padrões e metadados, o compartilhamento de dados e a retenção de dados e/ou preservação em longo prazo. Planos de gestão de dados (GDP) são geralmente necessários no contexto dessas políticas, já que obrigam os investigadores a pensarem sobre como eles irão gerenciar seus dados antes do projeto ter se iniciado, um requisito chave para as boas práticas de gestão de dados. As políticas também contêm consistentemente disposições para a proteção da confidencialidade, propriedade intelectual e dados sensíveis (SHEARER, 2015, p. 4).

Existe uma diversidade de pensamentos de como as políticas de GDP são implantadas e monitoradas. O certo é que esta é uma área nova para todos. Também coexiste nesse contexto a criação de ações de GDP mesmo sem a existência de políticas públicas. Certamente, tal cenário tende a mudar rapidamente nos próximos anos, com a participação mais efetiva dos agentes públicos.

É fato que precisamos de políticas públicas, mas também já é fato que algumas instituições<sup>22</sup> de pesquisa, ainda que de forma embrionária, usam sua autonomia para desenvolver políticas locais que atendam aos editais de fomento internacionais, bem como às necessidades de diretrizes quanto ao armazenamento, à preservação e à reutilização de dados. Também já se constata que pesquisadores começam a procurar apoio para realizar a curadoria de seus dados, pois algumas revistas internacionais exigem acesso aos mesmos para publicar determinado artigo.

Em termos de movimentação nas instituições de pesquisa em prol de impulsionar um movimento de apoio à ciência aberta, pode-se dizer que o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) vem liderando no país o movimento de acesso aberto à informação científica desde o início dos anos 2000. Hoje a instituição é considerada referência em projetos voltados ao movimento do acesso livre à informação científica e tecnológica. Exemplo desse compromisso é a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), lançada em 2002, que utiliza a tecnologia de arquivos abertos e integra sistemas de informação de teses e dissertações de instituições de ensino e pesquisa brasileiros. Além da BDTD, o instituto apoia, desde 2009, em conjunto com a Finep, a criação de repositórios institucionais abertos em universidades públicas e centros de pesquisas financiados com recursos públicos. Outra iniciativa de destaque é a formação da Rede de Serviços de Preservação Digital – Cariniana, responsável pela preservação dos periódicos eletrônicos na plataforma OJS/SEER no Brasil, e que possui planos de ampliar o projeto, abrangendo documentos de outros tipos e em variadas mídias.

---

<sup>22</sup> A exemplo, cita-se o ICBio, Museu Emilio Goeldi e Instituto de Energia Nuclear.

A respeito das iniciativas do Ibict de acesso aberto à informação científica, merece ser ressaltado que o instituto, apenas em novembro de 2015, se articulou e organizou o *workshop* **Desafios no Contexto Contemporâneo para promover a nova ciência baseada em dados de pesquisa**. Participaram com destaque no evento o Projeto DataONE, a Rede Nacional de Pesquisa (RNP), a Empresa Brasileira de Pesquisa Agropecuária, unidade de Satélites (Embrapa), o Instituto Nacional de Pesquisa da Amazônia (Inpa), o ICMBio, a Escola Politécnica da Universidade de São Paulo (USP) dentre outros. Durante o evento houve uma cobrança dos participantes para que o Ibict liderasse, perante o MCTI, a elaboração de um conjunto de diretrizes sobre a gestão de dados de pesquisa no Brasil. Desde então, o instituto tem, de maneira ainda tímida, tentado se articular perante os demais *stakeholders* para mapear as necessidades dos pesquisadores quando à gestão de dados, bem como os principais pontos de uma política que norteie a gestão dos dados de pesquisa.

Certamente a elaboração de uma política para a gestão de dados de pesquisa em nível nacional traz à tona a necessidade de se identificar quais tipos de dados de pesquisa o Brasil produz. Além disso, é preciso escolher qual tipo de dado e qual área de conhecimento se priorizará para armazenar e iniciar um programa de gestão de dados de pesquisa. Igualmente, mostra-se necessário refletir sobre a necessidade de produzir um repositório único de acesso aos dados científicos, ou produzir diversos repositórios temáticos, ou ainda, desenvolver, a exemplo da BDTD, uma interface única que busque em diferentes repositórios o dado pesquisado. Nesse caso, quem seria a instituição que apresentaria essa competência?

Uma política de gestão de dados ainda precisa abordar outros aspectos, tais como: a) definir as regras de compartilhamento e reuso dos dados; b) definir o prazo de carência para algumas categorias de dados; c) definir prazo de armazenamento para algumas classes de dados; d) definir padrões de metadados e interoperabilidade destes; e) exigir do pesquisador um plano de gestão de dados quando a pesquisa for fomentada pelo governo; f) definir os requisitos necessários para a implementação do DOI para dados.

A maior parte das indagações anteriores precisa ser refletida em diversos níveis da esfera governamental do um país. Parece coerente afirmar que um conjunto de respostas a essas indagações precisa ser elaborado por meio

de uma política pública, elaborada por um comitê interministerial, com a participação das agências de fomento e, na medida do possível, com a participação de membros da comunidade científica.

Em razão do exposto, é prudente refletir sobre a atual estrutura dos dados de pesquisa no Brasil e como essa gestão tem evoluído. Além disso, é pertinente identificar os principais repositórios de dados científicos do país, bem como os atores estratégicos envolvidos na gestão destes dados. Do ponto de vista estritamente técnico, faz-se necessário apontar soluções para um tratamento adequado dos dados científicos a fim de viabilizar o processo de curadoria, armazenamento, organização, busca, recuperação e difusão dos dados. Caso contrário, os dados coletados podem se tornar inelegíveis ou, o que seria mais drástico, se perder em grande volume de dados, por falta de tratamento e preservação adequados.

## CONSIDERAÇÕES FINAIS

A literatura revela que o bom uso das informações é uma necessidade. Já a história revela que, pelo menos uma vez, o mau uso da informação teve um fim trágico. Tal fato se deu quando Albert Einstein, em 1939, preocupado com as pesquisas sobre fissão nuclear, informou ao então presidente dos Estados Unidos, Franklin Roosevelt, que já existia a possibilidade de criação de uma bomba de alto poder de destruição. Roosevelt, de posse de uma informação altamente estratégica, reuniu um grupo de cientistas, do qual Einstein não fez parte, e deu início ao Projeto Manhattan, que produziu com sucesso um artefato atômico – a *bomba atômica*, posteriormente lançada em Hiroshima e Nagasaki nos dias 6 e 9 de agosto de 1945.

A respeito do assunto, Einstein, que sempre foi um pacifista, um tempo depois afirmou que o maior erro de sua vida foi ter enviado a carta a Roosevelt. Também declarou que não foi sua descoberta a causadora da tragédia da Segunda Guerra Mundial, mas o uso que fizeram dela. Aqui fica a sugestão para que cientistas reflitam sobre a forma ética no uso de seus dados.



É nesse contexto que os profissionais da informação do presente precisam trabalhar para que as informações disponibilizadas em um volume cada vez maior na internet sejam utilizadas para o bem comum, como, por exemplo, para promover o desenvolvimento de pesquisas em saúde, visando promover a cura do câncer, da AIDS, dentre outras mazelas da humanidade.

Os dados são um componente importante da pesquisa econômica e social – eles são a base para a pesquisa e o produto final da pesquisa. A qualidade dos dados de pesquisa e a sua proveniência tornam-se fundamentais no compartilhamento e na subsequente utilização como documento secundário. A gestão eficaz dos dados é uma condição essencial para a geração de dados reutilizáveis de alta qualidade. Os investigadores precisam ter o conhecimento e as habilidades para garantir que os dados que criam e gerenciam podem ser explorados ao máximo, sendo, portanto, um promissor potencial para futuras pesquisas.

Outra questão que não pode deixar de ser abordada é: de que maneira os profissionais da informação podem tratar os dados de pesquisa? Há lugar para eles nas bibliotecas? Os centros de informação como estão hoje talvez não tenham condição física, de infraestrutura e de pessoal capacitado para trabalhar com o conteúdo dos dados de pesquisa. Luce (2010, p. 3) argumenta que, para as bibliotecas, a evolução gradual dos dados de pesquisa (que ele chama de *e-Science*) provoca desafios profundos, e ao mesmo tempo proporciona a elas uma oportunidade de redefinir seus papéis e agregar valor ao seu portfólio de serviços. Hoje os laboratórios são os locais mais “populares” para se desenvolver os dados de pesquisa.

Como um objeto acadêmico, os dados continuam a crescer em importância na comunidade de pesquisa e, paulatinamente, os bibliotecários e demais profissionais da informação têm, crescentemente, maiores responsabilidades na gestão e curadoria de dados. As novas iniciativas de bibliotecas e arquivos incluem as tarefas de ajudar os pesquisadores a encontrar conjuntos de dados para reutilização; localização e hospedagem em repositórios para o arquivamento necessário, consultas sobre o fluxo de trabalho, planos de gestão de dados e melhores práticas para a preservação.

Os profissionais de informação, ao olhar para as opções de fornecer serviços e produtos a essa nova e vibrante área, precisarão saber como tratar e difundir esse tipo de documento até então negligenciado pela área de ciência da informação

A preservação de dados em formato digital é imperativa. As transformações pelas quais a ciência contemporânea vem passando já não permitem que os dados coletados sejam preservados para serem recuperados apenas nesta década, mas sim também em um futuro distante, com a mesma qualidade e confiança em sua autenticidade. A informação digital é um recurso vital na economia do conhecimento, valiosa para a pesquisa, para a educação, para o desenvolvimento tecnológico, assim como para as atividades culturais e o aprimoramento de políticas públicas.



## REFERÊNCIAS

ABBOT, D. *What is digital curation*. Edinburgh, UK: Digital Curation Center, 2008. Disponível em: <<http://www.era.lib.ed.ac.uk/handle/1842/3362>>. Acesso em: 30 de jan. 2013.

ALFONSO-GOLDFARB, A.M. *O que é história da ciência*. São Paulo: Brasiliense, 1994. (Coleção primeiros passos, 286).

ALVARO, E. et al. E-Science librarianship: field undefined. *Issues in Science & Technology Librarianship*, n. 66, p. 28-43, Summer 2011.

ATKINS, D.E. et al. *Revolutionizing science and engineering through cyberinfrastructure*: report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. Washington, DC, 2003. Disponível em: <<http://www.nsf.gov/cise/sci/reports/atkins.pdf>>. Acesso em: 10 abr. 2015.

BELL, G. Prefácio. In: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (Org.). *O quarto paradigma*: descobertas científicas na era da e-Science. São Paulo: Oficina de Textos, 2011. P. 11- 15.

BLUE RIBBON TASK FORCE ON SUSTAINABLE DIGITAL PRESERVATION AND ACCESS - BRTF. *Sustainable economics for a digital planet*: ensuring long-term access for digital preservation: final report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access. 2010. Disponível em: <[http://brtf.sdsc.edu/biblio/BRTF\\_Final\\_Report.pdf](http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf)>. Acesso em: 21 de abr. 2015.

BOLLIER, D. *The promise and peril of big data*. Washington: Aspen Institute, 2010. Disponível em: <[http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The\\_Promise\\_and\\_Peril\\_of\\_Big\\_Data.pdf](http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf)>. Acesso em: 15 maio 2014.

BORGMAN, C.L. *Big data, little data, no data*: scholarship in the networked world. Cambridge, Massachusetts; London, England: MIT Press Books, 2015. 383 p.

CASTELLS, M. *A galáxia da internet*: reflexões sobre a internet, os negócios e a sociedade. Rio de Janeiro: Jorge Zahar, 2003.

CÉSAR JÚNIOR, R.M. Apresentação à edição brasileira. In: HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin (Org.). *O quarto paradigma*: descobertas científicas na era da e-Science. São Paulo: Oficina de Textos, 2011. p. 7- 8.

CHECHIA, D. *Big data*: do conceito a prática. (slides). Palestra apresentada na 7a. Edição da Conferência O Outro Lado - Security BSides São Paulo, 2013. Disponível em: <<http://pt.slideshare.net/daniel.chechia/bigdata-da-teoria-pratica>>. Acesso em: 13 de julho de 2016.

CONWAY, P. *Preservação no universo digital*. Rio de Janeiro: Arquivo Nacional, 1997. (Projeto Conservação Preventiva em Bibliotecas e Arquivos). Disponível em: <[http://www.portal.arquivonacional.gov.br/media/CPBA\\_52\\_Preserva%C3%A7%C3%A3o\\_Universo\\_Digital.pdf](http://www.portal.arquivonacional.gov.br/media/CPBA_52_Preserva%C3%A7%C3%A3o_Universo_Digital.pdf)>. Acesso: 27 abr. 2005.

COSTA, M.M. *Política de gestão de dados científicos: panorama mundial e diretrizes para o Brasil*. [Projeto de Tese]- Programa de Pós-Graduação em Ciência da Informação da Universidade de Brasília, Universidade de Brasília, Brasília, 2016.

DATAONE: best practices. 2016. Disponível em: <<https://www.dataone.org/best-practices>>. Acesso em: 10 abr. 2016.

DAVENPORT, T.H. *Ecologia da informação: por que só a tecnologia não basta para o sucesso na era da informação*. São Paulo: Futura, 2001.

\_\_\_\_\_. *Big data no trabalho: derrubando mitos e descobrindo oportunidades*. Rio de Janeiro: Elsevier, 2014.

DIGITAL CURATION CENTER. *What is digital curation?*. 2016. Disponível em: <<http://www.dcc.ac.uk/digital-curation/what-digital-curation>>. Acesso em: 01 jul. 2016.

ESTADOS UNIDOS. White House. Executive Office of the President. *Big data: seizing opportunities, preserving values*. Washington, 2014. Disponível em: <[http://www.whitehouse.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_may\\_1\\_2014.pdf](http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf)>. Acesso em: 20 maio 2014.

FELDMAN, G.C. *Ocean color web: products definitions*. (2016). Disponível em: <<https://oceancolor.gsfc.nasa.gov/cms/products>>. Acesso em: 22 de setembro de 2017.

FOX, P.; HENDLER, J. E-science semântica: o significado codificado na próxima geração de ciência digitalmente aprimorada. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Org.). *O quarto paradigma: descobertas científicas na era da e-science*. São Paulo: Oficina de Textos, 2011.

\_\_\_\_\_; HARRIS, R. ICSU and the challenges of data information management for international science. *Data Science Journal*, v. 12, n. 10, Feb. 2013. Disponível em: <[https://jstage.jst.go.jp/article/dsj/12/0/12\\_WDS-001/\\_article](https://jstage.jst.go.jp/article/dsj/12/0/12_WDS-001/_article)>. Acesso em: 01 jul. 2016.

GIARETTA, D. *DCC approach to digital curation*. [Draft]. [S. l.]: DCC, 2004. Disponível em: <<http://www.dcc.ac.uk/sites/default/files/documents/DCCApproachtoDigitalCuration-20040827.pdf>>. Acesso em: 02 abr. 2014.

GRAY, J. *E-Science: a transformed scientific method*. Palestra apresentada no Conselho Nacional de Pesquisa dos Estados Unidos (NRC-CSTB), Mountain View, Califórnia, 11 janeiro de 2007. Disponível em: <[http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB\\_eScience.ppt](http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB_eScience.ppt)>. Acesso em: 30 ago. 2012.

GREEN, D. Infraestrutura-científica: introdução. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Org.). *O quarto paradigma: descobertas científicas na era da e-science*. São Paulo: Oficina de Textos, 2011. P. 129-130.

HEY, T.; HEY, J. E-Science and its implications for the library community. *Library Hi Tech*, v. 24, n. 4, 2006. P. 515-528.

\_\_\_\_\_; TANSLEY, S.; TOLLE, K. (Org.). *O quarto paradigma: descobertas científicas na era da e-Science*. São Paulo: Oficina de Textos, 2011. 261 p.

\_\_\_\_\_; TREFETHEN, A. E-Science and its implications. *Philosophical Transactions of the Royal Society (A)*, v. 361, p. 1809-1825, June 2003.

JANKOWSKI, N.W. Exploring e-science: an introduction. *Journal of Computer-Mediated Communication*, v. 12, p. 549-562, 2007.

INSTITUTO CHICO MENDES DE CONSERVAÇÃO DA BIODIVERSIDADE. *Portal da biodiversidade*. Brasília, 2015. Folder.

\_\_\_\_\_. *Instrução normativa nº 03, de 01 de setembro de 2014*. Fixa as normas para a utilização do Sistema de Autorização e Informação em Biodiversidade – SISBIO, na forma das diretrizes e condições previstas nesta instrução normativa e, regulamenta a disponibilização, o acesso e o uso de dados e informações recebidos pelo Instituto Chico Mendes de Conservação da Biodiversidade por meio do SISBIO. (Processo nº 02070.001067/2013-96). Brasília, 2014.

\_\_\_\_\_. *Instrução Normativa nº 2 de 25 de novembro de 2015*. Institui a política de dados e informações sobre biodiversidade do Instituto Chico Mendes de Conservação da Biodiversidade e dispõe sobre sua disponibilização, acesso e uso. Brasília, 2012.

MACHLUP, F.; MANSFIELD, U. (Org.). *The study of information: interdisciplinary messages*. New York: John Wiley & Sons, 1983.

MARCUM, D.B.; GEORGE, G. (Ed.). *The data deluge: can libraries cope with e-science?*. Santa Barbara, California: Libraries Unlimited, 2010.

MÁRDERO ARELLANO, M.A. *Crítérios para a preservação digital da informação científica* 2008. 356 f. Tese (Doutorado em Ciência da Informação)- Universidade de Brasília, Brasília, 2008. Disponível em: <<http://repositorio.unb.br/handle/10482/1518>>. Acesso em: 01 jun. 2016.

MAYER-SCHÖNBERGER; V.; CUKIER, K. *Big data: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana*. Rio de Janeiro, Elsevier, 2013. 256 p.

MEADOWS, J. *Understanding information*. München: KG Saur, 2001. 112 p.

NASA. *NASA Science Earth. Data Processing Level*. (2016). Disponível em: <<https://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products/>>. Acesso em: 13 jun. 2016.

PORTICO. *About us*. 2015. Disponível em: <<http://www.portico.org/digital-preservation/about-us>>. Acesso em: 11 jun. 2015.

RODRIGUES, E. et al. *Os repositórios de dados científicos: estado da arte*. Portugal: Universidade do Minho; Universidade do Porto, 2010. (Projecto RCAAP D24 – Relatório).

ROSENTHAL, D.S. et al. Requirements for digital preservation systems: a bottom-up approach. *D-Lib Magazine*, v. 11, n. 11, Nov. 2005. Disponível em: <<http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>>. Acesso em: 28 jul. 2012.

SALES, L.F. *Integração semântica de publicações científicas e dados de pesquisa: proposta de modelo de publicação ampliada para a área de ciências nucleares*. 265 f. Tese (Doutorado em Ciência da Informação)-Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014.

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

\_\_\_\_\_. Dados abertos de pesquisa: ampliando o conceito de acesso livre. *Revista Eletrônica de Comunicação Informação e Inovação em Saúde*, v. 8, n. 2, jun. 2014.

\_\_\_\_\_. *Gestão de dados de pesquisa para bibliotecários e pesquisadores*. Rio de Janeiro: CNEN: IEN, 2015. 90 p. Versão Preliminar.

SHEARER, K. *Comprehensive brief on research data management policies*. 2015. Disponível em: <[www.science.gc.ca/1E116DB8-E7F3-4B6F-BB44-83342BAAA030/Comprehensive%20Brief%20on%20Research%20Data%20Management%20Policies.pdf](http://www.science.gc.ca/1E116DB8-E7F3-4B6F-BB44-83342BAAA030/Comprehensive%20Brief%20on%20Research%20Data%20Management%20Policies.pdf)>. Acesso em: 05 jul. 2016.

SIMBERLOFF, D. et al. *Long-lived digital data collections: enabling research and education in the 21st century*. [S.l.]: National Science Foundation, 2005.

SOLLA-PRICE, D.J. de. *O desenvolvimento da ciência: análise histórica, filosófica, sociológica e econômica*. Rio de Janeiro: Livros Técnicos e Científicos, 1976.

STRASSER, C. *Research data management*. Baltimore: National Information Standards Organization, 2015. (NISO Primer Series).

\_\_\_\_\_. et al. *Primer on data management: what you always wanted to know*. DataONE best practices primer. [Tennessee]: DataONE, 2012. Disponível em: <[https://www.dataone.org/sites/all/documents/DataONE\\_BP\\_Primer\\_020212.pdf](https://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf)>. Acesso em: 10 jun. 2016.

TAPSCOTT, D.; WILLIAMS, A.D. *Wikinomics: como a colaboração em massa pode mudar o seu negócio*. Rio de Janeiro: Nova Fronteira, 2007.

UNESCO. *Carta para la preservación del patrimonio digital*. [Paris?]: UNESCO, 2003. Disponível em: <[http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/charter\\_preservation\\_digital\\_heritage\\_es.pdf](http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/charter_preservation_digital_heritage_es.pdf)>. Acesso em: 23 jul. 2004.

## ***Research data: what are the impacts of large volume produced, how to organize them and what to preserve***

### **ABSTRACT**

*It discusses aspects of the collaborative science of the XXI Century, science internationalization and virtualization that led to the explosion of research data collected online, giving rise to the phenomenon of big data and cyberinfrastructure, also called e-Science. It exposes the facts which brought to light the science of intensive use of research data, including the technological development of instruments for data collection and analysis. It conceptually explains what e-Science and cyberinfrastructure are. At the same time, it presents the expressions “research data” and “scientific data” and argues that because it is a new subject, there is still no consensus in the literature about which term should be used. It outsiders the concept of big data, and how it works with the use of social networking data to application development. It highlights the e-Science as a part of big data, which handles large-scale data in the scientific realm. It seeks to define the term “data” to then display the term “research data”, its peculiarities, ways of collection, treatment and preservation. It discusses storage and digital preservation aspects of research data, addressing aspects of technology infrastructure. It concludes the chapter with the presentation of reflections on research data management in Brazil.*

**Keywords:** *Big data. Data-driven science. Data curation. Cyberinfrastructure. Research data. E-Science. Research data management. Digital preservation of research data.*

## ***Datos de investigación: qué son, impactos del gran volumen producido, cómo organizarlos y cuáles preservar***

### **RESUMEN**

*En el caso de la ciencia colaborativa del siglo XXI, la internacionalización y la virtualización de la ciencia que culminaron con la explosión de datos de investigación recogidos en línea, dando origen al fenómeno de big data y cyberinfraestructure, también denominado e-Science. Expone los hechos que trajeron a la luz la ciencia de uso intensivo de datos de investigación, entre ellos el desarrollo tecnológico de los instrumentos de recolección y análisis de datos. Explica conceptualmente qué es e-Science y cyberinfraestructure. Al mismo tiempo, presenta los términos datos de investigación y datos científicos y argumenta que, por tratarse de un tema nuevo, aún no hay consenso en la literatura sobre qué término debe ser utilizado. Externó el concepto de big data, y cómo trabaja con el uso de datos de redes sociales para el desarrollo de aplicaciones. Evidencia a e-Science como una parte de la gran fecha, que se ocupa de datos a gran escala en el ámbito científico. Se busca definir el término dado para entonces presentar la expresión datos de investigación, sus peculiaridades, formas de recolección, tratamiento y preservación. Discuta sobre los aspectos de almacenamiento y preservación digital de los datos de investigación, abordando aspectos de infraestructura tecnológica. Finaliza el capítulo con la presentación de reflexiones sobre la gestión de datos de investigación en Brasil.*

**Palabras clave:** *Big data. Ciencia orientada a datos. Curaduría de datos. Cyberinfraestructure. Datos de investigación. E-Science. Gestión de datos de investigación. Preservación de datos de investigación.*



# GESTÃO DOS DADOS DE PESQUISA:

## OPORTUNIDADES E DESAFIOS

Anaíza Caminha Gaspar

Lillian Alvares

Maria de Nazaré Freitas Pereira

### RESUMO

Analisa casos de sucesso e lições aprendidas disponíveis na literatura sobre compartilhamento, reuso e preservação de dados abertos da ciência. O acesso aos dados abertos da ciência abre questões que necessitam de pesquisa, como por exemplo, novas concepções de bancos de dados onde pesquisadores avaliam a qualidade, procedência e confiabilidade dos dados. Neste contexto, torna-se decisivo o estabelecimento de infraestrutura eletrônica capaz de fazer face aos imensos desafios existentes. A complexidade do tema compreende igualmente a necessidade de desenvolvimento de abordagens que proporcionem a descoberta de dados em todas as disciplinas. Cabe aos provedores de infraestrutura de dados científicos e entidades de financiamento encorajar o fácil acesso aos recursos de informação, promovendo capacitação profissional e mecanismos de reconhecimento que incentivem o compartilhamento de dados. Diferentemente das publicações científicas, os dados de pesquisas exigem cuidados especiais para a articulação de uma agenda comum, envolvendo entidades financiadoras, pesquisadores e instituições produtoras de dados. O espaço de curadoria, incluindo ferramentas para descoberta dos dados, novas métricas de citações e políticas de compartilhamento, vem sendo pouco a pouco ocupado pelos centros e bibliotecas que desenvolvem repositórios e sistemas de informação coordenados para registro e divulgação de dados.

**Palavras-chave:** Ciência aberta. Repositórios de dados de pesquisa. Cientista de dados. Bibliotecários de dados. Infraestrutura eletrônica e serviços de informação.

## INTRODUÇÃO

Este trabalho objetiva apresentar uma visão geral dos desafios e oportunidades para uso e reutilização dos dados da pesquisa científica no âmbito da ciência aberta, conforme tendências e lições aprendidas em melhores práticas encontradas na literatura que se encontra reunida em uma bibliografia comentada com cerca de 700 referências bibliográficas (INSTITUTO BRASILEIRO..., 2014).

De acordo com a OECD, dados científicos são “registros factuais usados como fontes primárias na investigação científica, e que são geralmente aceitos na comunidade científica como necessários para validar os resultados de investigação” (ORGANISATION FOR ECONOMIC..., 2007), assumindo várias formas: texto, números, imagens fixas, imagens em movimento, etc. e dimensões. Eles representam “(...) um recurso valioso a longo prazo e (...) compartilhá-los e torná-los publicamente disponíveis é essencial para que o seu valor potencial possa ser realizado” (SWAN, 2008 apud SCHROEDER; DAVID; DEN BESTEN, 2008).

A reutilização dos dados de pesquisa é uma forma de compartilhamento que se insere na gênese da ciência aberta como o compartilhamento ideal de conhecimentos, recursos educacionais e informacionais que são viabilizados por poderosas infraestruturas eletrônicas, transpondo fronteiras institucionais de disciplinas científicas e de nações. Supõe-se que isso pode ser um passo importante para estimular, desde cedo, professores e alunos em suas carreiras de pesquisadores.

Para entendimento de como essas iniciativas se organizam, importa considerar vários aspectos da “abertura” em pesquisa nos quais as práticas contemporâneas de ciência eletrônica caracterizam em maior ou menor grau o ideal da “ciência aberta” (SCHROEDER; DAVID; DEN BESTEN, 2008). No momento atual, os dados abertos da pesquisa científica ocupam posição central nessa discussão, reconhecidos como insumo e produto de valor essencial para a constituição de um novo setor de atividades econômicas e sociais, amplamente discutido no marco referencial da ciência aberta.

Por dados abertos da pesquisa científica entendem-se “aqueles que estão disponíveis gratuitamente na internet pública, permitindo a qualquer usuário baixar, copiar, analisar, reprocessar, fazer a captura por *software* ou utilizá-los para qualquer outra finalidade, sem outras barreiras financeiras, legais ou técnicas além daquelas que dizem respeito à própria internet. Para este fim os dados que dão origem às publicações científicas devem ser explicitamente colocados em domínio público” (MURRAY-RUST; NEYLON; POLLOCK; WILBANKS, 2010).

Nesse sentido, recomendações da Comissão Geral da União Europeia incentivam o desenvolvimento de infraestruturas colaborativas, acordos e alianças para publicização dos dados de pesquisa, ressaltando a necessidade de encontrar o equilíbrio certo entre a regulamentação e a invenção, controle e liberdade, desempenho e custo, público e privado, internacional e local (MORAIS PIRES; HUDSON, 2012).

Esses são os principais desafios a serem enfrentados pelas agências de financiamento e provedores de dados governamentais, para criar condições de apropriação dos resultados da ciência por parte da sociedade em geral, sem restrições quando o financiamento da pesquisa é realizado com recursos públicos. Normas e práticas de abertura são apoiadas pelos pesquisadores por serem vitais para o trabalho das comunidades científicas modernas, ainda que restrições técnicas e institucionais delimitem os contornos dessa abertura (ALBAGLI; APPEL; MACIEL, 2013), por causa da natureza social da ciência. Mas a mesma natureza e a condição de modernidade da ciência colocam “(...) as tecnologias de informação e comunicação (TICs) em uma condição determinante para o estabelecimento de novas modalidades de socialização entre pesquisadores” (SALES; SAYÃO, 2012).

A apresentação das melhores práticas e das lições delas decorrentes é antecedida da ilustração de dois aspectos que moldam as práticas de compartilhamento de dados: a natureza dos dados científicos e a e-infraestrutura onde circulam.

## NATUREZA DOS DADOS CIENTÍFICOS E COMPARTILHAMENTO POR ÁREAS DE CONHECIMENTO

O mundo se amplia e se particulariza no conhecimento local que se torna cada vez mais global, reunindo dados de pesquisa geridos em diferentes centros de dados por pesquisadores individuais ao redor do mundo, além-fronteiras geográficas e institucionais. Pesquisas realizadas em áreas das ciências da terra e da vida são um bom exemplo disso.

A ordem implícita no funcionamento da natureza mostra a terra como uma entidade viva de sistemas de vida interligados, nos quais as espécies que habitam o planeta se sujeitam aos mesmos fenômenos físicos e naturais. Nesse contexto o intercâmbio de dados científicos é crucial para a vida de todos os habitantes do planeta, mas isso tem que ser realizado discriminando a origem dos dados: dados de observação (registros históricos, como medidas de precipitação, que não podem ser reproduzidos e necessitam de preservação permanente), dados computacionais (resultantes de simulações, que teoricamente podem ser reproduzidos se for preservada informação sobre o modelo e a sua execução) e dados experimentais (que também não são facilmente reproduzíveis) (NATIONAL SCIENCE FOUNDATION, s/d).

### CIÊNCIAS DA TERRA

Dados científicos procedentes de monitoramento sobre meio ambiente e ecologia, representantes das ciências da terra, transcendem fronteiras geográficas, unem diversos países para o compartilhamento de dados de pesquisa, e podem juntar para seu entendimento disciplinas como população e ecologia comunitária, ciência dos ecossistemas, biologia evolutiva, sistemática filogenética, ciências sociais e econômicas, ecologia urbana, oceanografia, matemática, ciência da computação e educação científica. Exemplo disso vem da The Long Term Ecological Research (LTER) Network<sup>1</sup>, rede criada pela National Science Foundation (NSF), a partir de ação liderada por cientistas, para realizar pesquisas sobre questões ecológicas que podem durar décadas e abranger grandes áreas geográficas.

---

<sup>1</sup>Disponível em: <<https://lternet.edu/network>> Acesso em: 24 set. 2014.

Tem participação de países europeus (Long-Term Ecosystem, s/d) e por mais de três décadas vem gerando pesquisa científica local que levou a resultados importantes em escalas regional e continental.

O princípio de compartilhamento integral e aberto de dados científicos (*data sharing*) foi discutido publicamente, na década de 90, no âmbito do Sistema de Dados e Informação sobre Mudança Global (*Global Change Data and Information System*, GCDIS). Naquela ocasião foi declarado que programas internacionais de pesquisa, abordando o tema de mudança global no monitoramento ambiental, dependem crucialmente do princípio de intercâmbio integral e aberto sem restrições, isto é, dados e informações devem estar disponíveis sem restrições em bases não discriminatórias, a não ser seu custo de produção e distribuição.

## CIÊNCIAS DA VIDA

Outro exemplo vem de uma disciplina que integra as ciências da vida. O Instituto Europeu de Bioinformática (European Bioinformatics Institute - EBI)<sup>2</sup> integra o Laboratório Europeu de Biologia Molecular (European Molecular Biology Laboratory, EMBL, Reino Unido). O EBI disponibiliza gratuitamente dados de experimentos de ciências da vida, realiza pesquisa básica em biologia computacional e oferece um programa extensivo de formação do usuário, apoiando pesquisadores na academia e na indústria.

Em 1970, o EMBL enfrentou o desafio de gerenciamento de dados de pesquisa e, em 1981, estabeleceu uma das primeiras bibliotecas de dados do mundo para dados de sequências de nucleotídeos. Iniciou extraíndo dados de revistas, mas com a aceleração de métodos para extração de DNA e a crescente eficiência das metodologias, o foco mudou para atrair diretamente o envio de dados por parte dos pesquisadores. Nos EUA, com o GenBank em 1986, iniciou-se um acordo de cooperação do qual surgiu a colaboração internacional com a base de dados de sequências de nucleotídeos (International Nucleotide Sequence Database Collaboration - INSDC). Essa cooperação bem-sucedida foi expandida para incluir o DDBJ no Japão, que somou-se ao GenBank, dos Estados Unidos, e ao Nucleotide Sequence do EMBL-EBI do Reino Unido.

---

<sup>2</sup> The European Bioinformatics Institute. Disponível em: <<http://www.ebi.ac.uk/>> Acesso em: 8 jul. 2014.

Os três bancos de dados sincronizam seus dados diariamente, tornando assim mais fácil para os pesquisadores acessar dados atualizadas em todo o mundo. Como os dados da pesquisa publicados são de domínio público, eles podem ser agrupados e vendidos para usuários comerciais. A decisão de colocar os dados de domínio público foi impulsionada para facilitar o acesso e reutilização da informação, considerando as necessidades das comunidades de ciências da vida para progredirem em seu campo de atuação. As atividades do EMBL-EBI estão estimulando o compartilhamento de dados entre os diferentes setores comerciais. No entanto, questões como o patenteamento ainda são consideradas restrições.

O último exemplo que ilustra como a disciplina científica que dá origem aos dados de pesquisa molda as práticas de organização e de compartilhamento vem das ciências sociais.

## CIÊNCIAS SOCIAIS

O Arquivo de Dados do Reino Unido (*UK Data Archive*) contém a maior coleção de dados digitais de pesquisas econômicas e sociais nesse âmbito. Ele adquire, faz curadoria e permite o acesso a bases de dados, fornecendo o suporte e infraestrutura técnica para a comunidade de pesquisa cumprir as exigências das políticas definidas pelos órgãos de financiamento e conselhos de pesquisa. Atualmente abriga vários milhares de conjuntos de dados, lidando com dados de pesquisa acadêmica, dados governamentais e dados comerciais. Os especialistas dessa instituição lidam diretamente com dados de pesquisa acadêmica produzidos por indivíduos e grupos de pesquisa no domínio amplo das ciências humanas e sociais. A necessidade de gerenciamento de dados de pesquisa nesse campo pode ser muito especializada quando existem informações pessoais nos dados, sobretudo quando se trata de dados qualitativos, como dados de entrevistas, que exigem tratamento rigoroso. O pré-processamento é necessário para tornar o conjunto de dados de pesquisa adequado para compartilhamento e para a publicação, observando detalhes do anonimato pessoal e garantindo que existe o consentimento para compartilhamento de dados ou publicação. O arquivo é em grande parte financiado pelo Economic and Social Research Council (ESRC) e o antigo Joint Information Systems Committee (JISC), e abrigado pela Universidade de Essex. “Com cada vez mais políticas de financiamento de dados de entidades e conselhos de pesquisa é ainda mais importante orientar os

pesquisadores sobre o ‘que fazer e não fazer em compartilhamento de dados, de modo a que estejam em conformidade com as diretrizes’ (BISHOP; VEERLE VAN DEN, 2011).

A promoção e a adoção geral das atividades de compartilhamento de dados dentro das diferentes comunidades científicas são consideradas como importante desenvolvimento estratégico na busca de mais ciência aberta, e uma boa prática científica em geral. Nesse sentido, a e-infraestrutura para os dados abertos da ciência vem sendo colocada como requisito crucial.

## **DESENVOLVIMENTO DE E-INFRAESTRUTURA INTERNACIONAL DE PESQUISA GLOBAL**

Programas e projetos de e-infraestrutura global para ciência aberta estão em andamento em vários países da União Europeia e nos Estados Unidos, Reino Unido, Canadá e Austrália, como forma de ampliar o alcance científico dos conjuntos de dados de pesquisas nacionais, institucionais e individuais. Essas iniciativas se organizam a partir de três vertentes principais: a) por iniciativas entre governos e acordos de caráter multinacional envolvendo diversas nações; b) por iniciativa de múltiplos agentes em atividades de pesquisa internacional e interinstitucionais, conforme, e c) por iniciativa de pesquisadores e indivíduos em comunidades de práticas colaborativas informais da ciência aberta (wikis, blogs, páginas web e repositórios de pesquisadores individuais).

A primeira vertente corresponde à criação da Aliança de Dados de Pesquisa (Research Data Alliance, RDA) com adesão dos países do G8 +6<sup>3</sup>. Nessa vertente, delegados da política científica ou equivalente, representando os ministros da Ciência e Tecnologia, são designados como altos funcionários responsáveis pela pesquisa científica em cada um dos Estados membros. Aumentar o escopo das infraestruturas de dados abertos, aprofundar e criar economias de escala no empreendimento científico é o principal objetivo, tendo em conta que os dados são catalisadores de soluções novas e inesperadas que surgem no âmbito da pesquisa global e multidisciplinar.

<sup>3</sup>G8 Science Ministers Statement. Disponível em: <<https://www.gov.uk/government/news/g8-science-ministers-statement>> Acesso em: 15 maio 2014.

Essa iniciativa ocupa-se de infraestruturas verdadeiramente globais, considerando os setores e projetos de pesquisa globalmente relevantes que se beneficiariam com esse trabalho, e também muitas instalações nacionais que funcionariam melhor se os recursos fossem reunidos. Uma megainfraestrutura global não surge da noite para o dia, ela se organiza a partir de alianças entre comunidades de pesquisa, programas e políticas que promovem o uso integrado de infraestruturas eletrônicas avançadas, serviços de informação que possibilitam acesso e processamento, pela curadoria dos dados, bem como na participação remota (interação) e acesso a experiências científicas, troca e interoperabilidade de dados (MORAIS PIRES; HUDSON, 2012).

A segunda vertente se orienta pela organização de infraestruturas nacionais para implementação de sistemas de intercâmbio global de dados entre instituições e países, a partir de alianças envolvendo associações científicas, profissionais e agências de financiamento (VAN DER GRAAFF; WAAIJERS, 2012), contando com políticas e projetos de cooperação que impulsionam o financiamento de estudos e projetos de redes de informação e centros de dados. O Serviço Nacional de Dados Australiano (ANDS)<sup>4</sup> é um bom exemplo de desenvolvimento de uma série de infraestruturas interoperáveis, incluindo instrumentos utilizados para capturar automaticamente metadados, juntamente com os dados. Constitui um serviço de repositórios de dados financiado por diversos organismos oficiais, tendo por objetivo favorecer a visibilidade na web de dados científicos australianos, promover a curadoria dos dados e contribuir para as políticas de gestão. Dispõe de ampla gama de instrumentos, armazenamento de metadados operacionais e descrição para esses metadados, incluindo ferramentas que permitem eficaz reutilização de dados de pesquisa. O ANDS é alimentado tanto pelas instituições de pesquisa quanto por detentores de dados do setor público, cada uma constituindo um nó de infraestrutura de pesquisa e armazenamento de dados.

---

<sup>4</sup> AUSTRALIAN NATIONAL DATA SERVICE. Disponível em: <<http://ands.org.au/>>. Acesso em: 14 jun. 2014.



A terceira vertente é formada por iniciativa de pesquisadores e indivíduos em comunidades de práticas colaborativas informais da ciência aberta (wikis, blogs, repositórios de pesquisadores individuais), tendo em conta que os dados e as questões de infraestrutura são universais para todos os tipos de instalações. Ilustra bem essa vertente o evento sediado pela Coordenação de Estudos e Pesquisa (COEP) do Ibict, de 18 a 22 de agosto de 2014. O seminário internacional Ciência Aberta, Questões Abertas<sup>5</sup>, que ocorreu em paralelo com o encontro do Grupo de Trabalho em Ciência Aberta, reuniu pesquisadores de diversas universidades, centros de pesquisa e hackerspaces nacionais e internacionais engajados na promoção de práticas abertas na ciência. Conforme divulgado,

“o movimento pela ciência aberta defende a adoção de práticas de pesquisa científica que priorizem a disponibilização de suas produções de forma aberta nas redes para os diversos públicos – de especialistas a leigos. Essa abertura pode envolver desde a publicação, com formatos e licenças abertos, de dados brutos e anotações de pesquisa, até a disponibilização de *softwares*, designs de instrumentos e a já mais difundida publicação de artigos científicos nas revistas de acesso aberto” (OPEN KNOWLEDGE BRASIL, 2014).

As ideias apresentadas nesse seminário apontam soluções de caráter não prescritivo que buscam construir a ponte entre os cientistas e os cidadãos, facilitadoras da confiança no processo científico. Reconhecem que o conhecimento virá a partir de dados criados fora de suas organizações e são regidas pela motivação dos pesquisadores para mudar a face da ciência. A “genialidade” é desbloquear valor, pesquisadores que se juntam em comunidades informais de pesquisa reconhecem que os sistemas que permitem o livre percurso de dados são fundamentais para o futuro da ciência e do planeta, além de formar cidadãos colaboradores na sociedade.

---

<sup>5</sup> Promovido pelo Grupo de trabalho em Ciência Aberta. Disponível em: <<http://www.cienciaaberta.net/>> Acesso em: 14 jun. 2014.

## BOAS PRÁTICAS E LIÇÕES APRENDIDAS

O papel cada vez mais destacado de grandes conjuntos de dados em pesquisas científicas tem importantes implicações para a forma como a pesquisa é realizada, como devem ser organizados e financiados os dados em si, e também como devem ser geridos e compartilhados. Um fato comum, percebido por todos, é que a gestão dos dados de pesquisa está no coração da construção de um novo modelo econômico e social que depende tanto dos incentivos para criação de infraestruturas eletrônicas quanto da capacitação e disposição dos pesquisadores em compartilhar seus dados de pesquisa. A ciência está mudando, esses dados pertencem a todos.

### DADOS DE PESQUISA EM DOMÍNIO PÚBLICO

Embora o compartilhamento de dados abertos nem sempre seja possível para certos conjuntos de dados que necessitam da aplicação de controles de acesso específicos para permitir o compartilhamento de dados confidenciais ou sensíveis, como regra geral a decisão sobre colocar dados de pesquisa de domínio público é um bom exemplo de “desbloqueio de valor”, pois o que determina sua vantagem competitiva são os serviços de informação agregados. Para se atingir o patamar da pesquisa global, deve-se necessariamente criar políticas e iniciativas nacionais consistentes. Em disciplinas tão variadas como medicina, ciências sociais e economia, dados e suas análises são partes essenciais de contribuições dos pesquisadores para seus respectivos campos. Enquanto o compartilhamento de dados de pesquisa para revisão e análise apresenta novas oportunidades para a promoção da pesquisa, a captura desses dados em formulários digitais e o fornecimento de infraestrutura digital para compartilhamento de dados e metadados representam vários desafios (COSTAS; MEIJER; WOUTERS, 2013).

Como vimos na seção 2, natureza dos dados e compartilhamento por áreas de conhecimento, áreas científicas diferem na natureza dos seus dados e em seus métodos e convenções sobre uso de dados e reutilização, mas os dados estão sempre no cerne da ciência, e para a produção de mais dados, mais bolsas de estudos serão concedidas para conduzir pesquisas empiricamente. Esses conjuntos de dados podem ser muito diversos entre si: material arqueológico ou biológico exige dados personalizados de pesquisa ligados a ele; ferramentas digitais de observação em astronomia explicam leis de

funcionamento do universo, realizando extensos e complexos cálculos matemáticos; bibliotecas de arte incluem música e muitas outras coleções de arte digitalizadas que são produtos da criação humana; e dados econômicos e estatísticos referentes a empresas ou países são produzidos como material de pesquisa.

Considerando o meio ambiente não digital, os dados da pesquisa podem ser classificados, em parte, como decorrentes de observação empírica e do esforço humano, e em parte como resultado do conhecimento produzido no ambiente digital automatizado, onde o processamento das coleções tornou os conjuntos de dados mais independentes dos pesquisadores que lhe deram origem. Uma vez coletado, o mesmo conjunto de dados pode ser usado por uma variedade de pesquisadores de diferentes institutos, disciplinas e nações por um período ilimitado de tempo para produzir nova ciência.

### **ACESSIBILIDADE, COMPARTILHAMENTO E CITAÇÃO DE DADOS**

Para a boa ciência e erudição, o acesso, o compartilhamento e a citação de dados agora são considerados fundamentais. Os conjuntos de dados devem ser publicados para que se tornem acessíveis e compartilhados. Eles são considerados também como resultados de pesquisa de primeira classe, do ponto de vista da política social e do financiamento. Analogamente à publicação dos resultados da pesquisa, os dados em que ela se sustenta devem ser citados para serem analisados e reanalisados em novas pesquisas com diferentes abordagens. Citações de dados foram definidas como citações formais, os recursos de dados que levaram a determinado resultado de pesquisa devem ser incluídos na lista de referências dos artigos publicados (COSTAS; MEIJER; WOUTERS, 2013).

Um componente crítico da comunidade acadêmica refere-se às normas universais para citação acadêmica, atribuição de crédito, localização e recuperação de artigos e livros.

Altman e King (2007), por exemplo, propõem um padrão universal semelhante para citar dados quantitativos que retém as vantagens de citações aos impressos, adiciona outros componentes segundo a forma digital e a natureza sistemática de conjuntos de dados quantitativos. E Brase e Farquhar (2011) comentam sobre DataCite, lançado em dezembro

de 2009, uma associação que reúne membros de 10 países, e permite aos pesquisadores localizar, identificar e citar os conjuntos de dados de pesquisa com confiança, promovendo uso de identificadores permanentes para conjuntos de dados. No Instituto de Ciências Sociais Quantitativa da Universidade de Harvard foi desenvolvido um *software* de código aberto para compartilhar, citar, preservar, descobrir e analisar os dados, chamado Dataverse Network (CROSAS, 2011).

Essas ferramentas são importantes porque propiciam a publicação, citação, extração e análise dos dados de pesquisa, incluindo questões legais e políticas de acesso a dados e informações. O principal objetivo dessas ferramentas é resolver os problemas de compartilhamento de dados mediante tecnologias que permitam que as instituições reduzam a carga para pesquisadores e editores de dados, e possam incentivá-los a compartilhar seus dados.

Para o pesquisador, diferentemente da pesquisa e de sua publicação, o compartilhamento de dados de pesquisa não está em sua agenda, porque a preservação dos dados e o compartilhamento não são considerados relevantes ainda para a promoção na carreira e avaliação da pesquisa.

## **O PAPEL DAS BIBLIOTECAS**

Durante os últimos anos, tem-se discutido na comunidade científica o papel das bibliotecas na curadoria, preservação e elaboração de repositórios de dados de pesquisa, especialmente quando a pesquisa é financiada com fundos públicos. A consciência crescente sobre a necessidade de registros dos dados se reflete na profusão de relatórios e publicações científicas sobre o tema. “A sobrevivência dos dados de pesquisa está em questão, uma vez que os dados não são abrigados em instituições de longa vida, como bibliotecas. Essa situação ameaça os princípios subjacentes de replicabilidade científica, já que em muitos casos, os dados não podem ser facilmente coletados novamente (HEIDORN, 2011).”

Novos papéis para as bibliotecas e novos perfis profissionais são demandados para garantir a autenticidade do registro e tratamento adequado dos dados, considerando que os pesquisadores, por si só, não são as pessoas adequadas para assegurar a preservação e o acesso continuado aos dados que produzem e disponibilizam.

Suas competências concentram-se essencialmente no domínio da pesquisa, e “novos” profissionais estão surgindo com conhecimentos no domínio científico dos dados (cientista de dados) e formação específica para tratamento do dado (bibliotecário de dados) (SWAN; BROWN, 2008).

O DATASHARE Repository<sup>6</sup> na Universidade de Edimburgo, Escócia, é um exemplo de serviço avançado de compartilhamento de dados no âmbito de uma biblioteca. Esse cenário é pouco usual, e decorre do fato de a universidade ter reunido, sob a designação de Information Services, o EDINA e a University Data Library. Nesse contexto, o Dspace foi citado como uma ferramenta de uso global e popular, para repositórios de acesso livre. Foi customizada pelo Edinburgh DataShare mediante seleção de campos para metadados padronizados para descoberta de conjuntos de dados, através do Google e outros motores de busca. Essa aplicação permite armazenamento de objetos digitais com suas descrições e gerenciamento de conteúdo, bem como preservação digital oferecendo workflows para submissão e revisão.

Khan e Steinhart (2011) apresentam a concepção da plataforma Repositório de Armazenagem Intermediária (Data Staging Repository, DataStaR) que se refere a partes específicas do ciclo de vida de curadoria de dados de pesquisa: dados e metadados de captura e de compartilhamento antes da publicação, e publicação em repositórios permanentes. O objetivo é apoiar o compartilhamento e publicação de dados e, ao mesmo tempo, permitir a criação de metadados, integrando diferentes esquemas de metadados específicos do domínio, de acordo com as necessidades dos pesquisadores.

Estudos de casos realizados por Rodrigues e Saraiva (2010) apontam os desafios que devem ser enfrentados e afirmam que a existência de várias centenas de repositórios institucionais em estágio de produção não assegura, contudo, sua capacidade de abrigar, preservar e dar acesso a conjuntos de dados científicos.

---

<sup>6</sup> DATASHARE Repository. Universidade de Edimburgo, Escócia. Disponível em: <<http://ie-repository.jisc.ac.uk/336/1/DataSharefinalreport.pdf>> .Acessoem: 10 jun.2014.

A necessidade de conjugar a dimensão institucional (muito ampla e multidisciplinar no caso das universidades) com a dimensão disciplinar (requisitos específicos) constitui um dos principais desafios à utilização dos repositórios institucionais como componente fundamental na infraestrutura global de curadoria dos dados científicos.

Nos Estados Unidos, na Universidade Johns Hopkins (CHOUDHURY, 2010), o repositório institucional (RI) está sendo desenvolvido como componente de uma arquitetura mais ampla de biblioteca digital que irá enfatizar a preservação em longo prazo. O RI é constituído por um conjunto de serviços que será utilizado para suportar as necessidades identificadas ou exigências de professores e alunos. Dado o ambiente de pesquisa intensiva na Johns Hopkins, uma área específica de interesse relaciona-se com conjuntos de dados de uma diversidade de disciplinas que vai das humanidades às ciências. Essencialmente, o IR está sendo criado como uma “porta de entrada” para o arquivo digital subjacente que apoiará curadoria de dados como parte de uma infraestrutura cibernética que se desenvolve com a presença de componentes abertos, modulares. Além dessa estrutura tecnológica, a universidade Johns Hopkins está elaborando novos papéis e relações entre a biblioteca e a comunidade acadêmica, principalmente através do desenvolvimento dos perfis profissionais de “cientistas de dados” ou “humanistas de dados.” Essas iniciativas de RI refletem a percepção de que ele é o primeiro passo em uma jornada mais longa, e para que os esforços institucionais sejam bem-sucedidos, eles devem ser integrados em um panorama maior de repositórios que servem a uma comunidade acadêmica distribuída e diversificada.

Ainda nos Estados Unidos, no intuito de começar a preencher as lacunas e elucidar métodos e processos de curadoria dos dados de pesquisa no contexto científico, um conjunto de instituições consorciadas: Association of Research Libraries (ARL), Association of American Universities (AAU) e Association of Public and Land-grant Universities (APLU), em resposta ao Memo OSTP da Casa Branca, estão desenvolvendo sistemas de informação com essa finalidade. Os repositórios institucionais, interligando a publicação de artigos e dados subjacentes à pesquisa, contam com novos sistemas associados, apoiados na ferramenta Share (ASSOCIATION OF RESEARCH LIBRARY, 2013). Nessa proposta, agências de financiamento e bibliotecas universitárias constituem peças-chave da infraestrutura eletrônica proposta. Adotando

breve conjunto comum de requisitos de metadados e expondo os metadados para motores de busca e outras ferramentas de descoberta de dados pela comunidade de pesquisa, essa proposta estadualiza os repositórios digitais universitários existentes, eliminando a necessidade de um repositório digital central, e alavanca os investimentos já realizados por universidades e suas bibliotecas ao longo da última década.

Repositórios de acesso aberto, cujo número tem aumentado com regularidade, são componentes importantes da infraestrutura de e-pesquisa global. O valor real de repositórios reside no potencial para interligá-los, criando uma rede que pode fornecer acesso unificado aos resultados da pesquisa e ser (re) utilizada por máquinas e pesquisadores. Para alcançar esse potencial necessita-se de interoperabilidade. Um roteiro da Confederation of Open Access Repositories (COAR)<sup>7</sup> está sendo elaborado para fornecer uma visão geral de alto nível de interoperabilidade dos repositórios de acesso aberto, identificar os principais problemas e desafios que precisam ser abordados e estimular o engajamento da comunidade de repositórios.

## CONSIDERAÇÕES FINAIS

Juntamente com as ações de registro dos dados de pesquisa em repositórios, devem-se incluir as ações de preservação. Preservação de dados trata das atuais estratégias globais realizadas em colaboração com os proprietários de dados e prestadores de serviços, entidades e instituições que têm o objetivo de garantir a preservação de dados e conhecimentos em longo prazo. No contexto referenciado encontram sérios desafios no tocante à mídia de armazenamento que aparece no mercado e em seguida desaparece.

O Ibict criou a Rede Brasileira de Serviços de Preservação Digital – Cariniana, que aderiu ao Programa LOCKSS da Stanford University. O serviço foi estendido a instituições com publicações de acesso livre, em que

---

<sup>7</sup> CONFEDERATION OF OPEN ACCESS REPOSITORIES (COAR). The case for interoperability for open access repositories for a more detailed description of the value of interoperability. Working Group 2: Repository Interoperability. Disponível em: <<https://www.coar-repositories.org/files/A-Case-for-Interoperability-Final-Version.pdf>> Acesso em: 10 jun. 2014.

foram incluídos 1.026 títulos de periódicos. A etapa seguinte tem o objetivo de ampliar os serviços da rede, incluindo a preservação de publicações eletrônicas no *software* DSpace, tais como livros, teses e dissertações em formato eletrônico. A Cariniana promove o compartilhamento de estudos e pesquisas, além da integração de conteúdos da memória institucional digital de forma consorciada e federada. Hoje conta com 11 grupos de pesquisa atuando em diversas áreas, tais como Big data, Curadoria, Preservação, Políticas, dentre outros.

Para se atingir o patamar da pesquisa global deve-se necessariamente criar políticas e iniciativas nacionais consistentes que abarquem tanto as questões atinentes à infraestrutura eletrônica quanto aos serviços de informação a ela associados. Dois projetos do Ibict, atualmente em fase de estudo, vão fornecer elementos e orientações nessa direção. O primeiro, o Big Data para Ciência Tecnologia e Inovação objetiva a definição de uma estratégia nacional para curadoria e gestão de dados científicos no Brasil. O segundo, o Mapa da Competência, baseado nos sistemas de informação do tipo CRIS (Current Research Information System) vai permitir o desenvolvimento de serviços de informação que se organizam em torno dos projetos de pesquisa financiados em suas múltiplas relações.

Os dois projetos se complementam. Pesquisadores podem preservar e compartilhar seus dados quando eles são obrigados a fazê-lo por regras impostas pelo financiamento, mas eles realmente não veem o benefício ainda, e políticas de reconhecimento da publicação dos dados de pesquisa precisam ser criadas. Esse é um processo de longo prazo cujas alterações ainda se fazem muito lentamente. Tal mudança precisa de mais tempo e mais orientação para os pesquisadores. Para incentivá-los a compartilhar dados, o mais importante continua sendo apoiar este objetivo com a infraestrutura de serviços correspondente. E nesse item o Ibict também registra sua presença, com a primeira pesquisa de doutorado sobre tese de doutorado sobre o assunto, mais precisamente a formulação de proposta de diretrizes para uma política nacional de curadoria digital e um modelo de publicação científica para a área de ciências nucleares, em que os dados são ligados às publicações acadêmicas por meio de relações semânticas sistematizadas em taxonomia construída para esta finalidade (SALES, 2014).



A realização plena do ideal da ciência aberta, por excelência precursora da ciência cidadã que se exercita dando vistas aos cidadãos das práticas dos cientistas nos laboratórios transparentes da ciência, depende da realização de projetos que ultrapassem as barreiras da academia e os limites do reino dos especialistas para uma sociedade política e tecnologicamente bem informada. Esse é o próximo passo a ser intentado pelo Ibict.



## REFERÊNCIAS

ALBAGLI, S.; APPEL, A. L.; MACIEL, M. L. E-science: ciência aberta e o regime de informação em ciência e tecnologia. E-Science e ciência aberta: questões em debate. In: *Encontro Nacional de Pesquisa em Ciência da Informação (Enancib)*. 2013. Disponível em: <<http://enancib2013.ufsc.br/index.php/enancib2013/XI/enancib/paper/viewFile/168/362pdf>> Acesso em: 5 jun. 2014.

ALTMAN, M.; KING, G. A proposed standard for the scholarly citation of quantitative data. *D-Lib Magazine*, v.13, n. 3/4, 2007. Disponível em: <<http://thedata.org/publications/proposed-standard-scholarly-citation-quantitative-data>>. Acesso em: 14 jun. 2014. DOI:10.1045/march2007-altman

ASSOCIATION OF RESEARCH LIBRARY (ARL). *Shared access research ecosystem (SHARE) proposal*. 2013. Disponível em: <<http://www.arl.org/storage/documents/publications/share-proposal-07june13.pdf>>. Acesso em: 4 jun. 2014.

BISHOP, L.; VEERLE VAN DEN, E. User's guide: do's and don'ts in data sharing. In: *The ODE project: ten tales of drivers and barriers in data sharing*. Dorset, UK: Alliance for Permanent Access (APA), 2011. Disponível em: <[http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/10/7836\\_ODE\\_brochure\\_final.pdf](http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/10/7836_ODE_brochure_final.pdf)>. Acesso em: 17 jul. 2014.

BRASE, J.; FARQUHAR, A. (Ed.). Access to research data: introduction from the guest editors. *D-Lib Magazine*, v.17, n. 1/2, Jan./Feb. 2011. Disponível em: <<http://www.dlib.org/dlib/january11/brase/01brase.html>>. Acesso em: 14 jun. 2014.

CHOUDHURY, G. S. Case study in data curation at Johns Hopkins University. *Library Trends*, v. 57, n. 2, p. 211-220, 2008. Disponível em: <<http://hdl.handle.net/2142/10669pdf>> Acesso em: 12 jun. 2014.

COSTAS, R; MEIJER, I; WOUTERS, P. (2013), opus cit.

COSTAS, R; MEIJER, I; WOUTERS, P. *The value of research data: metrics for datasets from a cultural and technical point of view*. Bonn: JISC; DFG; CSC; Deff; Surf, 2013. Disponível em: <<http://www.knowledge-exchange.info/datametrics>>. Acesso em: 14 jun. 2014.

CROSAS, M. The dataverse network: an open-source application for sharing, discovering and preserving data. *D-Lib Magazine*, v. 17, n. 1 / 2, 2011. Disponível em: <<http://www.dlib.org/dlib/january11/crosas/01crosas.html>>. Acesso em: 14 jun. 2014.

HEIDORN, P. B. The emerging role of libraries in data curation and e-science. *Journal of Library Administration*, v. 51, n. 7-8, p.662-672, 2011. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/01930826.2011.601269#.U5Uf0y35Q>>. Acesso em: 8 jun. 2014.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). *Documento técnico contendo revisão da literatura e análise de publicações sobre ciência aberta (open science) e temas correlatos para elaboração de bibliografia especializada com resumos traduzidos e comentados*. Brasília: Ibiict, 2014.

KHAN, H.; STEINHART, G. DataStaR: using the semantic web approach for data curation. *International Journal of Digital Curation*, v. 6, n. 2, p. 209-221, 2011. Disponível em: <<http://www.ijdc.net/index.php/ijdc/article/view/192>>. Acesso em: 8 jun. 2014.

LONG-TERM Ecosystem Research in Europe. Disponível em: <<http://www.lter-europe.net/>> Acesso em: 5 jun. 2014.

MORAIS PIRES, C.; HUDSON R. L. *Open infrastructures for open science: horizon 2020, consultation report*, 2012. Disponível em: <<http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/open-infrastructure-for-open-science.pdf>>. Acesso em: 14 jul. 2014.

MURRAY-RUST, P.; NEYLON, C.; POLLOCK, R.; WILBANKS, J. *Panton principles: principles for open data in science*, 2010. Disponível em: <<http://pantonprinciples.org>>. Acesso em: 24 set. 2014.

NATIONAL SCIENCE FOUNDATION. Disponível em: <<http://www.nsf.gov>>. Acesso em: 24 set. 2014.

OPEN Knowledge Brasil. Disponível em: <<http://br.okfn.org/2014/08/13/seminario-internacional-discute-ciencia-aberta/>> Acesso em: 20 jun. 2014.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). *Principles and guidelines for access to research data from public funding*. Paris, 2007. Disponível em: <<http://www.oecd.org/dataoecd/9/61/38500813.pdf>> Acesso em: 27 abril 2014.

RODRIGUES, E.; SARAIVA, R. *Os repositórios de dados científicos: estado da arte*. Minho: Universidade do Minho/Projeto RCAA, 2010. Disponível em: <[http://projeto.rcaap.pt/index.php?option=com\\_remository&Itemid=2&func=startdown&id=271&lang=pt](http://projeto.rcaap.pt/index.php?option=com_remository&Itemid=2&func=startdown&id=271&lang=pt)>. Acesso em: 14 jun. 2014.

SALES, L. F. *Integração semântica de publicações científicas e dados de pesquisa: proposta de modelo de publicação ampliada para a área de ciências nucleares*. 2014. Tese (Doutorado em Ciência da Informação) – Escola de Comunicação da Universidade Federal do Rio de Janeiro (UFRJ); Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), Rio de Janeiro. 2014.

SALES, L. F.; SAYÃO, L. F. O impacto da curadoria digital dos dados de pesquisa na comunicação científica. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 17, n. esp.2 – III SBCC, p.118135, 2012.

SCHROEDER, R.; DAVID, P. A.; DEN BESTEN, M. L. *Will e-science be open science?* Working paper series, 14 Dec. 2008. Disponível em: <<http://dx.doi.org/10.2139/ssrn.1317390>> pdf. Acesso em: 03 jun. 2014.

SWAN, A.; BROWN, S. *The skills, role and career structure of data scientists and curators: an assessment of current practice and future needs*. Technical Report, School of Electronics & Computer Science, University of Southampton. Truro, UK: Key Perspectives Ltd, 2008. Disponível em: <<http://eprints.ecs.soton.ac.uk/16675>> Acesso em: 11 maio 2010.

SWAN; B. apud SCHROEDER, R.; DAVID, P. A.; DEN BESTEN, M. L. *Will e-science be open science?* Working paper series, 14 Dec. 2008. Disponível em: <<http://dx.doi.org/10.2139/ssrn.1317390>> pdf. Acesso em: 03 jun. 2014.

VAN DER GRAAFF, M.; WAAIJERS, L. (ed). *A surfboard for riding the wave: towards a four country action programme on research data*. Bonn: JISC; DFG; Surf Foundation; Deff, 2012. Disponível em: <<http://www.knowledge-exchange.info/Default.aspx?ID=469>> .Acesso em: 12 jun. 2014.

## ***Research data management: opportunities and challenges***

### **ABSTRACT**

*It analyzes success stories and lessons learned available in the literature on sharing, reuse and preservation of open data science. Access to open data from science opens up questions that need research, such as new conceptions of databases where researchers evaluate the quality, source and reliability of data. In this context, the establishment of electronic infrastructure capable of coping with the immense challenges is decisive. The complexity of the subject also includes the need to develop approaches that provide data discovery in all disciplines. Scientific data infrastructure providers and funding entities are encouraged to encourage easy access to information resources by promoting professional training and recognition mechanisms that encourage data sharing. Unlike scientific publications, research data require special attention for the articulation of a common agenda, involving funders, researchers and data-producing institutions. Curatorial space, including tools for data discovery, new citation metrics and sharing policies, is gradually being taken up by centers and libraries that develop repositories and coordinated information systems for data registration and dissemination.*

**Keywords:** *Open science. Research data repositories. Data scientist. Data librarians. Electronic infrastructure and information services.*

## **Gestión de los datos de investigación: oportunidades y retos**

### **RESUMEN**

*Se analizan casos de éxito y lecciones aprendidas disponibles en la literatura sobre compartir, reutilización y preservación de datos abiertos de la ciencia. El acceso a los datos abiertos de la ciencia abre cuestiones que necesitan investigación, como por ejemplo, nuevas concepciones de bancos de datos donde los investigadores evalúan la calidad, procedencia y confiabilidad de los datos. En este contexto, se vuelve decisivo el establecimiento de infraestructura electrónica capaz de hacer frente a los inmensos retos existentes. La complejidad del tema comprende también la necesidad de desarrollar enfoques que proporcionen el descubrimiento de datos en todas las disciplinas. Cabe a los proveedores de infraestructura de datos científicos y entidades de financiamiento alentar el fácil acceso a los recursos de información, promoviendo capacitación profesional y mecanismos de reconocimiento que incentiven el intercambio de datos. A diferencia de las publicaciones científicas, los datos de investigación exigen cuidados especiales para la articulación de una agenda común, involucrando a entidades financiadoras, investigadores e instituciones productoras de datos. El espacio de curaduría, incluyendo herramientas para el descubrimiento de datos, nuevas métricas de citas y políticas de compartir, viene siendo poco a poco ocupado por los centros y bibliotecas que desarrollan repositorios y sistemas de información coordinados para registro y divulgación de datos.*

**Palabras clave:** *Ciencia abierta. Repositorios de datos de búsqueda. Científico de datos. Bibliotecarios de datos. Infraestructura electrónica y servicios de información.*

# PRESERVAÇÃO DO PATRIMÔNIO CULTURAL DA MÚSICA BRASILEIRA

Fernando William Cruz  
Juliana Faria Silva  
Luiza Beth Nunes Alonso

## RESUMO

É possível ver a música brasileira como um patrimônio porque faz parte da cultura do povo e ultrapassa os limites do concreto. Legitima-se, portanto, como expressão que ganha caráter de tradição e se contextualiza em sua época promovendo registros subjetivos da organização social e cultural vigente. Apesar da sua importância, percebe-se que não há ainda uma política estabelecida no Brasil para garantir os procedimentos formais e diretrizes de preservação do patrimônio de músicas brasileiras, em especial no que se refere a aspectos tecnológicos. A música brasileira, enquanto promotora de herança cultural imaterial, exige técnicas e procedimentos compatíveis e que vão além das iniciativas tradicionais de organização da informação. Nesse capítulo (i) discute-se a pluralidade e a diversidade da música brasileira, procurando destacá-la em relação a outros tipos de música, e (ii) apresentam-se algumas alternativas e sugestões que podem ser relevantes para garantir a preservação e o acesso a esse patrimônio imaterial por gerações futuras.

**Palavras-chave:** Patrimônio musical brasileiro. Preservação digital de músicas. Tecnologias musicais. Organização da informação musical.

## INTRODUÇÃO

A inclinação para a música é algo inato e essencial em todas as culturas e provavelmente remonta aos primórdios da espécie humana (SACKS, 2007, p. 10). Essa afinidade pode ser desenvolvida ou moldada pela cultura em que se vive, pelas circunstâncias da vida e pelos talentos ou deficiências de cada indivíduo. De fato, percebe-se que a música pode ser experimentada como uma continuação das tradições familiares, como as canções de ninar, os hinos e canções de campo cantadas alegremente no período de férias, estabelecendo uma relação profunda com as raízes culturais de um povo (CRUZ, 2008, p. 1).

Além dos efeitos emotivos que a música provoca nos ouvintes, ela também pode ser vista como forma de arte que ajuda na preservação da história dos povos e pode ser compartilhada por pessoas de diferentes origens, já que sua natureza permite atravessar barreiras linguísticas. Por exemplo, a música ocidental (principalmente a de origem brasileira) tem adeptos apaixonados no Japão, e, provavelmente, muitas pessoas na Europa adoram a música indiana. Presume-se que é possível ter prazer na audição musical sem a necessária compreensão do significado das letras, o que normalmente não ocorre com documentos textuais, por exemplo.

Nesse contexto, é possível ver a música brasileira<sup>1</sup> como um patrimônio porque faz parte da cultura do povo e ultrapassa os limites do concreto. Legitima-se, portanto, como expressão que ganha caráter de tradição e se contextualiza em sua época promovendo registros subjetivos da organização social e cultural vigente. A exemplo da música popular brasileira, Alves (2011) ressalta que esse tipo de música, nas suas mais diversas representações, (i) proporciona o acesso à expressão, ao rosto e à feição de quem fez e faz a fusão cultural brasileira, (ii) é elogiada por sua qualidade no mundo todo caracterizando-se como um dos expoentes da cultura brasileira, e (iii) é a mais abrangente e popular forma de expressão cultural do povo brasileiro.

---

<sup>1</sup> Em um dos primeiros escritos sobre a música brasileira, em 1928, Mário de Andrade classifica seu repertório em “artístico” – a música feita para ser apresentada para um público/audiência – e em “popular” – significando, no sentido genérico de música do povo, de qualquer lugar ou origem étnica.



Características que também podem ser encontradas ou delegadas a outros repertórios de música brasileira que, de algum modo, mescla elementos da cultura brasileira com os de outras culturas. De fato, mais que em qualquer outra expressão cultural, é na música que se pode vivenciar a presença das variadas vertentes étnicas que formam o Brasil, seja pelas letras que falam do cotidiano ou pela origem dos sons e ritmos (variando das músicas regionais às músicas nacionais que tocam nas rádios) que promove a agregação de pessoas em torno de determinados estilos musicais.

Mesmo dentro de gêneros aparentemente inconciliáveis da música brasileira, é possível perceber autores que captam simples expressões populares e sons que remetem às origens culturais, às produções de saberes locais, valorizando tradições caipiras, regionais, ao mesmo tempo em que busca romper com padrões limitadores da criação (ALVES, 2011). Por isso, a música brasileira é vista, ao mesmo tempo, como patrimônio material e imaterial. Patrimônio material enquanto registrada em documentos, a exemplo dos manuscritos musicais, e imaterial uma vez que os documentos são registros que dão suporte às práticas culturais intangíveis (COTTA, 2006). Portanto, essa música é alvo das discussões sobre a necessidade de preservação de herança cultural.

A Unesco (1972) ressalta a importância de se salvaguardar a herança cultural intangível, e propôs cinco grandes domínios sob os quais esse tipo de patrimônio é manifestado: (i) tradições orais e expressões, incluindo a linguagem verbal, (ii) expressões artísticas que envolvem algum tipo de execução, (iii) práticas sociais, rituais e eventos festivos, (iv) conhecimento e práticas relativas à natureza e o universo e (v) artesanato tradicional. A música brasileira se encaixa nesse contexto enquanto possibilidade de criação que carrega informações a serem trocadas entre gerações, dando a oportunidade para as comunidades explorarem livremente sua herança cultural mesmo diante das forças de mercado impostas pela globalização<sup>2</sup>. Apesar disso, a cultura popular brasileira precisa ser reconhecida e considerada nas “versões locais do fenômeno da globalização”, ou seja, a construção de significados vindo de contextos de tempo e espaço sócio-político-econômicos específicos devem ser estudados dentro da especificidade local (BEHÁGUE, 2006, p. 78).

---

<sup>2</sup> Defende-se que a globalização provoca estreitamento nas manifestações culturais porque tem um viés puramente comercial, econômico.

O Instituto do Patrimônio Histórico e Artístico Nacional (Iphan)<sup>3</sup> é o órgão que responde pela preservação do Patrimônio Cultural Brasileiro, em consonância com as recomendações da Unesco (1972). Cabe ao Iphan proteger e promover os bens culturais do País, assegurando sua permanência e usufruto para as gerações presentes e futuras, contando com uma estrutura de superintendências e escritórios técnicos espalhados pelo País. Nesse sentido, promove estratégias de preservação de diferentes formas de expressão cultural, dentre elas, a música enquanto expressão artística, como é o caso dos dossiês sobre o samba (e suas variações)<sup>4</sup> que descrevem detalhes do estilo em consonância com o contexto social subjacente.

Para reforçar a importância da preservação do patrimônio musical, a Constituição Federal brasileira, em seu artigo 216, define o significado deste patrimônio nacional: “Constituem patrimônio cultural brasileiro os bens de natureza material e imaterial, tomados individualmente ou em conjunto, portadores de referência à identidade, à ação, à memória dos diferentes grupos formadores da sociedade brasileira (...)” e atribui ao poder público a responsabilidade de promover, proteger e conservar o patrimônio cultural brasileiro, em seu parágrafo 1º: “O Poder Público, com a colaboração da comunidade, promoverá e protegerá o patrimônio cultural brasileiro, por meio de inventários, registros, vigilância, tombamento e desapropriação, e de outras formas de acautelamento e preservação.” Para Cotta (2006, p. 26), há uma dissociação do conceito tradicional de patrimônio cultural e do patrimônio especificamente musical presente nas orientações das políticas públicas no Brasil e, por isso, os acervos musicais estiveram em uma “espécie de limbo até recentemente, uma vez que não foram considerados nem patrimônio documental, nem patrimônio cultural”. Mas de acordo com o conceito inovador de patrimônio cultural imaterial – que permite o tombamento de bens culturais intangíveis, como a culinária, as danças, as festas populares e as diferentes práticas culturais – as práticas musicais são incluídas nesse patrimônio imaterial.

Boeres e Arellano (2005) argumentam que

---

<sup>3</sup> <http://portal.iphan.gov.br/>

<sup>4</sup> Um exemplo é o “Dossiê das matrizes do samba no Rio de Janeiro: partido-alto, samba de terreiro, samba-enredo”, Disponível em <http://portal.iphan.gov.br/uploads/ckfinder/arquivos/Dossi%20Matrizes%20do%20Samba.pdf>.

[...] a preservação (digital) requer não apenas procedimentos de manutenção e recuperação de dados, no caso de perdas acidentais, para resguardar a mídia e seu conteúdo, mas também estratégias e procedimentos para manter sua acessibilidade e autenticidade através do tempo, podendo requerer colaboração entre diferentes financiadoras e boa prática de licenciamento, metadado e documentação, antes de aplicar questões técnicas.

Do ponto de vista prático, não há ainda uma política estabelecida no Brasil para garantir os procedimentos citados, e as instituições acabam adotando mecanismos *ad-hoc* para preservar o patrimônio de músicas brasileiras.

Embora não se tenha uma noção exata de quanto já se produziu de cultura expressa em forma musical no Brasil, percebem-se algumas iniciativas de resgate e preservação desse tipo de expressão na sua forma material. Um exemplo são as técnicas de desbravamento, coleta e registro de patrimônios musicais em forma impressa promovidas por Sotuyo-Blanco (2004) que permitem descobrir e catalogar expressões musicais escondidas ou esquecidas em espaços familiares, mas que podem representar e contar um pouco da história da época em que essas músicas foram produzidas. Existem outras iniciativas de preservação que são atribuídas aos trabalhos de musicólogos brasileiros como (i) *O ciclo do ouro: o tempo e a música do barroco católico*, publicado em 1978, (ii) a publicação dos Volumes 2 e 3 dos *Manuscritos da Coleção Francisco Curt Lange*, do Museu da Inconfidência (Arquivo Casa do Pilar) de Ouro Preto, (iii) o catálogo temático dos manuscritos musicais de André da Silva Gomes, incluído no livro *Música na Sé de São Paulo colonial*, publicado em 1995 com autoria de Régis Duprat, e (iv) a catalogação da Seção de Música do Arquivo da Cúria Metropolitana de São Paulo (que pertenciam à Catedral São Paulo) por uma equipe coordenada pelo musicólogo Paulo Castagna entre 1996 e início de 1999<sup>5</sup>.

A organização de informações musicais brasileiras como requisito de preservação envolve a produção de CDs de autores consagrados e a digitalização de acervos (como a iniciativa encabeçada pela Biblioteca Nacional do Rio de Janeiro, por exemplo)<sup>6</sup>, além do repositório de partituras

<sup>5</sup> Informações retiradas do trabalho de Castagna (2000) e do site do Museu da Inconfidência – Acervo de Manuscritos Musicais, disponível em [http://www.museudainconfidencia.gov.br/interno.php?pg=musico-logia\\_colecao\\_de\\_manuscritos\\_de-fault&codigo=4](http://www.museudainconfidencia.gov.br/interno.php?pg=musico-logia_colecao_de_manuscritos_de-fault&codigo=4).

<sup>6</sup> <https://www.bn.br/explore/acervos/musica-arquivo-sonoro>

do Centro de Integração, Documentação e Difusão Cultural da Universidade de Campinas<sup>7</sup>. Em termos de abordagens para preservação de músicas brasileiras, existem iniciativas que variam desde a preservação tecnológica (no caso de formato digital) com a manutenção de estruturas de hardware e software, até a estratégia de migração de suportes tecnológicos para garantia de perpetuação de registros musicais em formato de áudio, como as conversões do vinil para o CD e do CD para o MP3.

É comum encontrar em instituições brasileiras de herança cultural, como bibliotecas, museus e arquivos, a utilização de plataformas para armazenamento de recursos digitais, tais como o DSpace<sup>8</sup> ou o Greenstone<sup>9</sup>, para formação de repositórios e bibliotecas digitais dando um caráter institucional para o aspecto da preservação.

Embora os espaços para discussão sobre preservação e curadoria digital sobre música em sentido amplo sejam limitados, há eventos científicos nacionais em que esse tema tem sido discutido: (i) o SBCM (Simpósio Brasileiro de Computação Musical), que é formado por uma comunidade multidisciplinar, e (ii) o congresso anual da ANPPOM (Associação Nacional de Pesquisa e Pós-graduação em Música), formado por profissionais ligados mais diretamente à área da música, que também têm buscado se integrar às iniciativas de organização e preservação de música brasileira, como o projeto BDB-Mus, relatado por Castro, Alonso, Ferneda, Cunha, Cruz e Brandão (2006).

Do ponto de vista da manutenção de artefatos físicos, o Brasil possui algumas iniciativas de preservação no contexto digital, considerando aspectos típicos de preservação do recurso de informação tais como o conteúdo, a fixidez, a referência, a proveniência e o contexto (YEUNG, CARPENDALE e GREENBERG, 2008).

---

<sup>7</sup><http://www.cididic.unicamp.br/>

<sup>8</sup><http://www.dspace.org/>

<sup>9</sup><http://www.greenstone.org/>

Nessa direção pode ser citada a Rede Brasileira de Serviços de Preservação Digital – Cariniana<sup>10</sup>, fundamentada numa estrutura descentralizada para preservação de documentos eletrônicos brasileiros de modo a garantir acesso continuado de longo prazo a esses documentos digitais, e que pode ser uma opção de preservação para mídias musicais brasileiras.

Não obstante a relevância das iniciativas citadas, percebe-se que a música brasileira promove herança cultural imaterial e, por esse motivo, exige técnicas e procedimentos compatíveis e que vão além das iniciativas apresentadas. Neste capítulo discute-se a pluralidade e a diversidade da música brasileira, procurando destacá-la em relação a outros tipos de música, e apresentam-se algumas alternativas e sugestões que podem ser relevantes para garantir a preservação e o acesso a esse patrimônio imaterial por gerações futuras.

## IDENTIFICAÇÃO DO PATRIMÔNIO A SER PRESERVADO

Na perspectiva de identificação de estratégias para preservação da música brasileira, uma questão que surge é com relação ao que preservar diante da diversidade disponível, principalmente num contexto onde vários artistas e intérpretes nacionais famosos têm produzido música para divulgar novas produções, para manter os fãs atualizados quanto aos lançamentos e notícias e até para realizar vendas de discos. Da mesma forma, grupos não tão famosos, compositores anônimos e bandas de garagem também se organizam para divulgar e vender suas produções musicais, seguindo o movimento conhecido como *Long Tail*<sup>11</sup>. Na mesma diversidade estão as músicas criadas a partir de um elemento musical do folclore brasileiro ou com influências de uma tradição musical de uma cultura local, além das que têm relação com aspectos ritualísticos envolvendo crenças e religiosidade que possam permear a composição e a divulgação de certo repertório musical.

---

<sup>10</sup> <http://www.ibict.br/pesquisa-desenvolvimento-tecnologico-e-inovacao/rede-brasileira-de-servicos-de-preservacao-digital>

<sup>11</sup> O Long Tail pode ser compreendido como um processo de democratização da divulgação e venda de músicas, onde inúmeros grupos musicais pelo mundo todo conseguem ofertar suas variadas produções musicais competindo de igual para igual com as grandes produtoras atuais em função da conectividade provida pela Internet (Anderson, 2006).

É sabido que o patrimônio musical é relacionado a aspectos culturais e sociais, principalmente que envolvem tanto a sua produção, distribuição e recepção. Portanto, quais músicas estão contidas no patrimônio da cultura brasileira? Quais repertórios ou quais requisitos distinguem essas músicas do restante das músicas ao redor do mundo? Esta discussão talvez encontre eco no ensaio sobre a música brasileira, de Mário de Andrade, realizado no ano de 1928. O *background* naquele período foi o nacionalismo – o que era ou não considerado genuinamente brasileiro na música e nas outras artes (pode-se fazer uma menção à Semana de Arte Moderna, que trouxe à tona a escultura, a arquitetura, a literatura – além da música – produzida no país).

Para Mário de Andrade, diferentemente dos demais que concebiam uma identidade musical a partir da noção de mestiçagem, a música produzida pelo povo brasileiro revelaria a unidade nacional, e em seu ensaio defende uma música brasileira que deveria: (i) ser toda criação musical nacional com ou sem caráter étnico; (ii) ser uma manifestação musical realizada por brasileiro ou indivíduo nacionalizado e que refletisse as características musicais de sua raça; (iii) ser feita por artistas brasileiros que se considerassem ou não gênios: “todo artista brasileiro que no momento atual fizer arte brasileira é um ser eficiente com valor humano”. No primeiro caso, esse artista “saberá encontrar fatalmente os elementos essenciais da nacionalidade” trazendo valor social à obra sem perder o artístico. No segundo, “o artista que faz parte dos 99 por cento dos artistas e reconhece que não é gênio, então é que deve mesmo de fazer arte nacional” porque, do contrário, sua arte internacional será inútil (ANDRADE, 1972, p. 19); (iv) não ser analisada sob o ponto de vista da opinião europeia – o que poderia falsificar a entidade brasileira; (v) conter uma pequena porção da música ameríndia, uma porção maior da música africana e muita quantidade da música portuguesa, além das influências espanholas, principalmente a hispano-americana – os gêneros da habanera e do tango de Cuba e Montevideu – e a europeia com suas danças – a valsa, a polca, a mazurca e a shottsh – que contribuíram na formação da modinha.

Andrade (1972, p. 13) entende que o fato de fazer sucesso na Europa não implica nenhuma importância para a música brasileira que estava, no século XIX, sendo apenas delineada para que, nos “derradeiros tempos do Império” abundasse. Além disso, a fusão com a música popular da Monarquia não é satisfatória por ser uma música feita por portugueses e africanos, e não brasileiros.

Para esse autor, desconhecemos a complexidade da música brasileira em razão de que a execução do ritmo escrito nas partituras daquela época serem totalmente distintas da grafia musical das impressões. Sua opinião se confunde no que diz respeito às influências que se devem utilizar na música brasileira – Andrade elogia algumas influências estrangeiras, como as do *jazz* que se infiltraram no maxixe, mas critica outras, as do tango na canção brasileira, defendendo que “está claro que o artista deve selecionar a documentação que vai lhe servir de estudo ou de base. (...) A reação contra o que é estrangeiro deve ser feita espertalhonamente pela deformação e adaptação dele. Não pela repulsa” (ANDRADE, 1972, p. 26).

O compositor brasileiro, para Andrade (1972, p. 26),

[...] tem de se basear (...) no folclore (...) [e] por isso não pode ser nem exclusivista nem unilateral. Si exclusivista se arrisca a fazer da obra dele um fenômeno falso e falsificador. E sobretudo facilmente fatigante. Se unilateral, o artista vira antinacional: faz música ameríndia, africana, portuguesa ou europeia. Não faz música brasileira não.

Em seu ensaio, Mário de Andrade descreve as características da música brasileira apresentadas no quadro 1.

Quadro 1 - Elementos que a música brasileira precisa conter para expressar a nacionalidade  
(Continua)

Ritmo	<ul style="list-style-type: none"> <li>• A síncope<sup>12</sup> é recorrente, mas não obrigatória foi “mais provavelmente importada de Portugal que da África”</li> <li>• Influência ameríndia: “rítmica de canto quase que exclusivamente fraseológica (...) coincidindo pois em muitas manifestações com a rítmica discursiva de Gregoriano”</li> <li>• Influência africana: “frases [rítmicas] oratórias [desprovidas de valores de tempo musical] aparecem e chegam mesmo a criar recitativos legítimos” que “contrastavam com a música portuguesa afeiçoada ao mensuralismo tradicional europeu”</li> <li>• Conflito entre as “nossas tendências” e a rítmica “quadrada” vinda da Europa que Portugal trouxe ao Brasil - entre a rítmica da música portuguesa com a prosódica das músicas ameríndias e africanas</li> <li>• O brasileiro se “acomodou” diante desse conflito, “fazendo disso um elemento de expressão musical”, adquirindo um jeito próprio de ritmar “fez do ritmo uma coisa mais variada mais livre e, sobretudo um elemento de expressão racial”</li> <li>• No populário musical brasileiro, algumas síncopes podem representar polirritmia ou ritmos livres</li> <li>• O artista deve atentar-se (i) à realidade da prática musical popular para empregar a síncope em suas composições; (ii) em não se restringir às síncopes que aparecem no “populário impresso”; (iii) em “empregar com frequência e abuso o elemento direto fornecido pelo folclore” pois a música enquanto arte “não é fenômeno popular, porém desenvolvimento deste”; (iv) em observar certos binários compostos que caracterizam a influência portuguesa que prevaleceu na música nordestina; (v) ao quaternário gaúcho; (vi) às valsas; (vii) às mazurcas e (viii) às modinhas ... para que a sua originalidade se dê “quando sair do característico popular a variedade métrica que o populário não fornece”</li> </ul>
-------	--

(Continua)

<sup>12</sup>“(...) o som da melodia na chamada parte fraca do compasso ou do tempo e se prolonga até uma acentuação seguinte” (ANDRADE, 1972).



## Quadro 1 - Elementos que a música brasileira precisa conter para expressar nacionalidade

Melodia	<ul style="list-style-type: none"> <li>• Inexpressiva quando baseada em melodia popular</li> <li>• Fazer música nacional esquecendo os “tesouros” de expressão musical como Lasso, Monteverdi, Carissimi, Gluck, Beethoven etc.</li> <li>• Atentar aos aspectos psicológicos (das letras, por exemplo) e de humor (tristonho, gracioso, elegante, apaixonado etc.)</li> <li>• Ambientar para os ouvintes o poema musicado que não implique liberdade individual, nem falta de caráter étnico ou utilize regras e fórmulas estreitas</li> </ul>
Harmonia	<ul style="list-style-type: none"> <li>• A harmonização acompanhante tem pouca importância na música popular para definir um caráter “nacionalizador”</li> <li>• A música artística não pode se restringir aos processos harmônicos populares, pois são “pobres por demais”</li> <li>• Precisa ser desenvolvido pelos compositores coincidindo com a harmonia europeia “a não ser que a gente crie um sistema novo de harmonizar”</li> <li>• A harmonização precisa se sujeitar às “leis acústicas gerais” e às “normas de harmonização da escala temperada”</li> <li>• Se houver um processo novo de harmonizar “não poderá assumir caráter nacional” já que é invenção individual e sem base no “populário”</li> <li>• A polifonia é a que possui maior caráter nacional refletido nos “contracantos e variações temáticas superpostas empregadas pelos nossos flautistas seresteiros, os baixos melódicos do violão nas modinhas, a maneira de variar a linha melódica em certas peças, tudo isso desenvolvido pode produzir sistemas raciais de conceber a polifonia”</li> </ul>

(Continua)

## Quadro 1 – Elementos que a música brasileira precisa conter para expressar nacionalidade

Instrumentação	<ul style="list-style-type: none"> <li>• Mesmo os instrumentos importados não impedem um caráter nacional (mesmo que atuem como solistas): “O próprio piano aliás pode ser perfeitamente tratado pelo compositor nacional sem que isso implique desnacionalização da peça. O violino se acha nas mesmas condições e está vulgarizadíssimo até nos meios silvestres. Numa fazenda de zona que permaneceu especificamente caipira, tive ocasião de escutar uma orquestrinha de instrumentos feitos pelos próprios colonos. Dominavam no solo um violino e um violoncelo... bem nacionais. Eram instrumentos toscos, não tem dúvida, mas possuindo uma timbração curiosa meio nasal meio rachada, cujo caracter é fisiologicamente brasileiro”</li> <li>• Não se trata de desafinação e sim de “caráter de sonoridade, de timbre”</li> <li>• Não comparar os instrumentos e as vozes nacionais tendo como base a “timbração europeia”</li> <li>• “O sinfonismo contemporâneo, que não é de nenhuma nacionalidade, é universal, pode perfeitamente ser brasileiro também (...) porque é justamente a maneira de tratar o instrumento quer solista quer concertante que nacionalizará a manifestação instrumental”</li> </ul>
Forma	<ul style="list-style-type: none"> <li>• “Quanto ao emprego de certas formas tradicionais, não vejo prejuízo nisso, embora não recomende. É uma inutilidade”</li> <li>• “Os nossos compositores têm demonstrado poder criador bem pequeno a respeito da forma, não se aproveitando das que o populário apresenta”</li> <li>• “O canto nacional apresenta uma variedade formal que sem ser originalidade dá base vasta pra criação artística de melodia acompanhada. Possui uma diversidade rica de formas estróficas com ou sem refrão”</li> <li>• Enfatiza as composições para coros por causa da “riqueza moderna em que a voz pode ser concebida instrumentalmente, com puro valor sonoro”</li> </ul>

Fonte: adaptado de Andrade (1972).

Em acréscimo, Blomberg (2011) relata que as histórias da música brasileira são poucas, pois não há muita informação sobre seus autores, métodos ou escolas, mesmo que os livros que as contam sejam referenciados e acessados pelo público leigo ou especializado. Para ela, a música praticada pelos primórdios da descoberta do Brasil – pelas comunidades ameríndias e, posteriormente, pelos

européus e africanos foi registrada por meio de relatos de viagens de estrangeiros, missionários ou administradores, e somente no século XX é que essas histórias foram abordadas sob um critério mais metódico. Para essa autora, os primeiros registros musicais no Brasil apareceram documentados em relatos de viagens por estrangeiros. Também afirma que houve pouco contato dos historiadores com a música no Brasil; todas as obras sobre as histórias da música brasileira que existem foram publicadas no século XX e possuem características comuns – são narrativas, biográficas e lineares – e, ainda,

[...] Enquanto os títulos propõem uma história da música brasileira ou no Brasil, de forma ampla, claramente seus conteúdos foram desenvolvidos de acordo com os recortes impostos ou, aos autores, por circunstâncias externas ou, pelos próprios autores. (BLOMBERG, 2011, p. 424).

Outras características encontradas por Blomberg (2011) a respeito das publicações que dizem respeito às histórias da música brasileira foram as seguintes:

- a gradual formação dessa música tendeu a relacionar-se com a formação do Estado Moderno – o início das histórias, embora citem o descobrimento do Brasil nas primeiras páginas, se dá a partir do período colonial (século XVIII) até à segunda metade do século XX;
- alguns historiadores preocupam-se com o papel social da música e suas definições: “nestes livros são formas e gêneros musicais e não obras, que são identificados na narrativa dos costumes brasileiros” (BLOMBERG, 2011, p. 427), primeiramente comparados aos dos europeus e depois confrontados quanto às suas origens – urbana ou rural;
- a importância da identificação das formas musicais para todos os autores, que são danças – batuque, fandango, quadrilha, samba, maxixe, frevo etc.; cantigas – de trabalho, de ninar, sentimentais como as modinhas e toadas, brejeiras como lundus e emboladas, religiosas, satíricas e fúnebres; e bailados – congos, maracatus, bumba-meu-boi etc.;
- intensificação da separação entre música erudita, popular e folclórica: “a popular (...) corresponderia à música urbana, seria posteriormente confrontada com (...) a música erudita, enquanto a folclórica seria definida ainda como música de origem rural”;

- identificam-se duas formas de “evolucionismo”<sup>13</sup> nas histórias: a da concepção de música que evolui do mais simples ao mais complexo, em que a noção de evolução teve como objetivo a formação de uma unidade nacional a partir das várias identidades musicais, e a da história baseada na evolução dos gêneros musicais ou dos estilos.

Napolitano e Wasserman (2000) estabelecem uma cronologia da produção historiográfica sobre a questão da música no Brasil nas décadas do século XX: (i) nos anos de 1920 e 30: preocupação com a brasilidade, a identidade nacional e os procedimentos de como deveria ser pesquisada e incorporada a música folclórica, além dos projetos ligados aos modernismos musicais (samba atua como o elemento principal da MB); (ii) anos 1940: consolidação de um “pensamento historiográfico sistematizado em torno da música urbana” a partir da evolução do samba; (iii) virada dos anos 1940 para os anos 50: a afirmação do samba como gênero de origem brasileira contra outros gêneros reconhecíveis que interferiram na audiência nacional (o *jazz*, o bolero e a rumba); (iv) anos 1950: “temor pela internacionalização e perda de referenciais da cultura nacional” gerado pela percepção de que a música brasileira estava perdendo espaço nos meios de comunicação fazendo como que “o debate nascido ainda nos anos 30, sobre a necessidade de se estabelecer a autenticidade do samba (...) ganhou nova força”; (v) final dos anos 1950: a Bossa Nova abala “toda a estrutura de criação e audição, baseada nos gêneros estabelecidos, na medida em que procurava uma renovação dentro da tradição do samba”; (vi) anos de 1970 e 80: a pureza étnica e social da origem do samba começa a perder vigor nas análises acadêmicas sobre o tema – ênfase nos novos padrões e identidades dos gêneros urbanos configurados para o consumo; e, (vii) meados dos anos de 1980: desafio dos trabalhos historiográficos – “favorecer a escuta simultânea e sincrônica das fontes musicais sem reduzir a história da música (brasileira) a um mero apêndice da história da sociedade ou da história das ideias” (NAPOLITANO; WASSERMAN, 2000, p.p. 172, 174, 178, 186).

---

<sup>13</sup>Toma-se como justificativa o conceito de historiografia, que “se refere à avaliação ou crítica de obras históricas enquanto documentos, como testemunhos de dimensões específicas de cada autor com relação à eleição de seus tópicos métodos e contextos” (BLOMBERG, 2011, p. 428).

A sigla MPB aparece no início dos anos 1960, mas não se sabe o momento exato, e um dos seus primeiros registros conhecidos é o nome do conjunto vocal MPB-4. A significação da sigla como etiqueta mercadológica é mais recente. Nos anos 1990 houve uma fragmentação do panorama musical, e a concepção de música-popular-brasileira como frente única e compactada é questionada por causa da afirmação de identidades musicais regionais ou estaduais (como o *mangue-beat* pernambucano e o axé baiano) e das músicas transnacionais (como o *rap* e o *funk*) ou daquelas de falsa popularidade (como o forró estilizado e o pagode romântico) (SANDRONI, 2010).

## PATRIMÔNIO MUSICAL BRASILEIRO VISTO COMO INFORMAÇÃO

O fenômeno do *Big Data*, caracterizado por grande volume de informações, ocorre também com a música popular brasileira, que é representada por artefatos (convertidos ou nascidos digitais) que envolvem uma variedade de possibilidades: (i) músicas impressas com partituras completas ou parciais, partituras vocais, livros de coro, materiais específicos para empréstimo para execução em concertos e outros formatos de música impressos; (ii) músicas em manuscritos feitos em papel; (iii) registros sonoros em formatos variados como cilindros, CDs, arquivos MP3 e outros; e (iv) *libretos* ou textos de música impressa ou manuscrita, produzida normalmente por ocasião de uma situação específica ou mesmo de uma *performance* de uma letra e frequentemente incluindo os nomes de todos os autores, cantores e pessoal de apoio, bem como a data e o lugar da *performance*.

A preservação desse patrimônio, de volume considerável e crescente, exige um olhar que vai além da percepção de que se está falando de uma expressão de arte e elemento cultural com raízes profundas. Percebe-se, portanto, a necessidade de se caracterizar a música brasileira como um tipo de informação passível de registro. Aliás, o interesse sobre o registro de informações musicais acompanha a história humana desde os tempos remotos. Um exemplo de registro está ilustrado na figura 1, contendo resquícios de uma partitura grega antiga, onde são usados símbolos taquígrafos (com notação fonética) e que faz parte da história das partituras atuais, mais complexas e descritas em notação CMN (*Common Music Notation*).

Figura 1: Fragmento de partitura grega



Fonte: <http://classics.uc.edu/music/michigan/index.html>

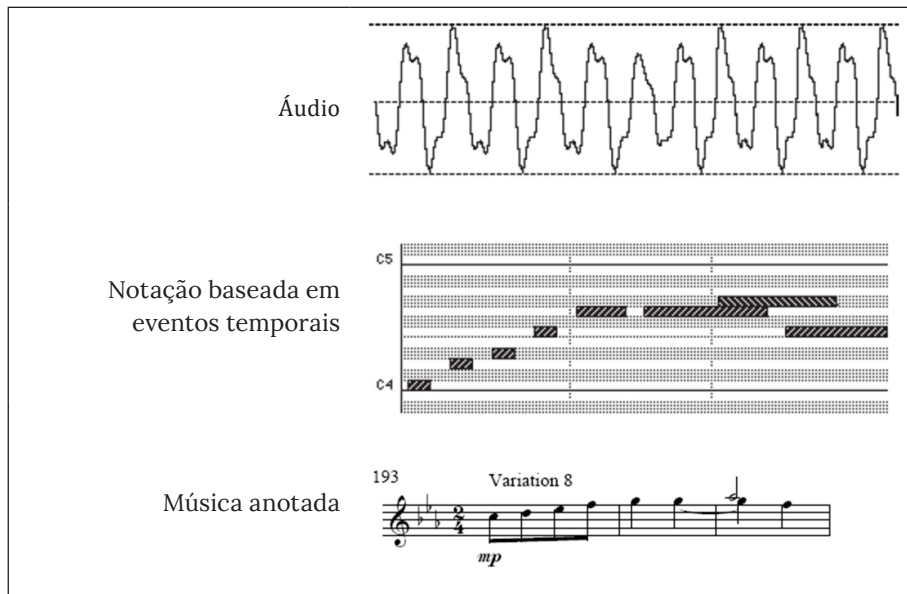
A compreensão da música enquanto informação é assunto ainda em aberto na comunidade científica, uma vez que não é trivial a percepção de significados numa obra musical. Embora haja exceções, a música em si pode ser considerada como um recurso cujas estruturas são livres de qualquer denotação (LESAFFRE, 2006, p. 14), o que a difere de outros tipos de informação que conseguem passar significados mais facilmente. Por exemplo, é possível perceber semântica em documentos textuais, e os recursos de uma imagem podem representá-la muito bem. Entretanto, passar um conceito numa música – principalmente se ela não contém letra – é uma tarefa difícil, e ainda não está claro que tipo de significado um trabalho musical é capaz de armazenar. De fato, a significação ocorre quando conteúdos musicais são interpretados pelo usuário e, por consequência, geram associações, às vezes muito pessoais. Em outras palavras, um conjunto de ondas sonoras que eventualmente viajam pelo espaço pode evocar diferentes significados e conhecimentos, dificultando a sua formalização.

Não obstante os problemas conceituais citados, existem definições na literatura sobre informação musical com vistas a permitir a aplicação de técnicas de organização por processos sistematizados (indexação, classificação, indexação, catalogação, etc.) e de preservação com uso de padrões e técnicas similares aos que são aplicados aos documentos textuais. Nessa perspectiva, uma das primeiras iniciativas de formalização foi feita por McLane (1996), que trouxe uma visão subjetiva para a música, permitindo enxergar a diferença entre o conteúdo em relação ao suporte que o contém. Em 2003, Downie refez considerações sobre a informação musical, procurando caracterizá-la como recurso informacional contendo dimensões temporal, harmônica, *pitch*, timbral, editorial, textual e bibliográfica. Mais recentemente, Cruz (2008, p. 279) atualizou a definição acrescentando elementos que dão a entender a relação da música com o seu contexto:

[...] a música possui as facetas estabelecidas por Downie (2003), mas a sua estrutura incorpora elementos adicionais que permitem defini-la como um objeto informacional musical mais amplo, dotado de conteúdo – atributos internos e metadados descritivos – e, de contexto – associações com outros objetos musicais e não musicais, e com situações ou eventos em que este objeto musical está inserido.

Embora tenha havido uma aproximação da música e do texto, o acesso às informações musicais (como requisito de preservação) possui uma complexidade adicional quando comparado ao cenário dos documentos textuais em função das dimensões apresentadas por Downie (2003). Para efeito de uso das TICs – Tecnologias de Informação e Comunicação e das recomendações e padrões de preservação, tais dimensões precisam estar associadas a alguma forma de representação. Byrd (2007) propõe o (i) o áudio, (ii) a notação baseada em eventos temporais, e (iii) a música anotada (CMN) como alternativas de representação completa da música (figura 2) porque possuem variados graus de estrutura e representatividade e podem ser combinadas para esclarecer melhor sobre o recurso musical.

Figura 2: Formas de representação da música



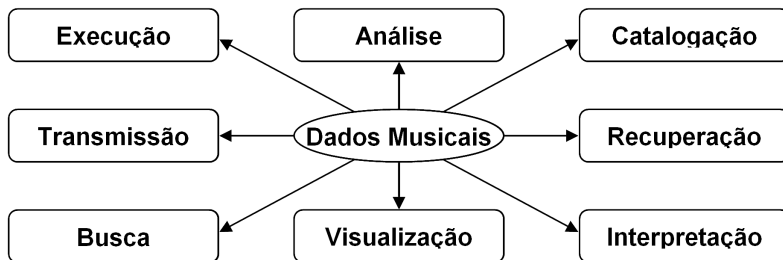
Fonte: Byrd (2007)

Enquanto as representações citadas são abstrações de diferentes tipos de informação musical, os formatos referem-se a como essas informações musicais são expressas. Sob esse ponto de vista, a mesma representação musical pode ter vários formatos ou codificações associados. Na literatura são apontados diferentes formatos, com propósitos distintos para cada uma das formas de representação citadas por Byrd (2007). Por exemplo, o áudio pode ser codificado em formatos WAV (modo não comprimido) ou MP3 (modo comprimido), e a representação baseada em eventos temporais possui um formato digital popular e conhecido como Musical Instrument Digital Interface (MIDI). Além dos formatos ligados às formas de representação musical, existem os que são ligados a interesses de uso específicos para a música, como apresentado na figura 3.



Alguns exemplos incluem (i) o Humdrum/\*\*Kern<sup>14</sup> – voltado para análise musical de músicas monofônicas, (ii) o DARMS (Digital Alternate Representation of Music) – para representar música anotada (impressa) e (iii) o Plaine and Easie – para indexação e catalogação de incipits musicais.

Figura 3: Funcionalidades possíveis para formatos de intercâmbio



Fonte: Schwartz (2003, p. 3)

Do ponto de vista da preservação, percebe-se que essa diversidade de formatos potencializa os problemas relacionados às estratégias recomendadas, tais como a migração tecnológica ou a emulação. Uma alternativa possível é a adoção de formatos baseados na metalinguagem XML que é extensível e declarativa (em oposição às linguagens procedurais) e que permitam derivar todas as necessidades de representação e uso citados. Um exemplo desse tipo é o MusicXML<sup>15</sup>, criado para permitir o intercâmbio de dados entre softwares de propósitos distintos.

## ESTRATÉGIA DE PRESERVAÇÃO DE MÚSICAS BRASILEIRAS

Dentre os requisitos de preservação digital está a garantia do acesso às informações musicais pela aplicação de ferramentas bibliográficas, tais como classificação, indexação, catalogação e outras técnicas voltadas para atender às necessidades de busca, incluindo repositórios e bibliotecas digitais. Existem, portanto, inúmeros sistemas para organização de informações já em uso que podem ser adaptados para as músicas brasileiras, incluindo

<sup>14</sup><http://www.musiccog.ohio-state.edu/Humdrum/>

<sup>15</sup><http://www.musicxml.com/>.

*softwares* mais elaborados que conseguem tratar a plasticidade<sup>16</sup> da música com interfaces sofisticadas relacionadas às diferentes dimensões citadas (como a melodia e o timbre em músicas monofônicas e polifônicas<sup>17</sup>).

No entanto, a simples aplicação de tecnologias e padrões às informações musicais brasileiras não garante a sua preservação, dado que esses sistemas nem sempre conseguem dar conta das necessidades de informação dos usuários. Portanto, o acesso de longo prazo de documentos musicais brasileiros pode ser feito considerando-se algumas sugestões sobre as peculiaridades das músicas brasileiras e sobre as novas tendências de registro bibliográfico que são discutidas mais adiante neste capítulo.

## **SUGESTÕES RELACIONADAS AO ASPECTO DE ORGANIZAÇÃO DA INFORMAÇÃO MUSICAL**

No Brasil é comum a utilização de metadados para descrever conteúdos musicais porque permite criar um número variado de pontos de acesso para encontrar, identificar, selecionar e obter o recurso. Essas descrições em geral são informações sobre a própria obra (gênero, tonalidade, etc.), sobre o documento em si (tipologia, publicação, dimensões, etc.) e descrições sistemáticas apoiadas por terminologias ou tesouros e de sistemas de classificação. Nesse sentido, a catalogação de recursos musicais brasileiros tem sido feita com uso dos formatos MARC e Dublin Core seguindo regras de catalogação mais gerais, como a AACR2, ou mais específicas, como as que são propostas pela IFLA (International Federation of Library Association and Institutions) e MLA (Music Library Association)<sup>18</sup>. Mesmo não tendo sido criado com esse propósito, o RISM (Repertoire Internationale des Sources Musicales) também pode ser colocado como um padrão normativo para catalogação de informações musicais para a compreensão e análise musical<sup>19</sup>.

---

<sup>16</sup>Segundo Downie (2003), a plasticidade da música se refere ao fato de ela admitir certas flexibilidades estruturais e ainda assim continuar sendo a música original. Dentre as flexibilidades citadas estão a transposição de tons, de ritmos, harmonia, letras, dentre outras.

<sup>17</sup>O ISMIR ([www.ismir.net](http://www.ismir.net)) e o ICMC ([www.icmc2016.com](http://www.icmc2016.com)) são exemplos de eventos científicos cuja comunidade se dedica a projetos ligados à aplicação de tecnologias para o tratamento de informações musicais.

<sup>18</sup> O endereço eletrônico da IFLA é <http://www.ifla.org>, e da MLA é <https://www.musiclibraryassoc.org>.

<sup>19</sup> Existe uma ramificação do RISM voltada para músicas brasileiras. Para mais informações, consultar no endereço <http://www.rism-br.mus.ufba.br/>.

Embora haja muitas fontes de informação sobre música brasileira com o propósito de preservação, percebe-se que muitas delas não dão conta de demandas mais específicas de consultas musicais, e esse pode ser considerado um dos problemas a serem resolvidos para garantir o acesso de longo prazo a esse acervo. A fim de perceber o quão preparados estão os sites de música brasileira na Internet para as estratégias de preservação digital, foi feito um experimento de análise de uma amostra dessas diferentes fontes, submetendo-as aos critérios de avaliação definidos por Sayão (2005).

Os critérios para manter os objetos digitais perenemente acessíveis para uso consideram os aspectos de: (i) preservação física das mídias e a sua renovação quando necessário; (ii) preservação lógica dos formatos e a dependência de *hardware* e *software* que os mantenham legíveis e interpretáveis; (iii) preservação intelectual do conteúdo incluindo sua autenticidade e integridade; (iv) preservação do aparato dos metadados para a localização, recuperação e representação da informação digital, e (v) monitoramento e instrumentalização da comunidade-alvo para que ela possa compreender plenamente a informação no momento do acesso.

Os resultados dessa experiência estão apresentados no quadro 2. Por exemplo, com relação à questão da preservação física, os sites de música brasileira visitados foram contemplados em alguma medida. No entanto, apenas o Instituto Tom Jobim e a Escola de Música da UFRJ apresentaram resultados para o critério de preservação lógica, e nenhum dos sites testados continha informações consistentes sobre o critério de preservação intelectual.

Quadro 2 - Exemplo de aplicação de critérios para preservação de música digital (Continua)

	Instituto Antonio Carlos Jobim	Instituto Paulo Moura	Escola de Música da UFRJ
Preservação física	Digitalização de partituras, fotos e documentos. Dispositivo para ouvir áudios de músicas, entrevistas e notícias de compositores de MPB (Tom Jobim, Milton Nascimento, Paulo Moura, Chico Buarque, Dorival Caymmi, Gilberto Gil)	Fotos, vídeos, áudios, materiais variados (entrevistas, relatos de pesquisas, etc.). Direciona muitos links externos para áudios, vídeos etc.	Partituras e manuscritos musicais digitalizados.
Preservação lógica	Problemas para ouvir o áudio.	Nada foi encontrado sobre o tema.	Software adaptado para a leitura das partituras e dos manuscritos.

<sup>20</sup>Acervo digital do Instituto Antonio Carlos Jobim (<http://portal.jobim.org/pt/acervos-digitais/>)

<sup>21</sup><http://www.institutopaulomoura.com.br/home/index.html>

<sup>22</sup>Biblioteca digital da Escola de Música da UFRJ (<http://www.docpro.com.br/escolademusica/bibliotecadigital.html>)

Quadro 2 - Exemplo de aplicação de critérios para preservação de música digital<sup>1</sup> (Continua)

	Instituto Antonio Carlos Jobim	Instituto Paulo Moura	Escola de Música da UFRJ
Preservação intelectual	Nenhuma. Não há o foco na descrição da obra musical. Parte de um item vinculando-o a um autor. Há um link na descrição de alguns compositores direcionando para páginas em que o conteúdo apresenta a linha do tempo da vida e obra do compositor <sup>23</sup> .	Não há uma linha de raciocínio que organize os conteúdos. São várias matérias (algumas com performances em vídeo), outras notícias de jornais etc. Falta padronização do conteúdo do site.	Não há.
Preservação do aparato	Os atributos para a descrição do material estão confusos. Por vezes não há a compreensão se os metadados estão apontando uma gravação ou uma partitura, por exemplo. Muitos qualificadores foram criados para os 15 metadados básicos do Dublin Core.	Textos comentando performances nas matérias.	Metadados descritivos das obras musicais: autor, título, dados da composição, descrição física (n. de páginas), biografia/histórico do item, nota de propriedade.

<sup>23</sup> Por exemplo, de Tom Jobim há um link direcionando para uma página externa <http://portal.jobim.org/jobim/>. Com fotos, documentos, áudios em baixo da capa de um disco (que não funcionam, provavelmente pode ser um problema relacionado aos direitos autorais de reprodução), linha do tempo etc.

	Instituto Antonio Carlos Jobim	Instituto Paulo Moura	Escola de Música da UFRJ
Monitoramento e instrumentalização da comunidade-alvo	Nenhum.	Nenhum.	Vídeos tutoriais para auxiliar o usuário a navegar no acervo.

(Conclusão)

A experiência demonstrou que há certa incompatibilidade entre as iniciativas de preservação e acesso de longo prazo em relação às características da informação musical brasileira. A aplicação das definições mais recentes de informação musical (DOWNIE, 2003 e CRUZ, 2008) à música popular brasileira parece ser uma alternativa interessante para reduzir os problemas citados, considerando-se que ela está presente no dia a dia dos ouvintes, e a descrição desse contexto exige uma abordagem do processo de catalogação de obras musicais que vai além dos registros de metadados oriundos do tratamento intelectual realizado no espaço das bibliotecas.

No entanto, percebe-se um questionamento da comunidade científica sobre o método tradicional de catalogação e um olhar para o *linked data* e os princípios subjacentes à Web semântica (ANTONIOU e VAN HAMERLEN, 2004).

Muitos pesquisadores (COYLE, 2010; SEEMAN e GODDARD, 2015; MURRAY, 2008; MILLER, 2011) consideram que na sociedade da informação apoiada por uso das TICs, a descrição de recursos informacionais deve considerar (i) um processo de registro catalográfico colaborativo, mas sem intercâmbio de metadados, (ii) com priorização das ligações entre os recursos de informação, (iii) realizado não apenas por bibliotecários, e (iv) apoiado por serviços de curadoria digital mais qualificados. Advoga-se aqui que a adoção desses princípios atende ao requisito de descrição do contexto musical e, portanto, potencializa o acesso de longo prazo às músicas brasileiras e os mecanismos de preservação digital como um todo.

Do ponto de vista dos modelos conceituais subjacentes ao processo de catalogação, parece que as alternativas *multi-entity* - que consideram vários conceitos relevantes para o processo descritivo (BAKER, COYLE e PETIYA, 2014) - são mais apropriadas para catalogação de músicas brasileiras do que

os modelos *mono-entity* tradicionais, que consideram apenas um conceito (música) com seus atributos relacionados. Uma característica comum dos modelos *mono-entity* é o foco nas funções do catálogo como produto final do processo e certa ligação com os suportes físicos que armazenam as informações, e não no conteúdo propriamente dito. Em contraposição, os modelos *multi-entity* parecem mais adaptáveis ao esquema de preservação de patrimônio de músicas brasileiras porque consideram o conteúdo musical sem, no entanto, desprezar os suportes materiais vinculados.

Um dos modelos *multi-entity* mais discutidos atualmente é o Functional Requirements for Bibliographic Records (FRBR), publicado pela IFLA (1998), que tem sido considerado adequado para as características da informação musical e também para as demandas subjacentes à preservação de herança cultural imaterial. No FRBR os documentos são descritos considerando sua própria composição e os possíveis relacionamentos, visando facilitar a interação do usuário com o objeto pesquisado. Por ser baseado no modelo ER (Entidades de Relacionamento) – largamente usado em sistemas gerenciadores de bancos de dados relacionais –, o conceito de obra como unidade de catalogação passou a ser associado a termos como “expressão”, “manifestação” e “item”, num único grupo (grupo 1), como está apresentado na figura 4.

Figura 4: Entidades do Grupo 1 do Modelo FRBR



Fonte: Tillet (2004, p. 2)

Além do grupo 1, o FRBR prevê ainda dois outros grupos que englobam conceitos muito apropriados para o registro catalográfico de documentos musicais. No caso do grupo 1, uma obra pode ter diversas expressões e cada expressão pode estar ligada a uma ou mais manifestações. Da mesma forma uma manifestação pode ser descrita por vários itens, num esquema de relacionamento como está ilustrado na figura 4. Um exemplo de aplicação do FRBR para catalogação de uma obra (w), suas expressões (e), manifestações (m) e itens (i) está apresentado na figura 5.

Figura 5 – Exemplo do modelo FRBR aplicado a um registro musical

w1 Variações Goldberg de J.S. Bach

e<sub>1</sub> performance de Glenn Gould gravada em 1981

m<sub>1</sub> gravação liberada em disco de 33 1/3 rpm em 1982 pela gravadora CBS

i<sub>1</sub> cópia mantida para SC

m<sub>2</sub> gravação pré-lançada em CD em 1993 pela Sony

i<sub>1</sub> primeira cópia mantida em HC

i<sub>2</sub> segunda cópia mantida em HC

e<sub>2</sub> performance de Ton Koopman gravada em 1987

m<sub>1</sub> gravação lançada em CD em 1988 por Erato

i<sub>1</sub> cópia mantida em HC

e<sub>3</sub> partitura para piano editada por Ralph Kirkpatrick

m<sub>1</sub> edição publicada em 1938 por G. Shirmer

i<sub>1</sub> cópia mantida em BMC

Fonte: Anderies (2004)

## PECULIARIDADES DA INFORMAÇÃO MUSICAL BRASILEIRA

As estratégias de preservação do patrimônio musical brasileiro devem considerar particularidades das categorias mais amplas (erudita, folclórica e popular) e o repertório associado a essas grandes categorias. Apresentam-se



aqui algumas peculiaridades técnico-musicais e estilísticas que devem ser consideradas para efeito de organização e preservação da informação musical brasileira, considerando os debates travados pela comunidade científica<sup>24</sup>:

- a forma mais acurada da sua representação é a sua *performance* (o som), uma vez que a obra musical existe primeiramente no tempo e é apreendida por uma audiência além de ser considerada existencialmente mais como um conceito abstrato no tempo do que uma entidade física particular no espaço;
- possui uma grande cadeia de produtores (compositores, intérpretes e todas as pessoas dedicadas à produção, distribuição, direitos autorais, gerenciamento etc.) e de usuários e seus variados perfis, desde o muito especialista em uma área bastante específica até um leigo que apenas ouve música;
- requer mais detalhes dos intérpretes nas performances, pois, de modo geral, os usuários requerem informações bastantes específicas para encontrar a música ou a versão da música que desejam ouvir ou executar;
- os catálogos *on-line* não estão adaptados às especificações técnicas dessa informação;
- há mau uso de termos e conceitos que levam a resultados que não correspondem às necessidades dos usuários ou os impedem de encontrar o que procuram. De modo geral, não se observa se os termos empregados na música erudita correspondem ou diferem do seu uso para a música popular ou folclórica;
- ausência de detalhamento – nas regras de catalogação – que são primordiais para materiais voltados para a performance musical, que não são primordiais para outros tipos de recursos como, por exemplo, aqueles que envolvem as especificações técnicas de partituras a serem executadas individualmente ou por diferentes formações instrumentais e/ou vocais;

---

<sup>24</sup> Refere-se aqui aos debates científicos sobre (i) música ocidental europeia (GENTILI-TEDESHI, 2015; GROVE, 2016; SMIRAGLIA, 2002; McLANE, 1996; TREIN, 1986; PACHECO; ALVARENGA, 2012), (ii) música popular norte-americana (TAGG, 1982; KANAI, 2015; GENTILI-TEDESHI, 2015) e (iii) música digital (HEMMASI, 2002; KANAI, 2015).

- limitação do formato de intercâmbio MARC para atender, principalmente a representação da música digital como a falta de estrutura adequada e de metadados administrativos, as limitações vindas dos catálogos *on-line* convencionais (como a de consultar questões musicais específicas), os cabeçalhos multifacetados e pré-coordenados (como é o caso do LCSH), o enfraquecimento das relações entre os elementos de cada obra (como aquelas entre os pontos de acesso) e as conexões ineficientes entre as versões de uma obra;
- faltam critérios bem estabelecidos para o reconhecimento de uma obra musical e suas instanciações (expressões e manifestações) que falham pela falta de análises ontológicas e epistemológicas;
- a diferenciação do arranjo para a música clássica e para a popular e a atribuição de responsabilidade de maneira a facilitar a busca do usuário pelas diversas versões;
- a catalogação de partituras musicais advindas da implementação do RDA: a instrução sobre o tipo de partitura – e aqui considerando aqueles mais comuns à MP – foi fundida com a menção de edição, além de ser dada no campo de notas gerais que, além de dificultar sua recuperação, duplica informações incluídas por vocabulários controlados em outros campos;
- ausência de padronização de termos musicais – principalmente dos gêneros e ritmos brasileiros – e inexistência de obras de referência nacionais, além dos problemas de tradução que falham na correspondência exata dos conceitos (se existem termos correspondentes ou não em outros idiomas) e na consideração dos significados técnicos aplicados ao contexto da MPB;
- pela interação do usuário com a MP que, por sua vez, influencia em sua maneira de buscar a informação da MPB, como por exemplo, pelo fato de ouvir música várias vezes e em diferentes versões; a capacidade de poder cantarolar ou tocar uma versão das músicas que gosta; não conseguir responder sobre qual assunto a música é; e levar em grande importância a data da gravação ou da composição, dos nomes de todos os intérpretes ou a quantidade e especificação dos instrumentos musicais das músicas que fazem parte do seu gênero preferido etc.;

- o problema da descrição bibliográfica fixa e imóvel de dados quando o recurso musical pertencente à MPB é criado em um contexto cultural e social, solicitando um conjunto de informações inter-relacionadas que podem ser vistas a partir de diferentes pontos de vista.

## CONSIDERAÇÕES FINAIS

Neste capítulo foi apresentado um panorama geral com algumas informações sobre a preservação do patrimônio musical brasileiro, em função de sua importância e pela herança cultural associada. Nessa perspectiva, uma das primeiras providências foi assumir a necessidade de identificação do acervo a ser preservado em função da diversidade de músicas no Brasil.

Nessa linha percebe-se que, embora haja divergências sobre estratégias de classificação de músicas, pode-se dizer que a música brasileira refere-se àquela que remete à condição ontológica do povo brasileiro, ou seja, aquela cuja definição vai além dos elementos característicos de uma música promovendo conexões com elementos de raiz, quer sejam os relacionados a situações de ócio, diversão, civismo, protestos e quaisquer eventos ou situações que envolvem a história da sociedade brasileira. Nesse sentido, o músico brasileiro é aquele que consegue se apropriar de um sentimento ontológico nacional e expressa esse sentimento na forma musical. Essa parece ser a música não midiática que tem conexões com o constituinte ontológico de um povo em contraposição àquelas que são geradas em função do estreitamento promovido pela globalização, de caráter mais comercial e que, em geral, promovem euforia (em alguma medida coercitiva), em vez de gerar sentimentos ligados ao constituinte ontológico do povo.

A visão da música enquanto informação permite um olhar mais técnico sobre as características dessa forma de expressão artística e orienta as decisões tecnológicas para garantir a preservação de longo prazo desejada para a música tanto no seu aspecto material quanto imaterial. Nesse sentido, foram traçados argumentos que justificam a necessidade de tecnologias de suporte compatíveis com o patrimônio de herança cultural musical brasileiro, já que esses contemplam especificidades que precisam ser atendidas. Ganham destaque as ideias de uso da catalogação baseada nos princípios da Web semântica e a

adoção de modelos conceituais *multi-entity* que acrescentam o imaterial na sua definição. De fato, o uso de tecnologias digitais como suporte para missões de herança cultural (HC) tem ressaltado a necessidade para sistemas de informação diferentes daqueles que são usados em negócios convencionais, governo e atividades de entretenimento (MURRAY, 2008). Na mesma perspectiva adotada por Nadal (2013) e outros pesquisadores, defende-se que a utilização de modelos conceituais como o FRBR são uma estratégia efetiva de preservação de herança cultural com suporte à preservação e acesso tanto de conteúdos digitais quanto analógicos, e refletem o entendimento habitual das instituições que fazem preservação digital dos recursos que possuem.

Os problemas subjacentes às características da música em geral e às especificidades da música brasileira deixam clara a necessidade de ter profissionais qualificados para cuidar do patrimônio musical brasileiro. Uma possibilidade é a adoção do *music library*<sup>25</sup> - um profissional para auxiliar no cuidado dos bens patrimoniais, em especial da música brasileira, para garantir acesso de longo prazo.

Num cenário ideal, imagina-se uma política nacional de preservação de dados com enfoque para expressões culturais que sejam compatíveis com a riqueza da arte musical brasileira. Além das garantias e diretrizes inerentes a uma política de preservação consolidada, é importante (i) a formação de pessoal especializado para dar suporte às atividades de curadoria digital de músicas nacionais, (ii) o fortalecimento de infraestruturas de preservação, como é o caso da Rede Cariniana e (iii) a sensibilização dos órgãos de preservação de patrimônio histórico sobre a importância de se associarem modelos conceituais como o FRBR, em especial para o caso da preservação e manutenção do patrimônio musical brasileiro<sup>26</sup>.

---

<sup>25</sup> Trata-se de um profissional com formação mista nas áreas de ciência da informação e biblioteconomia e com domínio sobre os processos de curadoria digital de acervos musicais. Mais detalhes podem ser encontrados no endereço <https://www.musiclibraryassoc.org/?page=MusicLibrarianship>.

<sup>26</sup> Nadal (2013) e Murray (2008) defendem o uso do modelo FRBR como pilar essencial para preservação de herança cultural e que pode ser usado em associação com o modelo de referência OAI (Open Archival Information System).

## REFERÊNCIAS

- ALVES, R. C. Música: patrimônio imaterial do Brasil. *Revista Lusofonia*, 2011. Disponível em: <<https://revistalusofonia.wordpress.com/2011/02/16/musica-patrimonio-imaterial-do-brasil>>. Acesso em: 12 jun. 2016.
- ANDERIES, J. Enhancing library catalogs for music. In: CONFERENCE ON MUSIC & TECHNOLOGY IN THE LIBERAL ARTS ENVIRONMENT, 2004. *Electronic proceedings...* [S.l.: s.n.], 2004. Disponível em: <<http://academics.hamilton.edu/conferences/musicandtech/Presentations/Catalog-enhance-ments.ppt>>. Acesso em: 02 jul. 2016.
- ANDERSON, C. *The long tail: why the future of business is selling less of more*. [S.l.]: Editora Hyperion, 2006.
- ANDRADE, M. D. *Ensaio sobre a música brasileira*. 3. ed. São Paulo; Brasília: Martins; INL, 1972.
- ANTONIOU, G.; VAN HARMELEN, F. *A semantic web primer*. Cambridge, Massachusetts (EUA): The MIT Press, 2004.
- BAKER, T.; COYLE, K.; PETIYA, S. Multi-entity models of resource description in the semantic web: a comparison of FRBR, RDA and BIBFRAME. *Library Hi Tech*, v. 32, n. 4, p. 562-582, 2014.
- BÉHAGUE, G. H. Rap, reggae, rock, or samba: the local and the global in Brazilian popular music (1985-95). *Latin American Music Review*, v. 27, n. 1, 2006.
- BLOMBERG, C. Histórias da música no Brasil e musicologia: uma leitura preliminar. *Projeto História*, v. 43, p. 415-444, 2011.
- BOERES, S. A.; MÁRDERO ARELLANO, M. Á. Políticas e estratégias de preservação de documentos digitais, In: ENCONTRO NACIONAL DE CIÊNCIA DA INFORMAÇÃO, 2005, Salvador. *Anais eletrônicos...* Salvador: ANCIB, 2005. Disponível em: <[http://www.cinform-antiores.ufba.br/vi\\_anais/docs/SoniaMiguelPreservacaoDigital.pdf](http://www.cinform-antiores.ufba.br/vi_anais/docs/SoniaMiguelPreservacaoDigital.pdf)>. Acesso em: 14 jul. 2016.
- BYRD, D. *Representation of musical information (lecture notes)*. [S.l.]: Indiana University, 2007. Disponível em: <<http://www.informatics.indiana.edu/donbyrd/INFO545Site-Spring07/I545.html>>. Acesso em: 22 maio 2016.
- CASTAGNA, P. *O estilo antigo na prática musical religiosa paulista e mineira dos séculos XVIII e XIX*. 2000. Tese (Doutoramento) - Faculdade de Filosofia, Ciências e Letras, Universidade de São Paulo, 2000.
- CASTRO, B. et al. BDB-MUS: a project for the preservation of Brazilian musical heritage. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL (ISMIR), 7., 2006, Victoria. *Electronic proceedings...* Victoria, Canadá: [s.n.], 2006. Disponível em: <<http://www.ismir.net>>. Acesso em: 09 out. 2016.

- COTTA, A. G. Fundamentos para uma arquivologia musical. In: COTTA, A. G.; SOTUYO BLANCO, P. (Org.). *Arquivologia e patrimônio musical*. Salvador, EDUFBA, 2006. P. 15-38.
- COYLE, K. *Yes, We can!*: libraries and the semantic web in bibliotheken. Köln, Germany: [s.n.], 2010. Disponível em: <[http://swib.org/swib10/vortraege/swib10\\_coyle.ppt](http://swib.org/swib10/vortraege/swib10_coyle.ppt)>. Acesso em: 12 jun. 2016.
- CRUZ, F. W. *Necessidades de informação musical de usuários não especializados*. 2008. Tese (Doutorado em Ciência da Informação) - Faculdade de Ciência da Informação, Universidade de Brasília, 2008.
- DOWNIE, S. Music information retrieval. *Annual Review of Information Science and Technology*, v. 37, p. 295-340, 2003. Disponível em: <[http://musicir.org/downie\\_mir\\_arist37.pdf](http://musicir.org/downie_mir_arist37.pdf)>. Acesso em: 12 jun. 2016.
- GENTILLI-TEDESCHI, M. Music presentation format: toward a cataloging Babel?. *Cataloging & Classification Quarterly*, v. 53, n. 3/4, p. 399-413, 2015.
- GROVE. *Oxford music online*. 2016. Disponível em: <<http://www.oxfordmusic-online.comsubscriber>>. Acesso em: 12 maio 2016.
- HEMMASI, H. Why not MARC?. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 2002, Paris. *Proceedings...* [S.l.: s.n.], 2002. P. 242-248.
- KANAI, K. Manually identifying the entities of work and expression based on music MARC data: towards automatic identification for FRBRizing OPACs. *Fontes Artis Musicae*, v. 62, n. 2, p. 118-128, 2015.
- LESAFFRE, M. *Music information retrieval: conceptual framework, annotation and user behavior*. 2006. Tese (Doutorado) - Faculty of Arts and Philosophy, Ghent University, Bélgica, 2006.
- MCLANE, A. Music as information. *Annual Review of Information Science and Technology*, v. 31, p. 225-262, 1996.
- MILLER, S. J. *Metadata for digital collections: a how-to-do-it manual*. [S.l.]: Ed. Neal-Schuman, 2011.
- MURRAY, R. J. The FRBR-theoretic library: the role of conceptual data modeling in cultural heritage information system design. In: INTERNATIONAL CONFERENCE ON PRESERVATION OF DIGITAL OBJECTS (IPRES), 1., 2008, London. *Electronic proceedings...* London: [s.n.], 2008. Disponível em: <[www.bl.uk/ipres2008/presentations\\_day1/26\\_Murray.pdf](http://www.bl.uk/ipres2008/presentations_day1/26_Murray.pdf)>. Acesso em: 12 maio 2016.
- NADAL, J. *Print and digital preservation*. 2013. Disponível em: <[http://www.jacobnadal.com/wpcontent/uploads/2013/01/Nadal\\_Chapter29\\_PrintDigitalPreservation.pdf](http://www.jacobnadal.com/wpcontent/uploads/2013/01/Nadal_Chapter29_PrintDigitalPreservation.pdf)>. Acesso em: 14 maio 2016.
- NAPOLITANO, M.; WASSERMAN, M. C. Desde que o samba é samba: a questão das origens no debate historiográfico sobre a música popular brasileira. *Revista Brasileira de História*, v. 20, p. 167-189, 2000.

CAP 7 | Preservação do patrimônio cultural da música brasileira

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA (UNESCO). *Convenção para a proteção do patrimônio mundial, cultural e natural*. Paris, 1972. Conferência Geral da Organização das Nações Unidas para a Educação, a Ciência e a Cultura, 1972, Paris. Disponível em: <<http://portal.iphan.gov.br/uploads/ckfinder/arquivos/Convencao1972%20-%20br.pdf>>. Acesso em: 14 jun. 2016.

PACHECO, K. L.; ALVARENGA, L. Manifestações de obras musicais: o uso do título uniforme. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 8., 2012, Rio de Janeiro. *Anais eletrônicos...* Rio de Janeiro: ANCIB, 2012. Disponível em: <<http://www.eventosecongressos.com.br/metodo/enancib2012/arearestrita/pdfs/19430.pdf>>. Acesso em: 12 jun. 2016.

RUYER, R. *A cibernética e a origem da informação*. Rio de Janeiro: Paz e Terra, 1972.

SACKS, O. *Alucinações musicais: relatos sobre a música e o cérebro*. [S.l.]: Ed. Companhia das Letras, 2007.

SANDRONI, C. MPB: um pouco de história. *Revista Cult*, 2010. Disponível em: <<http://revistacult.uol.com.br/home/2010/03/mpb-um-pouco-de-historia/>>. Acesso em: 22 jun. 2016.

SAYÃO, L. F. Preservação digital no contexto das bibliotecas digitais: uma breve introdução. In: MARCONDES, C. H. et al. (Org.). *Bibliotecas digitais: saberes e práticas*. Brasília: IBICT, 2005.

SEEMAN, D.; GODDARD, L. *Preparing the way: creating future compatible cataloging data in a transitional environment*. Canada: Memorial University Libraries, 2015. Disponível em: <<http://dx.doi.org/10.1080/01639374.2014.946573>>. Acesso em: 20 jun. 2016.

SMIRAGLIA, R. P. Music works and information retrieval. *Notes*, v. 58, n. 4, p. 747-764, 2002.

SOTUYO-BLANCO, P. O patrimônio musical na Bahia: um recenseamento arquivístico musical preliminar. *Revista da Bahia*, v. 32, n.39, p. 49-58, 2004.

TAGG, P. Analysing popular music: theory, method and practice. *Popular Music*, v. 2, p. 37-65, 1982.

TREIN, P. *A linguagem musical*. Porto Alegre: Mercado Aberto, 1986.

YEUNG, T. A.; CARPENDALE, S.; GREENBERG, S. Preservation of art in the digital realm. In: INTERNATIONAL CONFERENCE ON PRESERVATION OF DIGITAL OBJECTS (IPRES), 1., 2008, London. *Electronic proceedings...* London: [s.n.], 2008. Disponível em: <[www.bl.uk/ipres2008/presentations\\_day1/06\\_Au%20Yeung.pdf](http://www.bl.uk/ipres2008/presentations_day1/06_Au%20Yeung.pdf)>. Acesso em: 12 maio 2016.

## ***Preservation of cultural heritage of Brazilian music***

### **ABSTRACT**

*It is possible to see Brazilian music as a heritage because it is part of the culture of the people and goes beyond the limits of concrete. It is thus legitimated as an expression that gains character of tradition and is contextualized in its time promoting subjective records of the current social and cultural organization. Despite its importance, it can be seen that there is not yet a policy established in Brazil to guarantee the formal procedures and guidelines for the preservation of Brazilian music heritage, especially with regard to technological aspects. Brazilian music, as a promoter of immaterial cultural heritage, requires compatible techniques and procedures that go beyond traditional information organization initiatives. In this chapter (i) the plurality and diversity of Brazilian music is discussed, seeking to highlight it in relation to other types of music, and (ii) some alternatives and suggestions that may be relevant to guarantee the preservation and access to this intangible heritage by future generations*

**Keywords:** *Brazilian musical heritage. Digital preservation of music. Musical Technologies. Organization of musical information.*



## ***Preservación del patrimonio cultural de la música brasileña***

### **RESUMEN**

*Es posible ver la música brasileña como un patrimonio porque forma parte de la cultura del pueblo y sobrepasa los límites del concreto. Se legitima, por lo tanto, como expresión que gana carácter de tradición y se contextualiza en su época promoviendo registros subjetivos de la organización social y cultural vigente. A pesar de su importancia, se percibe que no hay todavía una política establecida en Brasil para garantizar los procedimientos formales y directrices de preservación del patrimonio de músicas brasileñas, en especial en lo que se refiere a aspectos técnicos. La música brasileña, como promotora de herencia cultural inmaterial, exige técnicas y procedimientos compatibles y que van más allá de las iniciativas tradicionales de organización de la información. En este capítulo (i) se discute la pluralidad y la diversidad de la música brasileña, buscando destacarla en relación a otros tipos de música, y (ii) se presentan algunas alternativas y sugerencias que pueden ser relevantes para garantizar la preservación y el desarrollo el acceso a ese patrimonio inmaterial por generaciones futuras.*

**Palabras clave:** *Patrimonio musical brasileño. Preservación digital de canciones. Tecnologías musicales. Organización de la información musical.*

## La preservación digital y la Red Cariniana

Miguel Ángel Márdero Arellano<sup>1</sup>

### RESUMEN

Las actividades prácticas de la preservación de documentos digitales tienen como característica principal la colaboración entre instituciones y especialistas. La constatación del grado de vulnerabilidad de ese tipo de material y la necesidad de que sean establecidas responsabilidades entre creadores y curadores determinará el tiempo de vida de los soportes físicos de la información en formato digital. La ciencia necesita de la preservación de los resultados de sus investigaciones para que puedan ser utilizados por otros científicos en el futuro; existe también la necesidad de salvar el patrimonio cultural digital, para garantizar el acceso a esos recursos con funcionalidades de registros electrónicos auténticos. La preservación de la información en forma digital necesita de un conjunto de prácticas técnicas y de gestión que cambian constantemente. Los repositorios digitales son una parte importante de la preservación digital y con ellos puede establecerse un grado de control de los procedimientos de creación, conservación y recuperación a largo plazo. La adopción de un modelo de red de preservación digital es una alternativa para las organizaciones que quieren coleccionar, almacenar, preservar y ofrecer acceso a su acervo en copias digitales. También es imprescindible la concordancia de esas redes con las normas internacionales ya probadas y que promueven la conservación digital de la producción científica a largo plazo. Este trabajo trata sobre el modelo de gestión de la preservación digital en red distribuida que tiene la característica de usar por lo menos seis instituciones donde se guardan copias de documentos digitales. Se presenta una experiencia regional de preservación digital distribuida, que sigue el modelo de una alianza internacional para salvaguardar la producción intelectual de un país. El modelo de negocios es de gestión de información en acceso libre, donde la perpetuidad de la recuperación de los datos está garantizada por la forma abierta de administrar las etapas de entrada, conservación y difusión en sistemas electrónicos de información confiables.

**Palabras clave:** Preservación digital distribuida. LOCKSS. CARINIANA. Red colaborativa. IBICT. Brasil.

---

<sup>1</sup>Coordinador de la Red Brasileña de Servicios de Preservación Digital — CARINIANA. IBICT. Brasil.

## INTRODUCCIÓN

En el avance vertiginoso de la tecnología, propiciado por el crecimiento del conocimiento científico, las sociedades contemporáneas esperan contar con el dominio de los recursos digitales, para garantizar el acceso a largo plazo de grandes colecciones de información. Esa dependencia tecnológica nos coloca frente al riesgo de perder los registros materiales de la investigación racional.

Al analizar el ciclo de la información (generación, tratamiento, preservación y conservación y, finalmente difusión de la información), se puede percibir que es necesario rever la preservación en el ámbito de la información digital.

Las técnicas de preservación que eran conocidas por las bibliotecas y los centros de información se modificaron y encontraron nuevas prácticas en un contexto de red de computadoras, en el cual la información no está solo en un lugar. La preservación de la información en forma digital necesita un nuevo conjunto de prácticas técnicas y de gestión que cambian constantemente.

Los documentos digitales representan la fijación de datos en unidades lógicas generadas por un procesador, su forma, contenido y soporte físico son variables que deben tomarse en cuenta cuando se establece su valor de uso y permanencia. Preservar los documentos digitales se ha convertido en una experiencia singular debido al esfuerzo que demanda mantener las características originales de esos materiales y su capacidad de servir como otro medio de comunicación humana.

En la gestión de la información, la preservación digital comprende todas las actividades que los contenidos representados digitalmente necesitan para asegurar su accesibilidad, almacenamiento a largo plazo e interpretación.

La preservación de la información científica, tecnológica y cultural a largo plazo es un asunto importante en el tratamiento y organización de bases de datos, repositorios y bibliotecas digitales, los cuales subsidian la productividad intelectual, estructurando datos sobre áreas de conocimiento, investigadores e instituciones en varios contextos.

La ciencia necesita la preservación de los resultados de sus investigaciones para que ellos puedan ser utilizados por otros científicos en el futuro. Llevando en consideración algunas resoluciones de la UNESCO (2003), como

“La Carta sobre la Preservación Digital” y “Las Recomendaciones sobre Software Libre para Repositorios y Sistemas de Preservación” (2007), puede afirmarse que existe también la necesidad de salvar el patrimonio cultural digital, para garantizar el acceso a esos recursos con funcionalidades de registros electrónicos auténticos.

El desafío de la gestión de la preservación a largo plazo es llevar en consideración todos los tipos de contenidos y formatos de objetos digitales. Este trabajo trata sobre el modelo de gestión de la preservación digital en red distribuida que tiene la característica de usar por lo menos seis instituciones donde se guardan copias de documentos digitales. También, se presenta la experiencia del IBICT, que con la creación de la Red Brasileña de Servicios de Preservación Digital (CARINIANA) adoptó el sistema de preservación digital de la Alianza LOCKSS para preservar los periódicos nacionales que usan la plataforma OJS, así como las tesis, los libros electrónicos, dos documentos archivísticos y los datos de investigación que están almacenados en repositorios digitales como, DSpace, Dataverse y Archivematica.

## **LOS PRIMEROS PASOS DE LA GESTIÓN DE LA PRESERVACIÓN DIGITAL**

El efecto del constante cambio tecnológico puede verse claramente con las editoras de periódicos científicos, que iniciaron la transferencia de sus acervos editoriales de una configuración analógica para una generación de tecnología computacional. A partir de eso, la migración periódica de materiales digitales ha asegurado la habilidad de los usuarios para recuperarlos y exponerlos. En el paso del tiempo, los contenidos publicados han sido copiados, recreados y actualizados con migraciones para nuevos formatos, en nuevas generaciones de software y hardware.

Los modelos de preservación comienzan a surgir con la finalidad de que los futuros usuarios puedan entender el ambiente tecnológico adoptado por las instituciones que producen hoy información científica. El soporte en papel sigue siendo objeto de prácticas de preservación, que garantizan la integridad física del documento, diferente del ambiente digital, donde la preocupación está en la generación, gestión y acceso al archivo electrónico, con el cumplimiento de normas que garanten su integridad intelectual.

La prevalencia de productos comerciales como los de la Microsoft y de la Adobe en el área de la producción de artículos científicos aumenta solamente el número de versiones de software y el riesgo de corrupción de datos y de las pérdidas esenciales de la información. Los asuntos relacionados con la preservación digital no están siendo tratados por las editoras de forma estratégica, ellas atribuyen esa preocupación a las bibliotecas, las cuales manifiestan más preocupación con la colecta y la inserción de esos contenidos en los sistemas actuales.

El trabajo de las bibliotecas en la preservación de las publicaciones periódicas envuelve la transferencia de documentos en formatos validos en el momento de sumisión en un repositorio digital. Pero las herramientas para repositorios digitales no fueron originalmente destinadas a proyectos de preservación digital o tienen como objetivo principal garantizar la longevidad de los contenidos digitales.

El cambio en el sistema de publicación actual en repositorios de acceso abierto puede llegar a forzar a los editores a que se tornen proveedores de servicios de evaluación por los pares y certificadores de conocimiento científico, en lugar de ser proveedores de acceso a productos digitales, cediendo así la misión de preservar las versiones originales a los gestores de los repositorios institucionales.

## EL ARCHIVO ELECTRÓNICO

El estatus de los sistemas de almacenamiento de contenidos digitales en el mundo ya está siendo acompañado por una iniciativa coordinada por el Centro Internacional ISSN. El servicio The Keepers Registry ofrece reportes globales de las organizaciones que realizan preservación a largo plazo de contenidos de periódicos electrónicos. La iniciativa cuenta con el apoyo de informaciones precisas registradas en los Centros Nacionales de ISSN, con los cuales el servicio The Keepers puede con, solamente 20% de las publicaciones seriadas están siendo archivadas de forma segura en alguna iniciativa de preservación digital.

Figura 1 — Página Web da Iniciativa The Keepers Registry

**The Keepers Registry**  
Supporting long-term access to journal content

Home | Journals | Publishers | Archiving Agencies | FAQ | About | Help

Discover who is looking after your e-journals

**Search**

Input a search term (e.g. free text, or a title or ISSN(s))

[Display example searches](#)  
[Glossary and definitions](#)

The following organisations are the Keepers of digital content, working on your behalf to ensure long-term access to the scholarly and cultural record. They provide the registry with information on their archival holdings, ordered by most recent update (date of which is unknown).

- [Library of Congress](#) (2 July 2014)
- [Archaeology Data Service](#) (1 July 2014)
- [British Library](#) (1 July 2014)
- [HathiTrust](#) (1 July 2014)
- [Scholaria Portal](#) (1 July 2014)
- [Perseus](#) (27 June 2014)
- [CLOCKSS Archive](#) (23 June 2014)
- [Global LOCKSS Network](#) (23 June 2014)
- [National Science Library, Chinese Academy of Sciences](#) (8 January 2014)
- [e-Depot](#) (24 April 2012)

The most up-to-date information may be available on the Keepers website.

**New!**  
[Sign up to test the new Member Services](#)

**Current Statistics**

- Serials with volumes reported as 'Ingested and archived' by at least one Keeper: 24,176
- Serials with volumes reported as 'Ingested and archived' by three or more: 9,956

Work on other serials (and volumes) is 'in progress'.

**News and Events**

**20th March 2014:** Data from the new keepers [\[View\]](#)

**2nd December 2013:** Library of Congress Joins The Keepers Registry [\[View\]](#)

**28th October 2013:** Generating Actionable Evidence on E-journal Archiving [\[View\]](#)

**25th October 2013:** Further International Exposure for The Keepers Registry [\[View\]](#)

**28th August 2013:** The Keepers Registry in Singapore [\[View\]](#)

**15th May 2013:** Two new Keepers report e-journal archiving [\[View\]](#)

**1st May 2013:** Tales from The Keepers Registry [\[View\]](#)

See the [The Keepers Registry blog](#) for archived news and our development roadmap.

Fuente: <https://thekeepers.org/>

Cumplir con los criterios de inclusión en una iniciativa como The Keepers permite que una red de preservación digital colaborativa y distribuida, contribuya con el mantenimiento de servicios de acceso a la información de bibliotecas y de publicaciones que están en peligro de perder sus contenidos digitales.

Los periódicos brasileños de acceso abierto ya aparecen impulsados por servicios como el LOCKSS, CLOCKSS, e-Depot y ahora también por la Red Cariniana, la cual fue incluida como la más nueva agencia de archivo digital registrada en el The Keepers.

## LA PRESERVACIÓN DIGITAL DISTRIBUIDA

Según Howell (2000), los métodos de preservación digital se encuentran en su primera fase, por lo que no se puede afirmar qué es lo que va a funcionar o no en esa área. Por otro lado, la velocidad de la producción de archivos digitales no está siendo alcanzada por los proyectos de preservación digital, algunos de ellos muy exhaustivos.

Las estrategias adoptadas por instituciones internacionales engloban el uso y elaboración de normas, establecimiento de consorcios, elección de metadatos específicos, aplicación de un método operacional o conjunto de acciones operacionales sobre los objetos digitales y el montaje de una infraestructura tecnológica, recomendando la centralización de los proyectos de preservación en aquellos contenidos que son de importancia vital para la institución o grupo de instituciones.

Varios estudios se enfocan en el modelo de preservación digital en redes colaborativas, apuntando para las necesidades de los usuarios actuales y del futuro que deberán contar con materiales auténticos y certificados por instituciones reconocidas.

La adopción de un modelo de red de preservación digital es una alternativa para las organizaciones que quieren coleccionar, almacenar, preservar y ofrecer acceso a su acervo en copias digitales autorizadas. También es imprescindible la concordancia de esas redes con las normas OIAS e ISO 16363 ya probadas y que promueven el archivo digital de la producción científica a largo plazo.

Las redes colaborativas están basadas en la perspectiva DDP (Distributed Digital Preservation), que propone la creación de copias distribuidas de archivos digitales en computadores geográficamente dispersos, para aumentar sus posibilidades de sobrevivencia. Las redes de preservación digital distribuidas consisten en múltiples lugares de preservación, que según Skinner (2010, p. 12-13) deben seguir los siguientes principios:

1. Los locales que preservan el mismo contenido no pueden estar a una distancia menor de 120 y 200 kilómetros entre ellas;
2. Los locales de preservación deben estar distribuidos fuera del área de transcurso de desastres naturales;
3. Los locales de preservación deben estar distribuidos a través de diferentes redes de energía eléctrica;
4. Los locales de preservación deben estar bajo el control de diferentes administradores de sistemas;

5. El contenido preservado en lugares dispersos debe estar en medios electrónicos actualizados y su funcionalidad verificada regularmente;
6. Se deberá replicar su contenido por lo menos tres veces siguiendo los principios ya citados.

Las redes de preservación digital distribuida comenzaron a surgir en los últimos años y específicamente un software emergió observando los seis criterios mencionados: el software LOCKSS (“Lots of Copies Keeps Stuff Save”) que permite la creación de las llamadas Redes Privadas LOCKSS (PLNs). En países de América del Norte, Europa y Asia<sup>2</sup> ya existen PLNs que preservan múltiples copias de contenidos digitales creados localmente por organizaciones culturales y de investigación científica.

## LA PRESERVACIÓN DIGITAL DISTRIBUIDA Y LA ALIANZA LOCKSS

Una preocupación actual en las bibliotecas de todo el mundo es la preservación del acceso al contenido de sus colecciones de periódicos digitales. La inscripción a bases de datos de periódicos es cara, y no garantiza el acceso una vez cancelado el contrato o cuando el periódico deja de ser publicado. Las bibliotecas están permitiendo que el contenido de los periódicos que antes se guardaban en acervos físicos, ahora esté accesible de forma remota en el servidor del proveedor de datos en Internet.

En la última década, muchos proyectos se desarrollaron con el objetivo de preservar documentos digitales a través de iniciativas que generaron conceptos y estrategias, contribuyendo a la identificación de problemas y a la creación de soluciones.

En varios países ya es un hecho la necesidad de solucionar urgentemente la preservación de periódicos electrónicos, cuya solución depende de la cooperación entre bibliotecas y editoras. Algunas de las acciones emprendidas resultaron en la formalización de una iniciativa internacional de cuño colaborativo y sin fines lucrativos, cuya misión es “crear herramientas

---

<sup>2</sup> Lista de Redes Privadas LOCKSS: [http://www.lockss.org/lockss/Private\\_LOCKSS\\_Networks](http://www.lockss.org/lockss/Private_LOCKSS_Networks)  
Recuperado en 28/06/2013.



y ofrecer soporte a las bibliotecas de investigación para que puedan, de una forma fácil accesible, crear, preservar y archivar colecciones electrónicas locales” (REICH, 2002).

Entre los proyectos de preservación en redes distribuidas, el proyecto LOCKSS se destaca cuando se lleva en consideración el bajo coste de su instalación y manutención. La preservación digital que el LOCKSS permite, no se reduce al acceso perpetuo de los materiales almacenados, mas también “se extiende a las funciones que garantizan la preservación intelectual de los documentos en cualquier forma digital.” (MÁRDERO ARELLANO, 2008).

La herramienta LOCKSS es un software de código abierto, que trabaja en sistema “peer-to-peer” permitiendo que unidades de información puedan colectar, guardar y preservar localmente material digital autorizado. Son copias de seguridad que pueden ser accesibles cuando la página de la revista no está disponible. De esa forma las instituciones que detienen los derechos de acceso a esas publicaciones pueden realizar actividades de custodia de documentos electrónicos en servidores locales.

Con el software LOCKSS las computadoras de uso doméstico se convierten en aparatos de preservación digital, denominados “cajas LOCKSS”. Esas cajas se pueden definir como los locales donde se almacenan (salvaguardados) los contenidos colectados directamente del sitio de la editora en Internet. En ellas se reproducen los contenidos publicados en Internet en su forma original los cuales se distribuyen para otras cajas de la misma red. Ninguna de las instituciones que forman la red de cooperación de recursos de información digitales descarta los registros (LOCKSS, 2008).

La Alianza LOCKSS (Lots of Copies Keep Staff Safe) creada por la Stanford University Libraries tiene como objetivo preservar el acceso a los periódicos publicados en Internet, ya sean de acceso libre o de editoras privadas. Bibliotecas y editoras de todo el mundo se han adherido a LOCKSS para trabajar juntos en la preservación del acceso autorizado a periódicos pago y gratuitos.

Establecida en 2004 y usada por bibliotecas y editoras de todo el mundo, la Alianza LOCKSS provee una infraestructura coordinada de desarrollo de software y gerenciamiento de colecciones digitales. La Alianza mantiene

dos tipos de red: la Red Global y las Redes Privadas. La primera se encarga de preservar libros y periódicos de más de 510 editoras comerciales, y el segundo tipo formada por más de 12 redes locales, regionales o temáticas preserva repositorios institucionales, documentos oficiales de gobierno, bases de datos, contenido de acceso libre, etc.

En instituciones de investigación de Europa, Asia y América del Norte, el sistema de preservación digital LOCKSS fue adoptado por tratarse de un estándar ISO (14721:2003), significando su consonancia con otras iniciativas que han comprobado el valor científico del sistema, como estándar internacional de preservación digital (UNESCO, Harvard University, Stanford University, M.I.T, y la Library of Congress, entre otras).

El modelo de preservación digital distribuida asumido por la Iniciativa LOCKSS de la Stanford University apoya la creación de redes de colaboración que guardan colecciones distribuyéndolas en diversas instituciones.

El modelo distribuido hace uso de la infraestructura local para garantizar que por lo menos una de las instituciones participantes tenga la última versión de un periódico, libro o tesis que fue copiado del original. El LOCKSS ejecuta cuatro funciones de gestión de la información importantes:

1. Coleccionar los contenidos provenientes de las publicaciones electrónicas usando un rastreador de Internet parecido con los que usan por los mecanismos de búsqueda.
2. Verificar periódicamente si el contenido almacenado tiene coherencia informacional.
3. Ofrecer páginas de Internet a donde pueden encontrarse los documentos preservados.
4. Hacer disponible una interface administrativa que permita que el administrador seleccione nuevos contenidos que puedan preservarse, con el monitoreo de los contenidos salvaguardados y controlando su acceso.
5. Migrar el contenido cuando sea necesario y entregar el documento al lector en un nuevo formato.

La herramienta LOCKSS permite la formación de una comunidad de bibliotecas comprometidas con el desarrollo de colecciones digitales. Cuando una red LOCKSS se establece, es indispensable la participación de los profesionales de la información, capacitados para administrar la colección preservada localmente y de esa forma, restituyen a la biblioteca su papel de local de custodia de la información.

## **LA PRESERVACIÓN DIGITAL DISTRIBUIDA EN LA RED CARINIANA**

En Brasil, la creación de sistemas intermedios de información pasó a ocupar un lugar estratégico en el momento en que el Estado asume la posición de agente principal en la gestión de programas, recursos y servicios de información y documentación. Esa postura hizo necesaria la construcción de infraestructura técnico científica, sumándola a la capacitación de personal calificado y a la producción, acceso y preservación de las informaciones en ciencia y tecnología. El gobierno de Brasil incumbió al Instituto Brasileño de Información en Ciencia y Tecnología (IBICT) la vocación para ser un núcleo de competencia en los procesos de tratamiento, acceso y difusión de esa información.

El IBICT actuando en sintonía con los cambios socio-técnicos contemporáneos en el campo de la información, promoviendo la competitividad y el desarrollo de recursos e infraestructura de información científica y tecnológica fuera y dentro de la institución, inició acuerdos de cooperación con instituciones de investigación en América Latina, con el propósito de formar una red colaborativa de servicios de preservación digital y de profesionales capacitados para el registro en forma digital de acervos culturales y científicos.

La preservación digital de largo plazo es un área estratégica para el IBICT. En 2012 el instituto convocó a cinco universidades públicas para formar parte de una red colaborativa, cuyo objetivo era su participación en un proyecto piloto de preservación de periódicos electrónicos, con una duración de seis meses.

En enero de 2013 se firmó el acuerdo entre LOCKSS y el IBICT; con el apoyo financiero de la institución brasileña de fomento a la investigación científica FINEP, la adquisición tuvo la particularidad de ser por inexigibilidad, pues el software era el único en el mercado dedicado exclusivamente para realizar la preservación de documentos digitales de forma distribuida.

Para consolidar el proyecto de construcción de una red colaborativa de preservación digital, la Alianza LOCKSS se comprometió a dar el soporte técnico necesario para la preservación de los periódicos electrónicos brasileños de acceso libre que usan la plataforma OJS, de las tesis y disertaciones y de los libros electrónicos que están en herramienta DSpace. Las instituciones participantes cuentan con el soporte por parte del LOCKSS Technical Support Service (LTSS) a través del equipo de la Red CARINIANA en el IBICT, el cual, provee el apoyo técnico, la capacitación y el asesoramiento en la negociación de los permisos con los editores y responsables por las revistas y por los portales de revistas en el OJS.

Un área importante dentro del proyecto de la Red CARINIANA es la provisión de servicios de preservación de tipos de documentos digitales usados por las instituciones participantes. Algunas cláusulas del acuerdo de cooperación que las instituciones participantes firmaron con el IBICT, incluyen determinaciones sobre el almacenamiento y el uso de los materiales guardados en las cajas LOCKSS. La etapa piloto permitió el intercambio de ideas sobre las cláusulas del acuerdo que incluyen las especificaciones del contenido almacenado, su obtención y acceso futuro.

El modelo de licencia de los materiales preservados en la Red CARINIANA es el de acceso libre, y la decisión de establecer esa red basada en la iniciativa LOCKSS fue el resultado de una serie de actividades de investigación previas, que incluyeron, análisis técnicos de software de preservación digital, de contactos con responsables por otras redes LOCKSS y la participación de representantes de centros de computación y de bibliotecas de cinco instituciones de enseñanza superior brasileñas: USP, UNICAMP, UFSM, UFPB y UEMA<sup>3</sup>.

---

<sup>3</sup>Universidade de São Paulo, Universidade Estadual de Campinas, Universidade Federal de Santa Maria, Universidade Federal da Paraíba, Universidade Estadual do Maranhão.

La Red CARINIANA tiene como objetivo salvaguardar los registros de la ciencia, tecnología y del patrimonio cultural del Brasil. A través de ella se crearán alternativas para que las instituciones puedan coleccionar, almacenar, preservar y promover el acceso al contenido seleccionado a través de copias autorizadas. También, la Red ofrecerá paquetes de software, aplicativos y ambientes multimedia desarrollados e implantados para preservar documentos digitales.

Figura 2 — Página principal del Portal de la red Cariniana



#### SOBRE

Início

IBICT

Preservação Digital

Allança LOCKSS

#### Parcerias da Rede

Parceiros Integrais

Parceiros Institucionais

Logos p/ Participantes

#### SERVIÇOS

### XV Congresso de Arquivistas de Catalunha - Espanha



Fuente: <http://cariniana.ibict.br/>

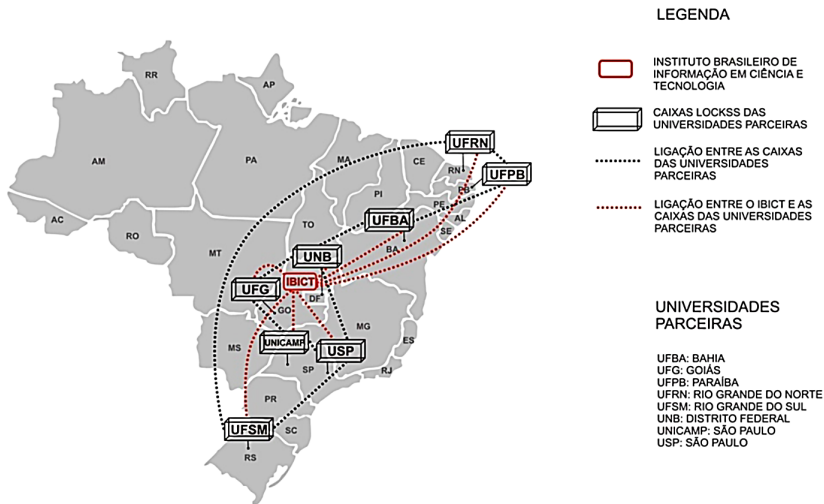
Las instituciones contactadas ya cuentan con documentos digitales en sistemas de información de acceso libre que el IBICT recomienda, por tanto, con la Iniciativa LOCKSS, el Instituto encontró mecanismos de integración de esos sistemas de información que usa y disemina, tales como la Biblioteca Digital de Tesis y Disertaciones, el Open Journal Systems, el Open Conference Systems, el Open Monograph Press, el Open Harvester y el DSpace.

Uno de los principales objetivos alcanzados en la etapa piloto fue el de concienciar a todos los participantes sobre la iniciativa LOCKSS. El IBICT creó materiales de apoyo para los responsables técnicos de las cajas de cada institución, tanto para coleccionar, informar los metadatos necesarios para el registro de los periódicos, como para poder instalar y administrar la interface técnica del software.

Con eso, se consolidó la etapa gerencial de la Red, la cual incluye las acciones de insertar, coleccionar y monitorear localmente los datos almacenados en cada caja LOCKSS. El alcance de la red se puede visualizar en la siguiente figura.

Figura 3 — El alcance de la red CARINIANA

#### CARINIANA: REDE BRASILEIRA DE PRESERVAÇÃO DIGITAL



Fuente: el autor

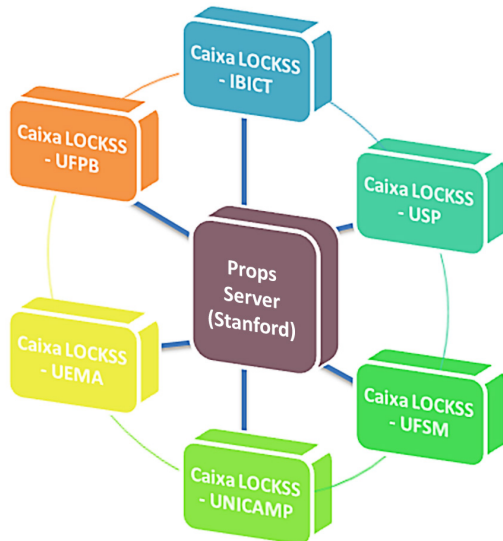
Las definiciones técnicas de funcionamiento de la Red recibidas por el equipo del IBICT sirven para estructurar el conocimiento necesario en la formación de especialistas en lengua portuguesa, que de alguna forma beneficiará a la comunidad internacional de usuarios del LOCKSS.

Para el funcionamiento sustentable de la Red, se estableció, en acuerdo de cooperación, que cada institución participante será responsable por la adquisición y configuración local del equipamiento, lo que podrá subsidiarse a través de negociaciones con fondos de financiación de investigación nacionales.

En la etapa piloto, se realizaron testes preliminares y análisis de las actividades posteriores a la instalación del software. Los representantes técnicos de cada institución recibieron capacitación vía Internet (video conferencias y chats online), además de visitas y reuniones técnicas. Se usaron todas las ediciones de 16 periódicos electrónicos de acceso libre que utilizan la plataforma OJS en el primer envío de contenido para todas las cajas LOCKSS de la Red.

En la etapa actual, la estructura de la Red cuenta con un servidor web localizado en la Stanford University, responsable por la verificación de los metadatos de las publicaciones a ser repasadas a las “cajas” de las instituciones consorciadas. Ese servidor cumple el papel de núcleo central o Props Server, que no interfiere con la administración de la Red.

Figura 4 — Red Privada Local (PLN) de periódicos electrónicos

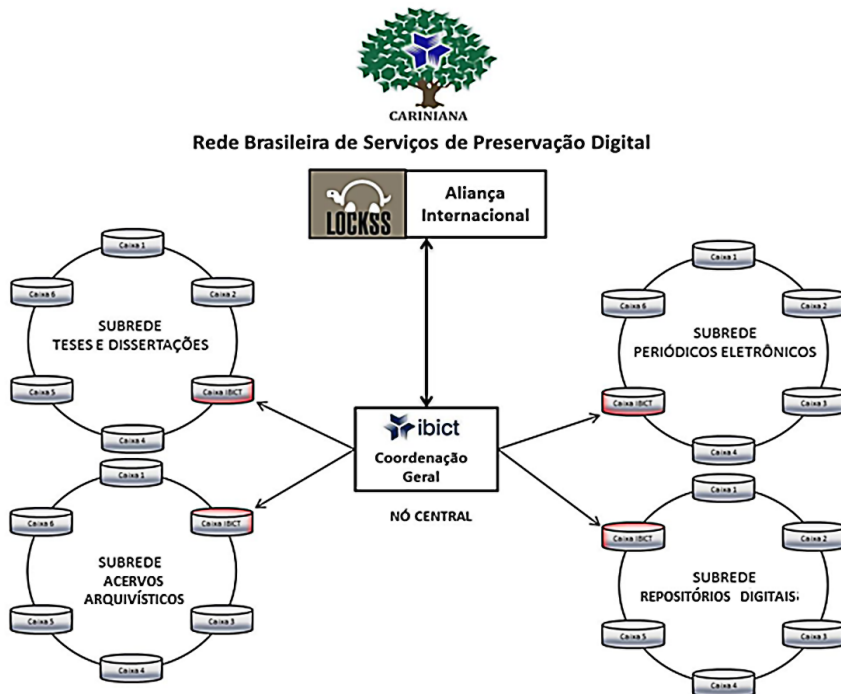


Fuente: el autor

Entre las funciones de las nueve instalaciones de las cajas existentes hasta hoy, además de almacenar, preservar y corregir daños a los contenidos, están las de comparar continuamente el contenido colectado, actuar como web proxy o cache para acceso en casos de emergencia y controlar el acceso a los periódicos preservados.

La red Cariniana archiva de forma distribuida más de 600 títulos de periódicos, acumulando más de 5 mil volúmenes y casi 11.000 volúmenes en procesamiento. En 2016 se incluirá la implantación de la sub-rede de repositorios digitales y la sub-red de tesis y disertaciones que usan la herramienta DSpace. La siguiente figura representa la formación de las sub-redes que serán parte integrante de la Red.

Figura 5 – Las sub-redes de la red CARINIANA



Fuente: <http://cariniana.ibict.br/>



Cabe mencionar que todo el proceso de preservación de periódicos electrónicos cuenta también con la colaboración de iniciativas internacionales, como la del Public Knowledge Project (PKP) y de la Stanford University. El IBICT coordina el Centro nacional del ISSN y también el Latindex.

Las cooperaciones de la Cariniana son una fuente de recursos de información de valor área la Red. Las actuales nueve instituciones cooperantes comparten los mismos términos de acuerdos de cooperación y cumplen el plan de actividades establecido para los años 2012 a 2018. La garantía de funcionamiento de la Red está descrita en su política y en las actividades de acompañamiento e investigación de sus colaboradores y equipos técnicos.

La Cariniana realiza dos encuentros anualmente con los representantes de las instituciones participantes, así como con sus líderes de su grupo de investigación, dividido en diez líneas de investigación registrados oficialmente en el Directorio de Grupos de investigación de una de las más importantes instituciones de fomento CNPq.

Figura 6 — Página del Directorio de Grupos de Investigación CNPq



Fuente: <http://lattes.cnpq.br/web/dgp>

## LA PRESERVACIÓN DIGITAL DISTRIBUIDA EN AMÉRICA LATINA

En el ámbito de América Latina, es necesaria la definición de políticas, obligaciones y metodologías más apropiadas a la preservación de documentos electrónicos, tomando en consideración la implementación del modelo de red distribuida, para que de esa forma se pueda verificar si los mismos atienden a las expectativas de los usuarios y de las instituciones de investigación nacionales.

La incorporación de instituciones de varios países de la región, posibilitaría la formación y capacitación dirigida a profesionales con intereses en diferentes aspectos de la preservación digital. El intercambio de experiencias en el uso de plataformas informáticas de código libre serviría para explorar de forma cooperativa los temas cruciales para los proyectos locales.

El modelo de preservación digital distribuida es importante para países en desarrollo, donde la mayoría de las instituciones carece de financiación para las actividades relacionadas con la preservación de sus archivos digitales. En países como los de América Latina, las redes distribuidas representan la posibilidad de que los documentos de acceso libre estén siempre disponibles y que cuenten con el apoyo de especialistas nacionales e internacionales.

## CONCLUSIONES

La preservación digital se puede considerar como un conjunto de actividades complejas, que requiere de años de planeamiento y gastos inmensurables. Con las soluciones tecnológicas orientadas a la creación de redes colaborativas esa percepción puede cambiar. La adopción de software de preservación digital para bibliotecas académicas, debe entenderse como el paso inicial para que solucione el problema de la salvaguarda de los documentos digitales.

El software LOCKSS es una iniciativa de extrema relevancia en el contexto tecnológico actual, principalmente para las áreas de la Ciencia de la Información, pues se presta a diversas actividades ejercidas en bibliotecas tradicionales y digitales. Al viabilizar el archivo, la preservación y asegurar el acceso a ese

tipo de registro, el sistema se muestra como aliado de los profesionales de la información, en lo que abarca la solución de cuestiones relacionadas con la gestión de documentos electrónicos.

Con el LOCKSS en Brasil, se abrió la posibilidad de que los profesionales de la información aplicasen las teorías sobre la preservación digital, iniciando ese trabajo con la creación de redes que preservarán informaciones de gran importancia, producidas actualmente en el país, para que las mismas puedan ser accesibles por generaciones futuras, tal como lo hacían los bibliotecarios de la antigüedad.

La red Cariniana intenta llamar la atención hacia la importancia de las prácticas de preservación digital distribuida, principalmente para que las instituciones participantes puedan contar con una opción más de archivo digital y que puedan desenvolver habilidades necesarias para gestionar sus propias redes distribuidas.

Una de las metas de la red Cariniana es promover el uso de herramientas de preservación digital en el Brasil, con las cuales se haga posible preservar colectivamente un número considerable de publicaciones electrónicas nacionales de acceso libre.

En países como los de América Latina es necesario que los profesionales envueltos en las actividades de preservación tomen conciencia de sus responsabilidades, para la implementación de iniciativas colaborativas a gran escala y, se creen redes de cooperación de salvaguarda de documentos digitales de ámbito nacional.

## RECONOCIMIENTOS

Este artículo está basado en la presentación de la ponencia para la III Conferencia Bibliotecas y Repositorios Digitales de América Latina (BIREDIAL 2013) que tuvo lugar en San José de Costa Rica, del 15 al 17 de Octubre de 2013. Está basada también en una serie de conversaciones con colegas de la Red Brasileña de Servicios de Preservación Digital – CARINIANA, entre ellos, María de Fátima Tavares, Rodrigo Rabello y Ramón Fonseca.



**REFERENCIAS**

HOWELL, Alan. Perfect one day: digital the next: challenges in preserving digital information. *Australian Academic & Research Libraries*, v. 31, n. 4, p. 121-141, 2000. Disponible en: <<http://dx.doi.org/10.1080/00048623.2000.10755130>>. Acceso en: 03 set. 2016.

LOTS OF COPIES KEEP STUFF SAFE (LOCKSS). *How it works*. [S.l.], 2008. Disponible en: <[http://www.lockss.org/lockss/How\\_It\\_Works](http://www.lockss.org/lockss/How_It_Works)>. Acceso en: 11 jun. 2013.

MARDERO ARELLANO, Miguel Ángel. *Critérios para a preservação digital da informação científica*. 2008. 354 f. Tese (Doutorado em Ciência da Informação)- Universidade de Brasília, 2008. Disponible en: <<http://hdl.handle.net/10760/12649>>. Acceso en: 11 jun. 2013.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA - UNESCO. *Carta para la preservación del patrimonio digital*. Paris, 2003. Disponible en: <[http://www.arquivonacional.gov.br/conarq/cam\\_tec\\_doc\\_ele/preservacao/cartapreservacao.asp](http://www.arquivonacional.gov.br/conarq/cam_tec_doc_ele/preservacao/cartapreservacao.asp)>. Acceso en: 23 jun. 2013.

----- . *Recomendações sobre software livre para repositório e sistema de preservação*. Paris, 2007.

REICH, Victoria A. *Lots of Copies Keep Stuff Safe as a cooperative archiving solution for ejournals*. 2002. Disponible en: <<http://www.istl.org/02-fall/article1.html>>. Acceso en: 09 jun. 2013.

SKINNER, Katherine; SCHULTZ, Matt (Ed.). *A guide to distributed digital preservation*. Atlanta, GA: Enducopia Institute, 2010. Disponible en: <<http://www.metaarchive.org/GDDP>>. Acceso en: 11 jun. 2013.

## **Preservação digital e Rede Cariniana**

### **RESUMO**

*As atividades práticas para a preservação de documentos digitais têm como característica principal a colaboração entre instituições e especialistas. A verificação do grau de vulnerabilidade deste tipo de material e a necessidade de estabelecer responsabilidades entre criadores e curadores determinará o tempo de vida dos apoios físicos da informação em formato digital. A ciência precisa da preservação dos resultados de suas investigações para que possam ser usadas por outros cientistas no futuro; Há também a necessidade de salvar o patrimônio cultural digital, para garantir o acesso a esses recursos com funcionalidade de registros eletrônicos autênticos. A preservação da informação em formato digital precisa de um conjunto de práticas técnicas e administrativas em constante mudança. Os repositórios digitais são uma parte importante da preservação digital e, com eles, pode ser estabelecido um grau de controle dos procedimentos de criação, conservação e recuperação a longo prazo. A adoção de um modelo de rede de preservação digital é uma alternativa para organizações que querem coletar, armazenar, preservar e oferecer acesso à sua coleção em cópias digitais. Também é essencial combinar essas redes com os padrões internacionais já testados e promover a preservação digital da produção científica a longo prazo. Este trabalho é sobre o modelo de gerenciamento de preservação digital em rede distribuída que tem a característica de usar pelo menos seis instituições onde as cópias de documentos digitais são armazenadas. Uma experiência regional de preservação digital distribuída é apresentada, que segue o modelo de uma aliança internacional para salvaguardar a produção intelectual de um país. O modelo de negócios é o gerenciamento de informações em acesso livre, onde a perpetuidade da recuperação de dados é garantida pelo modo aberto de gerenciar os estágios de entrada, conservação e difusão em sistemas de informação eletrônicos confiáveis.*

**Palavras-chave:** *Preservação digital distribuída. LOCKSS. CARINIANA. Rede colaborativa. IBICT. Brasil.*

## **Digital preservation and the Cariniana Network**

### **ABSTRACT**

*Practical activities for the preservation of digital documents have as main characteristic the collaboration between institutions and specialists. The verification of the degree of vulnerability of this type of material and the necessity of establishing responsibilities between creators and curators will determine the life time of the physical supports of the information in digital format. Science needs the preservation of the results of its investigations so that they can be used by other scientists in the future; there is also the need to save the digital cultural heritage, to ensure access to those resources with authentic electronic records functionality. The preservation of information in digital form needs a set of constantly changing technical and management practices. Digital repositories are an important part of digital preservation and with them can be established a degree of control of the procedures of creation, conservation and recovery in the long term. The adoption of a model of digital preservation network is an alternative for organizations that want to collect, store, preserve and offer access to their collection in digital copies. It is also essential to match these networks with the international standards already tested and to promote the digital preservation of scientific production in the long term. This work is about the management model of digital preservation in distributed network that has the characteristic of using at least six institutions where copies of digital documents are stored. A regional experience of distributed digital preservation is presented, which follows the model of an international alliance to safeguard the intellectual production of a country. The business model is information management in free access, where the perpetuity of data recovery is guaranteed by the open way of managing the entry, conservation and diffusion stages in reliable electronic information systems.*

**Keywords:** Distributed digital preservation. LOCKSS. CARINIANA. Collaborative network. IBICT. Brazil.

## **SOBRE OS AUTORES**

### **Miguel Ángel Mardero Arellano**

Formou-se em Antropologia Social do Instituto Nacional de Antropologia e História (1985), mestrado em Ciências da Informação da Universidade de Brasília (1998) e doutorado em Ciências da Informação da Universidade de Brasília (2008). Ele trabalha no Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) desde 1997, É Tecnologista Senior é coordenador da Rede Brasileira de Preservação Digital Serviços CARINIANA. Ele é editor pelo Brasil do repositório internacional E-LIS. Tem experiência na área de Ciência da Informação, com ênfase na preservação digital, em particular nas seguintes áreas: sistemas eletrônicos de gestão da informação, publicações científicas eletrônicas, Open Journal Systems OJS, Dataverse, ICA-AtoM, Archivematica, repositórios digitais confiáveis, sistemas preservação distribuídos, LOCKSS e bibliotecas digitais. ORCID0000-0001-5306-919X

*E-mail:* miguel@ibict.br

### **Maíra Murrieta Costa**

Doutoranda em Ciência da Informação pela Faculdade de Ciência da Informação da Universidade de Brasília, com doutorado sanduíche na University of Michigan – School of Information – onde tem se dedicado ao estudo de políticas de informação para a gestão de dados de pesquisa. É Mestre em Ciência da Informação pela Faculdade de Ciência da Informação da UnB – momento em que se dedicou a estudar sobre a recuperação e armazenamento da informação em redes sociais. Possui um MBA em Administração Estratégica de Sistemas de Informação pela Fundação Getúlio Vargas (2005). É Bacharel em Biblioteconomia pela UnB (2001). É servidora do Ministério da Ciência, Tecnologia, Inovações e Comunicações há 11 anos. Atualmente desenvolve suas atividades na Secretaria de Desenvolvimento Tecnológico e Inovação como gestora do Sistema de Informações Estratégicas do Programa SIBRATEC – Sistema Brasileiro de Tecnologia.

*E-mail:* mairamurrieta@gmail.com

## **Murilo Bastos da Cunha**

Minicurriculo PhD em Library and Information Science pela School of Information - University of Michigan (1982). Possui estágio de pós-doutorado na School of Information - University of Michigan (1996). Foi presidente da Associação dos Bibliotecários do Distrito Federal e do Conselho Federal de Biblioteconomia. Aposentou-se do cargo de professor titular da Faculdade de Ciência da Informação da Universidade de Brasília em 2014, permanecendo como professor permanente do Programa de Pós-graduação em Ciência da Informação da UnB. Entre as atividades administrativas exercidas na UnB estão as de diretor da faculdade de Estudos Sociais Aplicados, Chefe do Departamento de Ciência da Informação e Documentação e diretor da Biblioteca Central. É autor de diversos livros na área de ciência da informação, entre eles: “Manual de Fontes de Informação” (2010); “Dicionário de biblioteconomia e arquivologia” (2008) e “Para saber mais: fontes de informação em ciência e tecnologia”. (2001).

*E-mail:* murilobc@unb.br

## **Sonia de Assis Boeres**

Doutoranda em Ciência da Informação pela Faculdade de Ciência da Informação da Universidade de Brasília. Tem se dedicado ao estudo sobre competências de profissionais da informação para preservação digital. É Mestre em Ciência da Informação pela UnB (2004), onde à época já trabalhou com as questões inerentes a políticas de preservação digital. Possui graduação em História pelo Centro Universitário de Brasília (UNICEUB, 1987) e também é graduada em Biblioteconomia, pela Universidade de Brasília (UnB, 1995). Foi professora substituta na Faculdade de Ciência da Informação. É consultora de gestão da informação com experiência nas áreas de preservação digital, biblioteca escolar, acesso à informação, profissional e gestão da informação.

*E-mail:* sonia.boeres@gmail.com



## **Anaiza Caminha Gaspar**

Mestrado em Ciência da Informação, UnB; DEA em Veille Scientifique et Technique, Université Aix-Marseille III, Inteligência Competitiva UFRJ, Coordenadora de Projetos do Instituto Brasileiros de Informação em Ciência e Tecnologia - IBICT (1987 - 2014); Consultora Unesco - Projeto Open Science, 2014.

*E-mail:* anaizagaspar@gmail.com

## **Lillian Maria Araújo de Rezende Alvares**

Doutora em Ciência da Informação pela Universidade de Brasília e pela Université du Sud Toulon-Var (Doctorat de Sciences de l'Information et de la Communication) em regime de cotutela. Especialista em Inteligência Competitiva pela Universidade Federal do Rio de Janeiro (UFRJ), Instituto Nacional de Tecnologia (INT) e Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) em 2001. Mestre em Biblioteconomia pela Universidade de Brasília (1997) e Graduada em Engenharia Mecânica (1990) pela mesma universidade. Trabalhou na Fundação Centros de Referência em Tecnologia Inovadora (Certi); Ministério das Relações Exteriores (MRE); Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict); Empresa Brasileira de Pesquisa Agropecuária (Embrapa); e Ministério do Desenvolvimento, Indústria e Comércio Exterior (MDIC). Professora da Faculdade de Ciência da Informação da Universidade de Brasília (UnB), desde 2006, onde também exerceu e exerce cargos administrativos como Coordenadora do Curso de Bacharelado em Museologia de abril de 2010 a abril de 2012; e Coordenadora do Programa de Pós-Graduação em Ciência da Informação (PPGCI/UnB) de junho de 2012 a janeiro de 2014. Membro do Conselho Fiscal da Ancib na gestão de 2006-2008. Vice-Presidente da Ancib na gestão 2012-2014. Atualmente é Coordenadora-Geral de Pesquisa e Manutenção de Produtos Consolidados do Ibict.

*E-mail:* lillianalvares@ibict.br

## **Maria de Nazaré Freitas Pereira**

Possui graduação em Biblioteconomia pela Universidade Federal do Pará (1967), mestrado em Ciência da Informação pela Universidade Federal do Rio de Janeiro (1981) e doutorado em Ciências Humanas: Sociologia pelo Instituto Universitário de Pesquisas do Rio de Janeiro (1997). Tem cursos de especialização em Transferência de Tecnologia e Desenvolvimento, Science Policy Research Unit (SPRU) – Sussex University, Brighton – Inglaterra (1989 /1990) e em Sistemas e Redes de Informação, Universidade de São Paulo/ Escola de Comunicação e Artes (1973). Atuou na implantação do Programa Sociedade da Informação no Brasil, na condição de Coordenadora de Inclusão Digital (2001/2003). Realiza estudos de consultoria para o IBICT sobre sistemas CRIS (Current Research Information System), no período 2014-2015. Sua experiência é nas áreas de Sociologia da Ciência e de Ciência da Informação, com ênfase em comunicação científica, produtividade de pesquisa e tecnologias da informação e comunicação (TICs) para o desenvolvimento, atuando principalmente, na condição de consultora independente, nos seguintes temas: metodologia da pesquisa acadêmica, através de ensino de método de sua criação – Carpintaria do Trabalho Acadêmico, inovação em ciência da informação, sistemas de informação de pesquisa em andamento (CRIS), desenvolvimento de conteúdo, editoração científica e inclusão digital para o desenvolvimento.

*E-mail:* marianazare@uol.com.br

## **Charlley dos Santos Luz**

Charlley Luz é arquivista, mestre em ciência da informação, professor da pós em gestão de documentos da FESPSP e consultor em estratégia digital e negócios digitais do Sebrae e Feed consultoria.

*E-mail:* charlleyluz@gmail.com

### **Fernando William Cruz**

Possui graduação em Processamento de Dados pela UnB (1988), mestrado em Informática pela UFPB (1992) e doutorado em Ciências da Informação pela UnB (2008). Tem experiência de mercado na área de TIC, com ênfase em desenvolvimento e suporte de sistemas de informação. Atualmente é professor adjunto da Universidade de Brasília com as seguintes atuações: (i) na Graduação em Engenharia de Software - FGA em disciplinas básicas de computação (sistemas operacionais, sistemas distribuídos, estruturas de dados, arquitetura e redes de computadores), (ii) na Pós-Graduação em Ciência da Informação - FCI realiza pesquisas voltadas para Organização e Representação da Informação, Estudos de Usuários e Bibliotecas Digitais. Participações recentes (desde 2012) em projetos ligados a sistemas colaborativos e web semântica.

*E-mail:* fwcruz@unb.br

### **Juliana Rocha de Faria Silva**

Possui licenciatura em Música pela UnB (2000), mestrado em Educação Musical pela UnB (2010) e doutoranda em Ciência da Informação da UnB desde 2013. Atualmente é professora de ensino básico, técnico e tecnológico do Instituto Federal de Educação, Ciência e Tecnologia de Brasília (IFB) -Campus Planaltina. Atua no ensino de Música no Ensino Médio e em projetos de pesquisa e extensão no ensino coletivo de instrumentos musicais e canto.

*E-mail:* jurfsilva@gmail.com

### **Luiza Beth Nunes Alonso**

Bacharel em Ciências Sociais pela Universidade de São Paulo/Brasil. Mestrado e Doutorado pela Harvard Graduate School of Education, Cambridge, MA/USA na área de Administração e Planejamento de Políticas Públicas. Pós-Doutorado em Epidemiologia e pesquisa na área de Prevenção e Promoção da Saúde, pela Universidade da Califórnia/ USA. Atualmente é professora, pesquisadora e coordenadora do Mestrado em Gestão do Conhecimento e Tecnologia da Informação da Universidade Católica de Brasília, com pesquisas em andamento na área de Gestão Social do Conhecimento, em especial sobre a disponibilidade de informações e seu uso no encaminhamento de propostas para as áreas de educação básica, saúde e ambiental. Membro do Centre International de Recherches et Études Transdisciplinaires – CIRET, e membro fundadora do Centro de Educação Transdisciplinar – CETRANS.

*E-mail:* lualonso@ucb.br

### **Vicente González García**

Engenheiro de sistemas, especialista en formulación y evaluación de proyectos y magister en gestión de proyectos.

Areas de actuación: gerencia de proyectos, planeación estratégica, sistemas de información, inteligencia de negocios.

*E-mail:* vgonzalez@unisalle.edu.co

### **Johann Enrique Pirela Morillo**

Licenciado en Bibliotecología y Archivología. Magister en Museología, Magister en Educacion. Mencion: Currículo. Doctor en Ciencias Humanas. Postdoctorado en Ciencias Humanas.

Areas de actuación: Investigación y docencia de pregrado y postgrado en Bibliotecología, Archivología y Ciencia de la información, diseño y evaluación curricular, patrimonio cultural.

*E-mail:* jepirela@unisalle.edu.co

### **Nelson Javier Pulido Daza**

Bibliotecólogo y Archivista, Especialista en Sistemas de Información, Abogado, Candidato a Doctor, Doctorado en Líneas de Investigación en Información y Documentación..

Áreas de actuación: Gestión de proyectos de consultoría en documento electrónico, docencia e investigación de pregrado y postgrado, diseño y evaluación curricular en Bibliotecología y Archivística,

*E-mail:* npulido@lasalle.edu.co.

### **Maria Janneth Alvarez Alvarez**

Profesional en Bibliotecología y Archivística de la Universidad de La Salle. Especialista en Archivística de la Pontificia Universidad Javeriana. Magister en Docencia de la Universidad de La Salle. Estudiante del Doctorado en Historia en la Universidad Nacional de Colombia. Profesor del área de archivística con énfasis en gestión documental, historia de los archivos, fundamentos de la archivística y práctica profesional.

Área de actuación: Docencia e investigación en archivística, gestión documental.

*E-mail:* mjalvarez@unisalle.edu.co

### **Irene Sofía Romero Otero**

Ph.D Información y Documentación en la Sociedad del Conocimiento por la Universidad de Barcelona y Comunicadora Social y Periodista por la Universidad Surcolombiana. Realizó su tesis en el campo de la Comunicación, Cambio Social y Desarrollo en la Universidad Complutense de Madrid. Profesora Asociada de La Universidad de La Salle, Bogotá, Colombia.

Áreas de actuación: Edición científica del libro electrónico, recursos electrónicos, sector editorial.

*E-mail:* isromero@unisalle.edu.co

## **Adriana Otalora-Buitrago**

PhD Estudios Políticos, Universidad Externado de Colombia; MSc Educación y Desarrollo Humano, CINDE-Universidad de Manizales; Politóloga y Economista, Universidad del Rosario, Bogotá, Colombia.

Profesora Asociada de La Universidad de La Salle, Bogotá, Colombia.

Áreas de actuación: Ciudadanía, Paz, Víctimas de la Violencia, Subjetividades Políticas, Políticas Públicas con enfoque diferencial, Capacidades Humanas.

*E-mail:* aotalora@unisalle.edu.co, adrianaotalora@gmail.com

